Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Export Decision Support Model for the Czech Republic

Author: **Carlos Couceiro Vlasak**

Supervisor: **prof. Ing. Michal Mejstřík, CSc.**

Academic Year: **2015/2016**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently; using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 31, 2016

Signature

Acknowledgments

.

I am grateful to prof. ing. Michal Mejstřík CSc for supervising my work on this thesis and the EEIP team for their valuable comments

Abstract

In this paper, an Export Decision Support Model applied to the Czech Republic is developed, with the aim of finding export opportunities. The model functions using a filtering process in which a stream of data composed of numerous socio-economic indicators representing the world trade is analysed. For their construction, an extensive literature review was developed as at the moment, no explicit rule exist describing its appropriate composition. Then, if a given market, determined by its associated matrix of indicators, fulfils the conditions of the model, then it is retrieved as an export opportunity. After the model construction, it is supplied with two streams of data, for 2010 and for 2014 and, the hypothesis that for both years the output is equal is evaluated. With the intention to infer if the constructed model needs periodical recalibrations for its appropriate use. Finally, a local sensitivity analysis is deployed uncovering the behaviour of the different parameters of the model, a novel approach not yet implemented in an EDSM tailor made for the Czech Republic.

JEL Classification F10, F13, F23, M31

Keywords export opportunity, entrepreneurship, international marketing, sensitivity analysis, trade

Author's e-mail karel.couceiro@gmail.com

Supervisor's e-mail mejstrik@fsv.cuni.cz

Contents

List of Tables

List of Figures

List of Abbreviations Used

AIC Akaike Information Criterion

ANN Artificial Neural Networks

BIC Bayesian Information Criterion

CETA Comprehensive Economic and Trade Agreement

DSS Decision Support Systems

EDSM Export Decision Support Models

ESI Export Similarity Index

FVA Foreign Value Added

GDP Gross Domestic Product

GNP Gross National Product

GVC Global Value Chain

HHI Herfindahl index

HS Harmonized System

IT Information Technology

LPI Logistic Performance Index

MZV Ministry of Foreign Affairs of the Czech Republic

OECD Organization for Economic Development and Cooperation

ONDD Office National Du Ducroire

RCA Revealed Comparative Advantage

R&D Research and Development

SITC Standard International Trade Classification

TTIP Transatlantic Trade and Investment Partnership

WITS World Integrated Trade Solution

YoY Year over year

1. Introduction

"In God we trust, all others must bring data"

William Edwards Deming

The task to find export opportunities for the exporters of a given country is a constant issue in economic history. But since the development of computer technology, some new instruments make this task easier. More in particular, Decision Support Systems have appeared, allowing for extensive and deep data processing so that final users can face fewer obstacles when assessing a final choice (for instance, the election of a profitable market) based on rational arguments through logical inference. Examples of these tools in the current literature are numerous and increasing; some of them will be described in the literature review section.

Acknowledging that still present necessity, the purpose of this Master Thesis is the development of an EDSM (Export Decision Support Model) for the Czech Republic. It will be supplied with two datasets for 2010 and 2014 with the initial hypothesis that for both years, the set of outputted opportunities will be equal (an explicit definition of what is understood as equality is provided) inferring, if the model needs regular actualizations or not. An export opportunity is understood at a given time interval, as a foreign market for a given product which an exporter could join and be successful in it, therefore the hypothesis states that the model will select long term opportunities rather than short. This task requires the analysis of all the markets in the entire world. However, as it is too effortful to be developed in detail, an EDSM (Export Decision Support Model, Decision Support System targeted to find export opportunities), which will implement a data filtering process, is established. Without going extensively into details, only some economic indicators which represent each market (country-product combination) are assessed (the idea is that if the markets are similar based on a given number of parameters, then they will also be similar in the remaining ones). At the end, two groups of markets are retrieved by the model: those whose indicators do not fulfil the requirements stated in the model, and those which pass and are assessed as export opportunities.

That implemented methodology was designed thanks to the ground provided by previous researchers as it will be clear in the literature review section with some additions in order to improve its accuracy as much as possible. More in particular, this thesis relies strongly on the EDSM designed by Urban et al (2014a; 2014b) as it was tailor made to the Czech Republic and it is quite recent. In addition, since its publication, new studies have appeared such as a report by the Ministry of Foreign Affairs of the Czech Republic (Tlapa, Klepácek and Svoboda, 2015) and new data are available. It is necessary to mention that until now, all the previous methodologies including that one, even relying as much as possible on empirics and theory, have a subjective component. The reason behind is that no scientific rule has been established yet in that research area as it is not possible yet to predict the future accurately. As such, this thesis will try to construct the most objective EDSM as possible, comparing the results through two datasets and providing a robustness assessment; a novel approach not yet implemented in an EDSM tailor made for the Czech Republic.

In the first section, Literature Review, a critical assessment is done regarding the most prominent market selection tools existing in the current literature and highlighting positive and negative points and, finally choosing the best framework for the purpose. In the second section, Methodology thanks to the ground provided by previous researchers in addition to some contributions gathered in order to reflect the concrete characteristic of the Czech Republic, the particular procedure implemented is described.

After the first implementation of the EDSM and a description and interpretation of the results; a sensitivity analysis is carried out to check the overall robustness of the model and explore it more profoundly given the available datasets, the behaviour of this tool and a possible scope for improvement; acknowledging the inherent and inevitable subjectivity of the model. And finally, both results (for 2014 and 2010) of the EDSM are evaluated and compared, inferring if the model needs periodical actualizations.

2. Literature Review

In an increasingly globalized world, companies are in constant search of profitable foreign markets in which they could expand and acquire economies of scale, improving their productivity, while diversifying their portfolio of customers. As it is widely acknowledged, the success of the internationalization process and, to a large extent the overall subsequent success of the firm depends on the correct selection of foreign markets (Papadopoulos et al, 2002).

In the current academic literature, a large variety of different approaches exist for the selection of profitable foreign markets. On the one hand, small firms lacking resources usually apply the opportunistic view (Bilkey, 1978) for their market expansion, joining foreign markets thanks to some request for their products; or, they may follow the recommendations drawn by public agencies. On the other hand, larger firms in their market expansion strategies, usually apply a more proactive approach. As such, they frequently engage actively in their expansion process through the implementation of explicit decision support models (among other processes). The aim of the majority of those models is to determine a set of relatively attractive markets (Papadopoulos, et al, 2002), before a final in-depth examination is carried forward (others, directly retrieve a group of export opportunities) in which export opportunities are determined (understanding by export opportunity a market for which an export venture is worth starting). As an in-depth examination of all the markets of possible interest would be too costly and time consuming, it is preferred to select an *a priori* group of promissory markets which could be further scrutinized later on.

A review of existing models for market selection was carried out by Papadopoulos and Denis (1988) classifying each of them to a different category. Being that proposed framework (a structure by which existing models are classified in categories) described in recent papers such as Cuyvers et al. (2009). This framework (figure 1) divides all the existing models into two distinct classes, regarding if the models are qualitative or quantitative. Nevertheless, other authors have delineated different schemata by which they categorize all existing models, actually being used by international marketing researchers and corporate management boards. For instance, Dimitratos et al. (2010) take the view of dividing all the existing models into systematic or non-systematic. Those authors (Dimitratos et al., 2010 p. 591) identify systematic models that include objective criteria (Brouthers and Nakos, 2005) and activities such as the following: the visit of foreign countries, the monitoring of the international business press for product-related activities and the use of primary and secondary data sources. These authors argue that those models, "have received little empirical support" because of "the difficulty in empirically testing them" (Brouthers and Nakos, 2005 p. 366). Besides that, those models are still frequently used by a large number of different firms, governments and international organizations around the world.

Accordingly, the degree of attained objectivity as will be made clear in the next sections where the most relevant models found in the literature will be described, drastically varies. An empirical ground could be given by trade theories, but given their limitations, they are not the basis through which those models are constructed. The reason for this lies in the fact that trade theories are broad frameworks that merely describe trade in ideal situations as they are abstractions of reality (Onkvisit and Shaw, 2004). Even though, all those models are based on theory, at least in part; they include variables which, according to economic theory are suitable and/or related for the selection of possible export opportunities. However, for those included variables considered, at present there is no general consensus regarding which ones to use, how they should be weighted in order to reflect their relative importance and/or how they could show their interaction effects (Papadopoulos, et al, 2002 p 167; Andersen and Strandskov, 1997; Gould, 2002). Therefore, at the moment no existing model appears to be superior over all the remaining ones (Papadopoulos, et al, 2002), in all circumstances.

Leaving all those weaknesses, since it is not possible to predict the future (there are always risks and unpredictable factors which we are not yet capable to assess), a structured approach for the market selection prior to an in-depth assessment is really necessary as there is evidence that it increases the export performance for different companies (Brouthers and Nakos, 2005). Those corporations which expand their business abroad usually invest large amounts of resources. As a result, managers, do need to have the best possible information. Those models are still in place in order to help as much as possible the final decision makers. In addition, many of those models usually incorporate internal knowledge and experience of the agents of interest as they are tailor made for their specific needs.

On the other hand, non-systematic approaches are described in such a way that they are implemented through "rules of thumbs", relying on "personal feelings and intuition" and involving little or no information, targeting markets which "minimize some psychological distance" (Brouthers and Nakos, 2005, p 364; Dimitratos et al., 2010; Andersen and Buvik, 2002 p 348). Those models are used when firms do not have enough resources. They are considered to be much more unreliable and those firms that use them, have worse export performance (Brouthers and Nakos, 2005, p 364) than those which use more systematic approaches; however they are better than nothing.

As explained earlier, the review done by Papadopoulos and Denis (1988) (since then, different models have appeared; nevertheless this framework is still applicable nowadays) classifies all existing models into qualitative and quantitative. Within those groups we encounter further delimitations as can be observed in figure 1. Those further delimitations will be explained in detail in subsequent sections.

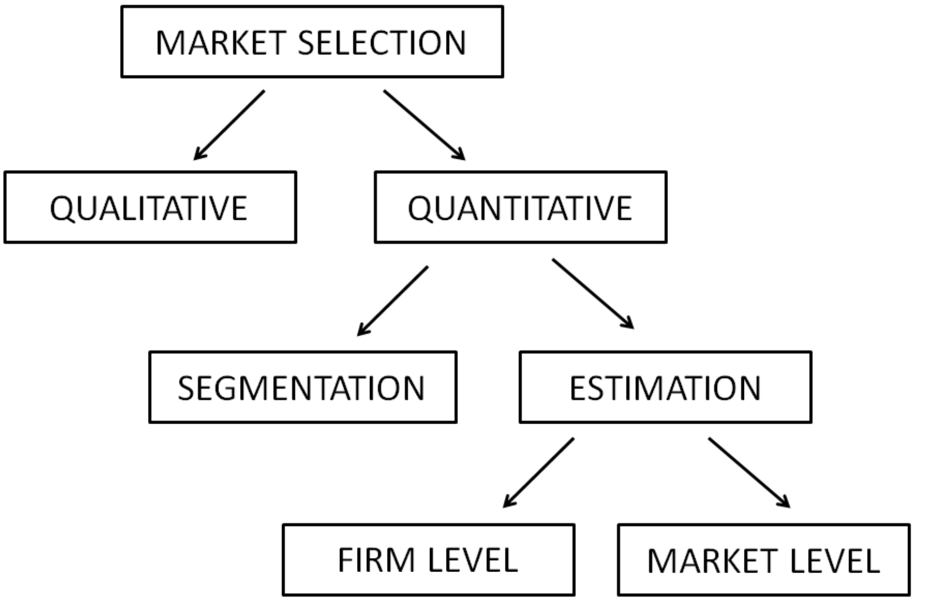

Figure 1: Market Selection Methods

Source: Author based on Papadopoulos and Denis (1988)

Qualitative models as their name indicates are based on opinions coming from internal sources of the company or from international organizations. As can be easily expected, those models can lead to inconsistencies (for instance, they can incorporate many cognitive biases[1] inherent in human nature and excessive reliance on heuristics[2]) and/or imprecisions, so they are not advisable. Inherent the subjectivity in the international market selection process, those models potentially increase it.

Contrastingly, quantitative approaches (those based primarily on objective and quantifiable criteria), comprise a large number of different models (Cuyvers et al., 2009), two main categories are identified. On the one hand, there can be observable

[1] "Systematic error in judgment and decision-making common to all human beings which can be due to cognitive limitations, motivational factors, and/or adaptations to natural environments" (Wike and Mata, 2012, p.531).

[2] "Judgment or decision-making mechanism or cognitive shortcut that relies on little information and modest cognitive resources" (Wike and Mata, 2012, p.531).

market grouping methods based on market segmentation (Andersen and Strandskov, 1997). On the other hand, we find in the literature many examples of market estimation methods which try to assess the potential demand for each concrete market (for instance, country-product combination). The basic idea behind market grouping methods, is that following the Uppsala model (Forsgren, 2002) of international market expansion (this is the most widely used model of international market expansion; other two widely cited models are the economic or eclectic and the network approach (Chandra et al., 2009).), the most similar markets in terms of culture, social distance and economy are usually targeted first (Steenkamp and Ter Hofstede, 2002). The targeting of near markets is intended to reduce the most possible uncertainties. Distant markets for an unexperienced company (not taking into account for this analysis, transportation neither other barriers) in terms of the previous cited variables are harder to enter and the firms are neither confident nor capable to operate in them effectively. The reason behind is that they tend to lack appropriate knowledge[3]. The crucial statement of the Uppsala model is that firms are engaged in a learning process through their foreign market strategy, as they progressively enter more distant markets (Forsgren, 2002). There are many possible statistical techniques for the implementation of those grouping models, ranging from cluster to factorial analysis. The number of quantitative variables selected will depend on the data availability as well as on the final chosen technique[4] (Cuyvers et al., 2009). Unlikely in supervised learning (where it can be computed an error rate), as there do not exist any explicit criterion; it is difficult to consider a given subset of variables to be preferred over another so it is usually up to the researcher to decide which ones to use. The result of this methodology is a grouping of all the existing markets into different categories, according to their similarity, defined by a set of chosen variables. As a result, the most similar markets are identified and the

[3] In addition, habitually for firms distant markets *ceteris paribus* involve greater costs because, for a group of similar markets, companies can enhance economies of scale and scope. For instance, they can implement the same marketing strategy, as they operate in the same cultural environment.

[4] See figure 10 for an example of a clustering method; different corporations can use different variables and choose other markets of interest. The idea is that the expansion of the corporation will be performed sequentially, from the most to the least similar markets.

corporation can target them following the recommendation drawn by the previously depicted Uppsala model.

This approach has been criticised because it usually relies only on macro level variables so trends and heterogeneities are neglected (Sakarya et al., 2007; Cuyvers et al., 2009) as the definition of what a market is tends to be broader than in the case of the majority of other models. As a result, explicit market opportunities and possible niches within countries are not easily identified. Accordingly, this method is appropriate just for the assessment of the overall environment, as a first exploratory step, not at the product level.

A second category within the quantitative approaches is formed by, the market estimation methods (Cuyvers et al., 2009). Those models first evaluate all the possible markets and then select those suited according to their potential and attractiveness. Further segregation is performed encountering afterwards, different categories. According to Cuyvers et al. (2009), two broad categories are encountered: those which take the view at the firm level and those which take the view at the country level, targeted to various industries (as we can encounter a huge variety, only the most relevant will be described). A concrete category, within those which take the view at the firm level and, which deserves a closer look because of its implementation and popularity, is the portfolio techniques. Those models share similarities with the grouping techniques as they also tend to assign markets to different categories. Nevertheless, they are more detailed as they tend to take into consideration, firm and market specific indicators (as for them it is frequent to define a market in the narrowest sense, country product combination or even city product combination); furthermore, their purpose is not a grouping but the estimation of the potential demand so niches can be identified. A first example which only uses two variables, is the growth share matrix (Keegan and Schlegelmilch, 2001), first developed by the Boston Consulting Group (BCG) in the 1960's. That method is represented by a 2*2 matrix (figure 2) in which all markets (of possible interest for a corporation) are assessed according to their growth rate and the current relative market shares the interested entity for which this model is developed in the given market. Obtaining finally a division into four different groups (cash cow, star, dog and question mark) (Hindle, 2003, p.111).

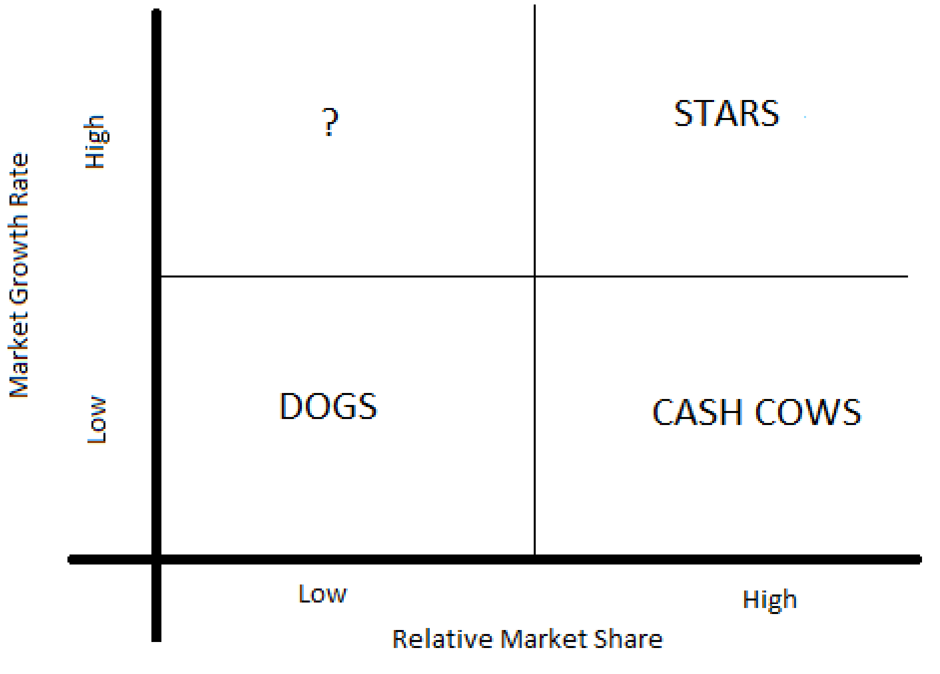

Figure 2: Growth Share Matrix

Source: Author based on Hindle (2003, p111).

Another approach is given by the multifactor portfolio techniques, for instance, the market attractiveness - business position assessment, or GE-McKinsey nine-box. Following the same reasoning as for the previous model, markets are assessed regarding their attractiveness and business position. In this case, those indicators do not represent simple variables but are composite indicators of many different variables. Thus, each indicator is a linear combination of several weighted factors, tailor made for each situation. Each company chooses its concrete factors and weights, with the possibility that some factors are determined subjectively. It is also worth mentioning that this framework has many possible implementations and can be employed even by countries, not only by firms (being used as well at country level). An example of this, is the matrix developed by the Spanish Secretary of Commerce (Estrategia, 2005) in which this framework is applied for the entire Spanish economy; here markets are defined not as country product combinations but only as countries (the major weakness still prevails so that the possible omission of significant variables and the weights were constructed *ad hoc*).

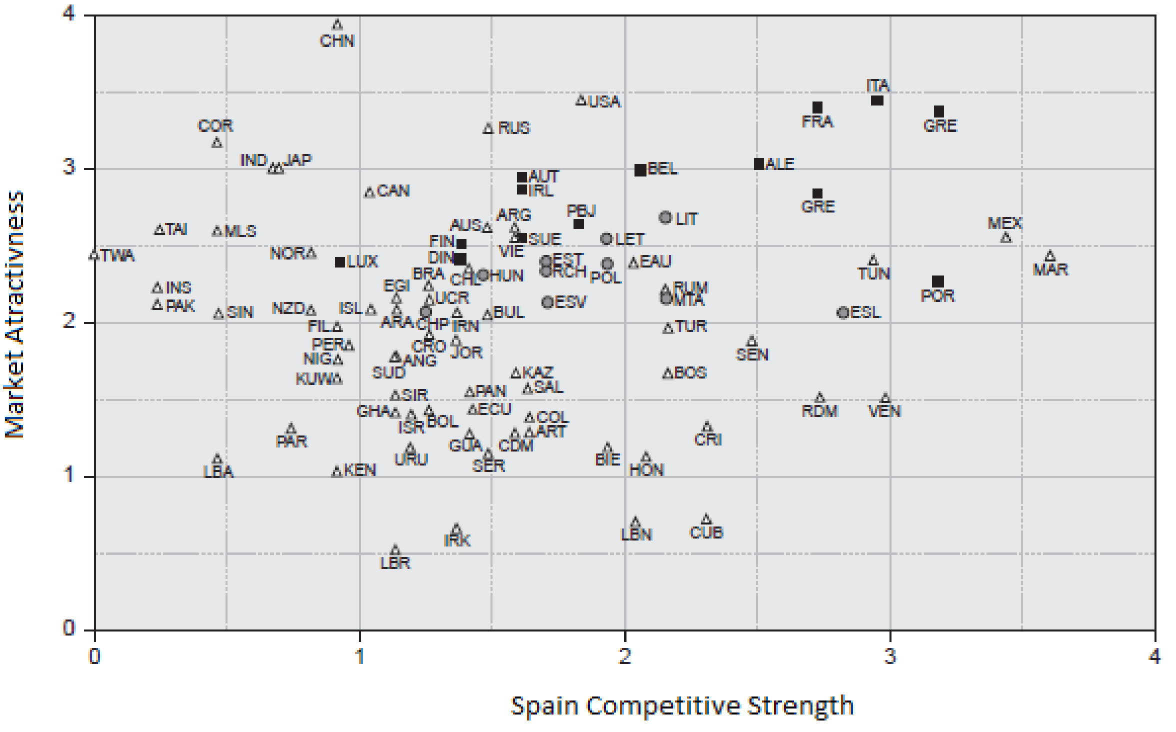

Figure 3: Spain's export markets assessment

Source: Estrategia, S. G. D. A., (2005)[5]

An important category[6] of support models at the firm level is that one which comprises the models which use in a systematic way the internal knowledge of the company for the selection of export opportunities, structuring their acquired information in order to make the best possible predictions for the future. A good example is the one provided by Ozorhon et al (2006) where a case base reasoning model for international market selection is developed. This model was conceived with the idea that past experience brings solutions to new and, novel situations. Thus, data was collected from various firms, where different historical cases (export

---

[5] As the original article from which this figure is extracted is in Spanish, the abbreviations for the different countries are also written in Spanish.

[6] There are a huge number of different models with their respective implementations, as such in this review only the most important ones are described.

ventures) were stored in a database. The model assessed proposed prospective markets by retrieving the most similar case from the past, predicting their profitability and the competitiveness of the company in that market. An advantage of that model is that it allows for continuous updating. Weaknesses remain that no clear indication on which variables to use exist, with the possibility that the cases are similar regarding all the selected variables but not similar with respect to the ones omitted by accident or simply because they are not measurable. And, as the dataset is comprised of historical experiences, there exist the possibility that some necessary (for the purpose of this selection model) export ventures would be missing as the management board should be reluctant to implement them into practice if their aim is to maximize the value of the firm. As a result, given a prospective market, the most similar case retrieved by the model could be one that does not match correctly (being the most similar in the database but not in *sensu stricto*); finally drawing incorrect predictions. Consequently, no guarantee exists of correct assessment and correct model generalization yet as it is not guaranteed that the model would be capable to correctly capture the required complexity.

A related model at the firm level is a neural net proposed by Dikmen and Birgonul (2004). Using a supervised machine learning algorithm (a neural network[7]), it predicts the attractiveness of and the possible competition encountered in a market (still being based on historical data). The biological analogy of that algorithm is not casual as it is designed to learn (similar to a brain) from a dataset and then apply that knowledge.

Provided a dataset comprised by labelled (opportunity, not opportunity) variables accompanied each by a set of features, the algorithm first is trained computing a matrix of weights (there are various ways how to train a neural network, the one which will is described is backpropagation). In such a way that afterwards, provided the set of features for each prospective market, the algorithm predicts its

[7] The more detailed description of that methodology goes beyond the scope of this thesis. For further information and exact notation, please consult specialized research articles such as Dreiseitl and Ohno-Machado (2002) or Friedman et al. (2001).

corresponding label. The learning phase is performed in such a way that the error rate is minimized, predicted versus real the label of each variable.

In other words, the aim of that algorithm is to be able to generalize the knowledge acquired through the initially provided dataset. So, given new data entries, new markets to assess determined by a set of features (this time not being the labels visible), correctly assess the corresponding labels (in the case of the example provided by Dikmen and Birgonul (2004), attractiveness and the possible competition of a market, inferring if it is an export opportunity or not).

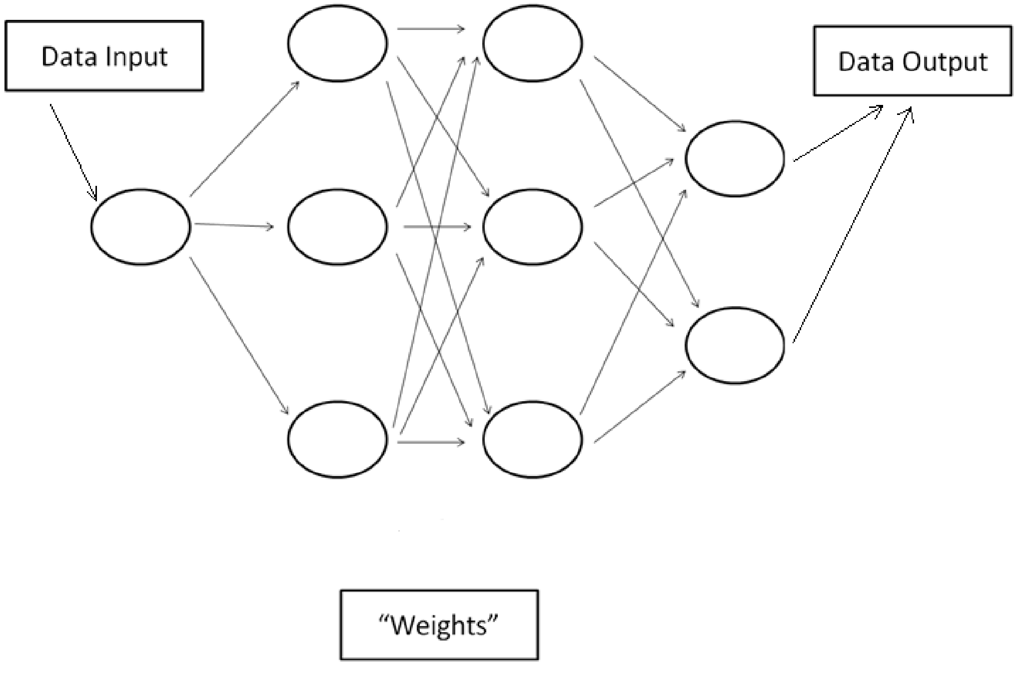

Figure 4: Neural Network

Source: Author based on Friedman et al. (2001)

In figure 4, we observe a graphical example of that methodology[8], the data is entered into the first perceptron (depicted by a "circle") and then it is retrieved as an output, in this case through two perceptron. The data "travels" as observed; in each perceptron, it is transformed through a function, each having a set of weights or parameters. As explained, those parameters are recursively modified through a learning process during the training of the algorithm, in order to minimize the error rate as much as possible. The common approach is to use for each perceptron the same function which is usually the sigmoid. In the case presented (1), Θ is the matrix of weights and X is the matrix of the input being $H_{\Theta}(x)$ the output.

$$H_{\Theta}(x) = \frac{1}{1 + e^{-\Theta^t x}} \qquad (1)$$

Sigmoid Function

In this algorithm, the vector of the inputs can be any combinations of useful features which the author seems convenient, normally taking into account field knowledge; some machine learning implementations use up to thousands of features. The features are usually selected in order to achieve the minimum possible in and out of sample error rate (in sample error rate is understood as the error committed over the dataset used for training, while out of sample is the error rate the algorithm makes when it is supplied with a new dataset) while maintaining a parsimonious number of them, in order to achieve generalization. Moreover, in that field numerous sophisticated approaches for the feature selection task as indicated in Verikas and Bacauskiene (2002) were developed. Even, some machine learning algorithms such as decisions trees are capable of choosing the most informative features (Guyon and

[8] In figure 4 we observe a particular representation, usually being up to the researcher to determine the exact number of perceptron.

Elisseeff, 2003) and to neglect the remaining ones, based on the reduction in entropy (information uncertainty) that they provide.

As can be observed, the matching methodology presents, in fact, improvement over the method presented by Ozorhon et al (2006). This improvement is derived from the complexity and demonstrated accuracy of that classification method, and that the selected features and its usefulness can be evaluated. At the same time, it still allows for updates, as given new data, this algorithm can be verified (observing if out of the sample error rate is excessive) and if necessary, recalculated.

In the case presented as illustration by Dikmen and Birgonul (2004), the algorithm was trained thanks to the stored knowledge of 600 Turkish construction firms (the last two presented methods can be easily generalized to the particular domain of that research). Nevertheless, in that paper we encounter proposed further methods for the obtaining of training data such as: the creation of the dataset thanks to domain experts, acquiring them from simulation results or from hypothesis results (Dikmen and Birgonul, 2004). For this category of models, subjectivity needed for their construction is minimal. But, a major issue remaining is the lack of enough and unbiased data which would map all the possible scenarios correctly, allowing therefore to the model to produce accurate results. In addition, the generalization capability of those methods is not guaranteed, given that the international environment is in constant transformation.

More in particular, the dataset is believed that will be biased as logically, (for solvency issues, if the corporations not want to bankrupt) unsuccessful cases there would be much less than successful. And, a domain adaption would be needed. Concretely, the dataset would be composed by attempted markets by the corporations, but prior the attempt a company idiosyncratic filtration process is believed that is applied (for instance, they would not attempt markets they believe that will be unsuccessful if they are risk averse); so, there will be an exogenous variable. Some approaches have been developed recently tackling both problems, for a review of them, it is recommended the study done by Patel et al (2015).

The majority of the remaining models at the firm level use the framework depicted in Andersen and Strandskov (1997) and Cavusgil (1985); a screening technique (also called filtering method or *go no go* analysis), figure 5 illustrates an example at the country level (as observed, this method is both used at the country and at the firm level), the name being highly illustrative. According to this, three structured stages prior to the final decision of the management board are followed. Concerning the first, a screening phase of unrequired markets (a criterion is stated and those markets which do not fulfil it, are eliminated from further consideration), then a selection phase or ranking of the remaining ones and a final identification phase where those remaining markets from the previous phases are comprehensibly evaluated and each firm chooses those which fit best their interests (as indicated by Brouthers et al. (2009), those multinational enterprises which select international markets taking into consideration their firm-specific advantages, display on average better performance). There is another technique that is both used at the country and at firm levels for the evaluation of prospective markets. It is an econometric approach, an example is Armstrong (1970) in which an econometric model is performed by estimating the future size of a market in order to assess their sales potential. As the ulterior intent of this research is to find export opportunities for the whole Czech economy (but not for the service sector); this framework will not be implemented. As this framework can be used to estimate a new variable for each market, but only that variable (for instance the future size of the market) is believed to not be enough to categorize a market as export opportunity (Gould (2002, p 270), "a well-informed selection decision could consider over 150 measures of markets competitive, economic, political, legal, cultural, technical and physical environment"), and, perform separate regressions for each market is not feasible as there exist over 1000000 possible country-combinations; implied the necessity to guarantee for each regression a correct model specification, being the omitted variable bias a constant threat. In addition, that method shares the inconveniences of the two previous approaches based on accumulated knowledge. Such as, for the training of an explicit econometric model predicting if a market is an opportunity or not; it would be necessary to supply the model with a training set containing as well data defining markets explicitly as non-opportunities not only markets identified as opportunities, being its time-stability not guaranteed.

On the other hand, the country level models are targeted to be used by governments and international organizations in their effort to find export opportunities for entire countries, there are many actual proposed frameworks in this line. According to Papadopoulos et al. (2002, p. 168), the only model within that framework which appears to be "simple, industry specific, generalizable and empirically validated", is the shift-share model but they finally assert that the model lacks predictive power. It is based on the setting of two points in time and the calculation of the import growth for each market, then obtaining the expected growth for each market (as the average of all markets). As a result, the promise markets are identified so that their actual growth (given those two selected points in time) is greater than their expected growth (given all markets within their category). After that, a ranking is performed and those markets for which the shift is the highest (actual versus expected growth), are assessed as export opportunities. The assertion that the model lacks predictive power is connected with the fact that it is highly dependent on the two points in time chosen and that it is sensitive to outliers.

Apart from the previously mentioned models, there are also many others within that category; it is important to point out that they are variations of the already explained ones with minor amendments (portfolio technique and screening process). For instance, the trade-off model proposed by Papadopoulos, et al. (2002) uses two composed indicators and divides all markets (country product combinations) into four categories. As can be observed, it is a clear analogy to the portfolio framework already proposed by McKinsey-GE, this time using different composite indicators (markets attractiveness and trade barriers).

Innovative in their article is Papadopoulos, et al. (2002) who tries to verify their model empirically. For a given year, they find correlations between the country scores for each product and in subsequent year's variables such as import growth, market share growth, import size and market share size. The results finally obtained cannot be regarded as conclusive regarding either acceptance or rejection off the validity of the model. As mentioned above, those types of models are almost impossible to be validated. Nevertheless, it is extremely necessary to make those models as objective as possible as the future of many enterprises relies on their accuracy. However, this validation framework, should not be used in the final model as it relies only on correlations of a composite index with only one variable at each time so it does not provide definitive results (what it is important is a multitude of variables and their interactions effects).

It is worth mentioning the recently published report by the Ministry of Foreign Affairs of the Czech Republic (Tlapa, Klepácek and Svoboda, 2015). This report is freely available and provides at the country level a complete list of export opportunities. If it were classified within the depicted framework, it would then be a quantitative estimation model at the country level. However, it relies heavily on the corroboration provided by the Czech Foreign Offices and on the cooperation of the business sector (for which the model is intended). The methodology consists, first, of a selection, from the world perspective, of promissory markets not yet extensively exploited. The variables considered are the following: the growth dynamics (the market is growing steadily, at a good rate and not decreasing), the potential of the Czech Republic (if there are less exports than it would be expected) and the export competence (if the Czech Republic has a relative comparative advantage measured through the Laffay index and if the partner has a disadvantage). With those variables considered, the markets chosen are those which show favourable values regarding all those three aspects. Finally, this list is evaluated and revaluated meticulously (it is assumed that further variables are taken into account) thanks to the extensive net of bureaus that the Czech Republic has abroad and thanks to the cooperation of business associations. It is explicitly targeted towards sectors with low current export penetration, and its reliance on external corroboration is not feasible to be applied by firms or other organizations without such vast resources.

The most extensively used framework for estimation at the country level, is a variation of an already proposed model; a screening process (Jeannet and Hennessey, 2005). The screening process at the country level is described in numerous papers such as Cavusgil, (1985), Cuyvres et al. (1995), Cuyvres et al. (2009), Urban et al. (2014a) (2014b) and Peterson, (1990), already above partially explained using the figure 5 as illustration. Habitually, their application is tailor made according to the needs of the interested organizations which apply them. Nevertheless, as can be gathered from those mentioned papers, the process usually begins by assessing each market from the broadest perspective first (using macro level variables comparing different parameters which categorize those markets with predetermined thresholds) and then in subsequent filters (those filters can be understood as "barriers" in which some markets pass, and others remain) other more specific features are taken into account (such as micro founded variables) for the evaluation of each market. After that, some markets are eliminated while afterwards, more detailed measures are taken into account. It is worth mentioning that as it is tailor made; each entity usually defines what it is understood as market for their analysis. The most widely accepted notion and the one which will be used in this thesis can be stated as follows; a market is understood as a country-product combination.

In figure 5; we observe the prototype of the screening process at the country level, first presented by Richard Wayne Walvoord in 1980, in his article "Export Market Research" in the journal *Global Trade Magazine*. The cut off variables are usually determined by the management in order to fit their particular needs (Jeannet and Hennessey, 2005) and a particular product under scrutiny. Cuyvres et al., (1995, 2009) decided to use as the main cut off criterion (determination of a threshold for each parameter), the point where a large number of markets fall apart if a given parameter increases further.

Although still no rule exists regarding which variables to use (inherent the possibility that important variables are not included). Neither explicit corroborated guidance exists regarding the explicit form of the depicted sequential process; this framework is the one which will be used in this thesis as Urban et al (2014a; 2014b) did (during the process, the algorithm will eliminate markets, not countries). The main reason is that, as the objective of this work is to find export opportunities for the Czech Republic at the market level (country product combination), this framework is the only one capable of handling such amount of information, in a cohesive way and produce a legible and coherent result. For obvious reasons, to make a global ranking of all the world markets or assess them in a matrix is not practical (as a remainder, there are approximately 1000000 country product combinations). Another argument in favour for this schema is that as we found in Wood and Robertson (2000), export managers in their export decision usually use an "elimination by aspect" heuristic. In other words, they do not tend to compensate markets by weaknesses in some parts with strengths, they rather tend to establish a set of minimum thresholds with each considered dimension under scrutiny and they tend to eliminate those which do not surpass those minimums. The filtering method partially resembles this step by step elimination process.

It is necessary to underline and remember that, model and parameter uncertainty is constant through all proposed models. Given that, some attempts exist in the literature to rationalize the variable selection procedure, as we encounter in Gould (2002, p 48). Where are here explained how Russow in 1989 with data from 1970 and 1980 given 6 randomly selected commodities, obtained four constructs as the most important ones for the selection of worth foreign markets through statistical analysis (no explicit confirmation exists): the overall market size (measured through GDP [Gross Domestic Product] or population), the economic development (measured, for example, through GDP per capita), the product market size growth (growth of imports) and the product trade (exports and imports of the product). Another study which attempted to rationalize the selection procedure is the one published by Wood and Robertson (2000) in which thanks to detailed interviews to managers of exporting corporations, a set of important dimensions and a ranking of them were established under consensus. The most important dimension under scrutiny according to this study is firstly the market potential (if there is sufficient demand for the product), secondly in order of more to less important the dimensions of legal aspects were generally found (tariffs, non-tariffs measures and regulations), then politics, infrastructure, economics and, finally, cultural aspects of the territory.

Every effort has been made to select the most appropriate variables[9] for the constructed model, described in subsequent sections. Papadopoulos et al (2002, p. 171), use the following set of criteria for variable selection ("relevance, frequency of use in past research, evidence of satisfaction, data availability, reliability and comparability"). Those principles are as well applicable in that circumstance, being highly relevant.

[9] In other research fields such as econometrics or machine learning, the variable selection procedure is usually carried out observing a fitting criterion such as the percent of correct answers given by the model, the R squared, the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).

As a remark, it is considered convenient to clarify some concepts. The ulterior aim of those reviewed models is to identify entrepreneurial opportunities whether at firm or at country level, “those situations in which new goods, services, raw materials, and organization methods can be introduced and sold at greater than their costs of production” (Shane and Venkataram, 2000, p.220); within this whole spectrum, explicitly the aim of those models is to identify new international markets in which it could be possible to enter, export opportunities[10]. Furthermore, “an asymmetry of beliefs and information is a precondition for the existence of entrepreneurial opportunities” (Shane and Venkataraman, 2000, p.221) because otherwise, they would attract other economic agents; consequently the economic profit and their existence could disappear (in general, nevertheless it is acknowledged that some corporations are immune to the effective competition in some markets as they are in possession of market power). Therefore, the vivid recommendation is that the results outputted by the finally constructed EDSM of this thesis (explained in detail in subsequent sections), would be extensively reviewed, as depending on the nature of the organization, some markets could be in fact presented as export opportunities while others not. What’s more, at the moment there is not consensus in the research community regarding if entrepreneurial opportunities are entities waiting to be discovered or if they are created through the actions of the entrepreneurs[11] (Suddaby, et al., 2015., p 3) (in press), in fact neither there is a consensus regarding what entrepreneurship is (Chandra et al, 2009, p.37). The chosen view in this thesis is that export opportunities are objective, therefore waiting to be discovered. Nevertheless, a determined sequence of actions for their final successful exploitation tends to be necessary.

---

[10] The article written by Chandra et al. (2009, p.37), explicitly argue that export opportunities fall within the domain of entrepreneurial opportunities, differencing them from other similar actions.

[11] Therefore, it is understandable that no scientific law at the moment exist for the development of an EDSM.

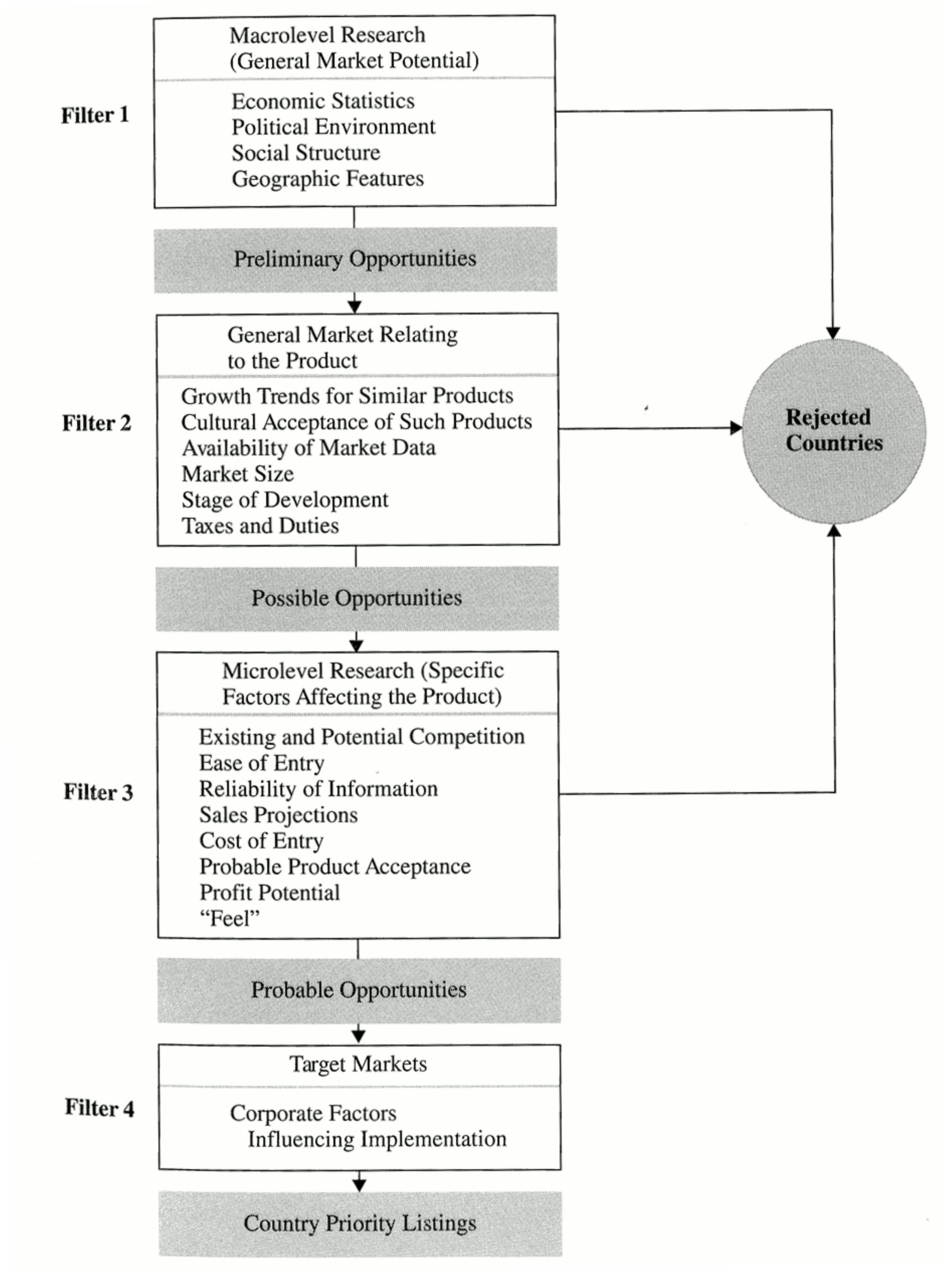

Figure 5: Market Screening

Source: Onkvisit, and Shaw (2004 p 145)

3 Methodology

3.1 Variable Selection

Taking into consideration the previously depicted principles, facts and arguments, this chapter provides an argumentation regarding the variables used in the filtering process, prior their implementation. More in particular, the principles formulated by Papadopoulos et al (2002, p. 171) seem to be appropriate ("relevance, frequency of use in past research, evidence of satisfaction, data availability, reliability and comparability"). Consequently, the model will be constructed on the basis of the previous work done by Cuyvres et al. (1995), Cuyvres et al. (2009), Urban et al. (2014a) and Urban et al. (2014b) with some additions as no scientific law is yet established guiding the model construction. Stressing that, each model has a subjective component, being up to the research to decide its final composition.

The encountered common feature is that the macro perspective is analysed first and then the micro perspective is carried out. So, except for the EDSM applied to the Czech Republic, the economic size of the countries in which a given market is situated tends to be taken into account. In this updated implementation of an EDSM targeted to the Czech Republic, it is chosen to continue with the approach developed by the previous version. Since, no clear benefit seems to exist in including parameters such as GDP given the risk that poor countries may be erroneously eliminated which in fact could contain markets potentially determined as export opportunities and, the economic size of the markets is already explicitly measured at the product level as will be made clear in the next sections.

Another popular phase during the filtering process is to discard countries based on their perceived country risk; the common approach is that the user of the model selects the level of the country risk that he wishes to tolerate at the end of the filtering process; so another filter is performed, this time the threshold being determined by the user. Given that, there are various sources regarding the possibility of using the OECD (Organization for Economic Development and Cooperation) country risk indicator or the ONDD (*Office National Du Ducroire*) country risk databases as the EDSMs constructed by Ludo Cuyvers do; in this thesis, no country index will be included. This thesis concentrates on the construction of an EDSM targeted to the selection of a subsample of promissory markets, with a strong recommendation to assess afterwards from all considered perspectives (such as the country risk or other risk index) that achieved subsample of export opportunities. Each organization has specific competitive advantages; therefore its internationalization strategy should be determined by its intrinsic nature in order to achieve the highest possible benefits (Brouthers et al., 2009).

In the next step, the micro perspective is carried out (see appendix C for the whole illustration of the filtration process finally implemented in this thesis). In the highlighted literature, country-product combinations (markets) are analysed via three indicators: short-term growth of imports, long-term (medium-term) growth of imports, and relative size of the market proxy by the amount of imports in monetary terms, differentiating them according to the RCA[12] (Revealed Comparative Advantage, indicating those products for which relative efficiency of production is higher, therefore a lower value can be tolerated achieving similar results) for a better calculation of the critical thresholds. Those measures target the potential of the market under scrutiny, indicating if there exists demand for a product. The economic logic behind is that, the more positive those measures are, then the higher the probability that those markets are export opportunities, *ceteris paribus*.

The papers reviewed differ in the data sources; the most appropriate is the HS[13] 6 (Harmonized System 6 digit level of desegregation) (there also exists HS 8 or even HS 12 but the problem with these even more disaggregated sources is that they

---

[12] Of the country of interest (for which is constructed the EDSM) in them and product under scrutiny, in that case Czech Republic.

[13] The source of data will be the COMTRADE database, which is dependent on the provision of information from countries all around the world, see section 4 (data) for further details.

Nowadays there exists an actualized version of the HS from 2007 but given that a big amount of countries report their statistics using the old version from 1992, it will be assumed that the common version is the one published in 1992.

are not homogenized across countries). One reason is that the HS classification is widely used in export ventures by exporters (Cuyvres and Viviers, 2012). Other sources which were previously used such as the SITC (Standard International Trade Classification) are rarely used by exporters, so the results obtained through the HS 6 become more meaningful and practical (even though, conversion tables exist, it is preferable not to carry out any conversion because the risk of error increases). The problem with the use of HS 6 is that given that their primordial purpose is for tariff calculation, some products are overrepresented and others underrepresented with respect to the actual world trade patterns (Bachetta et al; 2014). In other words, some categories are such detailed disaggregated being discernible products for which there does not exist a huge volume of trade, being other categories less detailed disaggregated than it would be convenient.

The addition made by Urban et al. (2014a), as its model is tailor made for the Czech Republic, scaling the RCA with the FVA (Foreign Value Added) is necessary and it will be included in this version, reflecting the particular Czech economic reality properly. As this index, has been criticized for not showing the comparative advantage properly (Leromain and Orefice, 2014). The main reason is that as the Czech economy is fully integrated into the European Union and in the global economy, through this country many products transit while the major part of the embedded value added being produced elsewhere. To corroborate it, in figure 6 we observe different countries and the proportion of foreign value added in their exports. In 2009 and 2011, the Czech Republic had one of the highest participations overall; this situation is expected to endure.

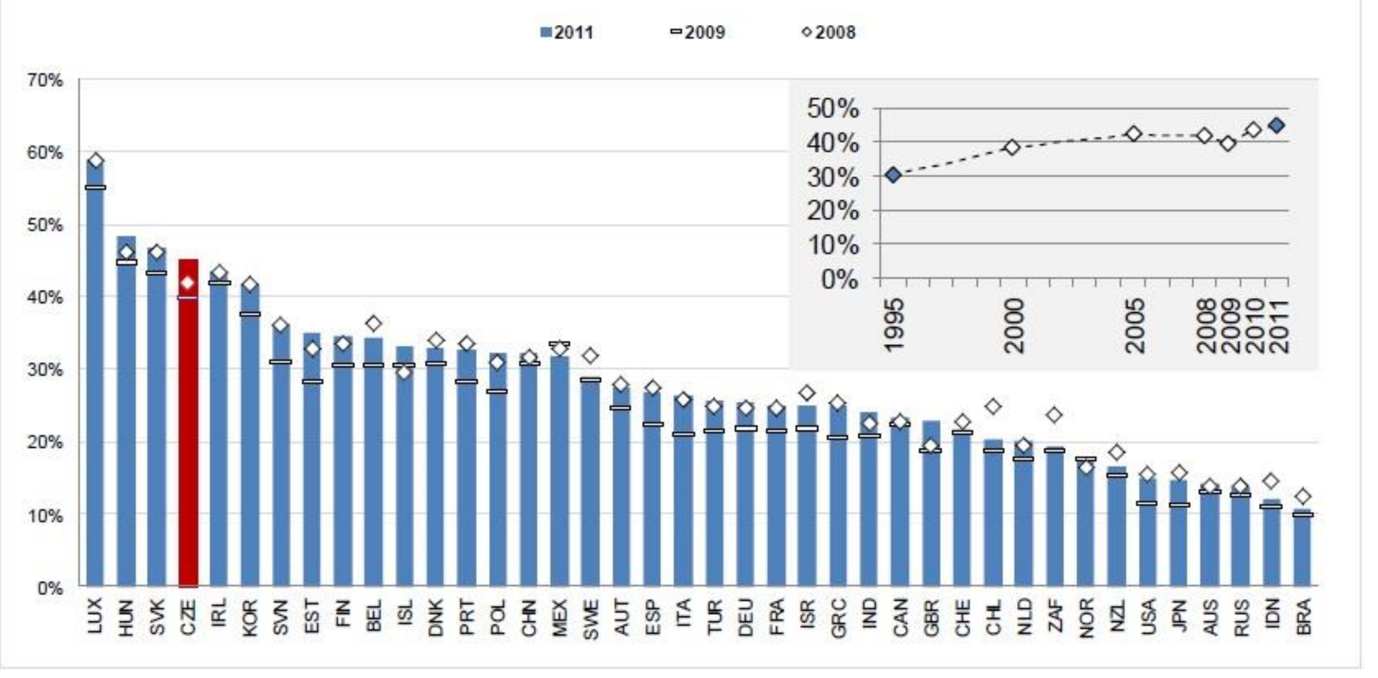

Figure 6: Foreign Value Added

Source: OECD (2015)

In the next leg of the elimination process, trade barriers and the market concentration are evaluated for each market. The common approach is to take into account the previous filters that provided a market classification (figure 7). Graphically shown are combinations which were eliminated and which remained, labelled as 0 or 1 if they surpassed the determined thresholds. This approach is considered to be correct because relative "flaws" of a market in one area can be compensated by relative "strengths" in other parts (that methodology partially contradicts the procedure previously mentioned, described and suggested by Wood and Robertson (2000) consisting of, a pure elimination but the author considers it appropriate as it permits a better identification of worth foreign markets because the concrete parameters are not only important issues but also the interactions among them and this is one way to represent them). The inclusion of this classification into the model as a basis for the elimination process is relevant as it allows for more flexibility, balancing relative weaknesses with strengths encountered in each market.

Consequently, there is a reduced risk of mistakenly elimination of worth foreign markets.

Table 4: Categorising of country-product combinations

Category	Short-term market growth	Long-term market growth	Relative market size
0	0	0	0
1	1	0	0
2	0	1	0
3	0	0	1
4	1	1	0
5	1	0	1
6	0	1	1
7	1	1	1

Source of data: excerpt from the DSM for South Africa.

Figure 7: **Categorizing country-product combinations South Africa**

Source: Cuyvers (2009)

At this stage, an analysis of the barriers to trade is crucial for the determination of the export opportunities as current evidence on international trade suggests. Those barriers are strong enough since only relative productive firms can overcome them. More specifically, “an estimate of the ‘representative’ international trade costs for industrialized countries reported in their ad valorem tax equivalent is 170%, [21% transportation costs, 44% border related trade barriers and 55% retail and wholesale distribution costs; 2.7=1.22*1.44*1.55]” (Anderson and Wincoop, 2004, p 3).

In all papers, the market concentration is first measured using the HHI[14] (Herfindahl Index). Different industries have embedded different structures so this measure in many cases provides an over-simplifying picture, further details would be needed to deduct precisely the “effective competition” of a market. But it is understandable that, this procedure being highly recommended, in our case, the EDSM would become intractable, given that, in the vast majority of industries this parameter provides relevant information regarding the level of “effective competition” encountered. A generally accepted definition of “effective competition” is the absence of market power, in such a way any company can increase and sustain prices profitably above the competitive level for a long period of time. Consequently, their absence suggests that there would be barriers to entry or expansion in that market among other implicit costs.

The major discrepancy found in the literature is concerned with the explicit measurement of trade barriers. The majority of the existing papers apply the methodology of “revealed absences of trade barrier”[15]. Basically, the philosophy of that implementation is that if the neighbouring countries with similar economic structures as the one of interest were successful entering a given market, then that should be the case for “our” country and the barriers and costs for the involved firms should be similar for both of them.

The parameter for each country product combination is provided by Urban et al (2014a), taking Germany as reference, as it is the neighbour nation with the highest ESI (Export Similarity Index) (table 2):

[14] As a remainder, the highest value for that parameter means, that the market is monopolistic. It would signify that even being attractive, there exist high barriers which detract the entrance of new competitors.

[15] One reason for this implementation is generally the lack of appropriate data.

$$MS\,(i, j) = \frac{\frac{Xg,i,j}{Xg,j}}{\frac{Xw,i,j}{Xw,j}} \qquad (2)$$

German Market Share

Where:

Xg,i,j : German export of product j to country i

Xg,j : German export of product j

Xw,i,j : World export of the product j to country i

Xw,j : World export of product j

Being the ESI an indicator first proposed by Finger and Kreinin (1979), aimed to measure the similarity among trade patterns for different countries, the following formula is formulated, taking values between 1 and 0:

$$\text{ESI (ab, c)} = \{ \textstyle\sum_{i=1} min\ [Xi(ac),\ Xi(bc)] \} \qquad (3)$$

Export Similarity Index

Where:

$Xi(ac)$: Share of commodity I of exports of a to c, usually being c the world

$Xi(bc)$: Share of commodity I of exports of b to c, usually being c the world

Actually, further in the literature some alternatives which measure trade barriers can be found. Thus, other studies construct explicitly an index for a better identification at the market level. The index implemented by Cuyvers et al (2009), possibly evaluates those barriers appropriately as it includes tariffs and non-tariff measures so it is possible to discern the level of protectionism among different products inside countries (being a widely held view that different industries, even inside countries, present different levels of protectionism) and includes a refined distance measure (not just air separation but sea distance taking into account that, "maritime transport handles over 80 per cent of the volume of global trade and accounts for over 70 per cent of its value[16]") among other parameters considered to be useful. But its actual form is rather useless as the aggregation among the variables in the index is not yet resolved. On the other hand, more recent studies based on Cuyvres et al (2009) with South Africa as the interest country, have developed alternative indexes. One of them is proposed by Steemkamp (2011) in his PhD thesis where a comprehensive amount of different indicators are used at the product rather than at market level. The indicators used to assess the product trade barriers are the following: the international shipping time per country, domestic time to import per country, international shipping cost per country, domestic cost to import per country, the LPI (Logistic Performance Index, explained afterwards), ad valorem equivalent tariffs per product and ad valorem equivalent non-tariff barriers per product. Finally, the index is constructed for each market extracting three principal factors and summing them (Steemkamp, 2011, p 73). This study correctly explores the barriers at the product level, at the same time the variables included target the factors behind costs of trade as described by Anderson and Wincoop (2004).

Taking into consideration all the previous statements, acknowledging its importance, this EDSM will also take into account trade barriers not only through the approximation provided by the German relative market share (filter revealed absences of trade barrier) as researchers in previous studies did. Within the previous summary of the current state of the art, Steemkamp provided the most promising

[16] Source: *World Economic Situation and Prospects* (2012)

measure, even though no absolute certainty that his aggregation procedure is correct[17] is provided and during the process of its construction, some information is always lost. As a result[18] of the given considerations, the chosen approach is to include additional variables into the model assessing each of them, different aspects concerning barriers to trade.

At first, it is preferable to include in the filter revealed absences of trade barrier, not only of Germany but construct this measure adding another country as previous models did, such as Cuyvres and Viviers (2012) where the country of interest was Belgium or Thailand. Certainly, Germany has the highest ESI (0, 56 in 2010 and 0, 513 in 2014, Table 1) with respect to the Czech Republic, is a neighbour country and is culturally similar. However, in the Global Value Chain (cross border production network) it position tends to be more upstream (more distant to the final demand). As we encounter in ECB (2013, p 15), "CEE countries are usually located further downstream in global value chains" than their euro area partners. While, Germany being similarly displayed through the ESI indicator, it is in another development stage, characterized by a higher per capita capital allotment, overall higher salaries and higher gross national income per capita and higher gross national product per capita. In addition, its main industries tend to be national, a different scenario than in the V4 countries; and it is a much bigger economy than the Czech Republic. Therefore, it is needed to include another country.

The additional country chosen is Hungary which also has overall a great ESI (0, 49 in 2010 and 0,493 in 2014; third just after Poland and Germany as it can be observed in Table 1), has an even greater share of foreign value added (figure 6 and figure 8) still being that measure highly similar and it is of similar size to the Czech Republic, being as well a small open economy with a similar historical background (their economy suffered as well a transformation when the communist regime collapsed). In addition, as can be observed in figure 8 extracted from Grodzicki

[17] For further information regarding aggregation methodologies, it is recommended to review the excellent handbook on constructing composite indicators written by Nardo, et al. (2005).

[18] Taking into consideration the explicit nature of the EDSM which will be discussed in subsequent parts.

(2014, p 24), the proportion of the sectors contributing to the total GVC derived income, is similar and displays a similar development pattern for both countries.

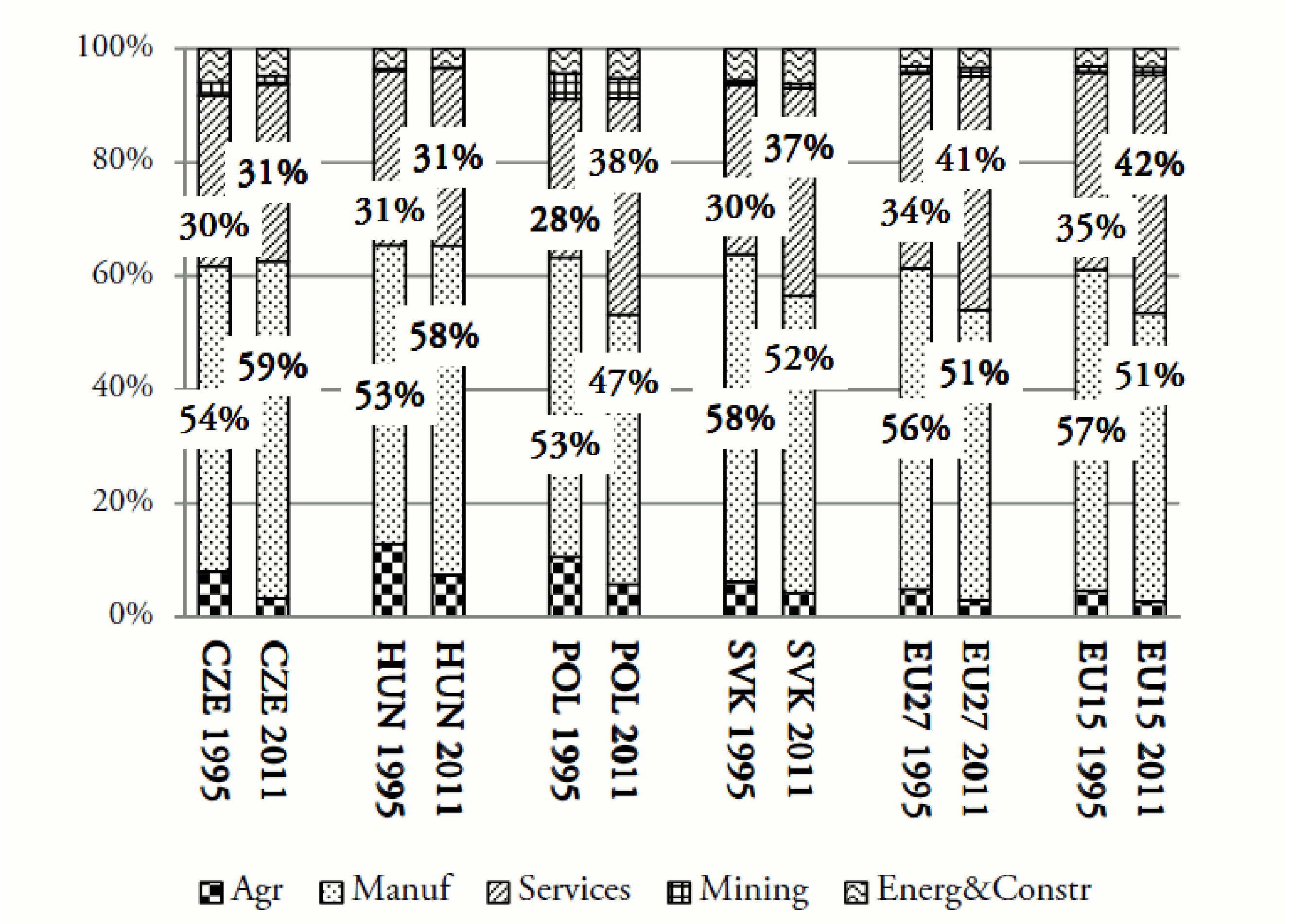

Figure 8: Sectorial contribution in the income derived from GVC

Source: Grodzicki (2014, p 24)

Country	Export Similarity
Czech Republic	1,000
Germany	0,513
Poland	0,508
Hungary	0,493
Slovakia	0,485
Slovenia	0,449
Austria	0,448
Mexico	0,430
Spain	0,424
Romania	0,423
France	0,408
USA	0,403
Sweden	0,400
Italia	0,392
Canada	0,386
Portugal	0,382
Russia	0,377
Netherland	0,362
Belgium	0,360

Table 1: ESI with respect to the Czech Republic in 2014

Source: Author based on COMTRADE database 2014

In addition, as observed through the observatory of economic complexity[19], although Czech Republic being more involved in the international trade, for both countries it is observable a similar pattern in terms of products exported and markets of destination (at least for those with major importance). And, both countries are not yet in the Eurozone, but in the European Union.

The previously provided arguments tend to be sufficient to catalogue a country as structurally similar. Nevertheless, an empirical method is next provided in order to gauge similarities among countries. More in particular, an unsupervised

[19] http://atlas.media.mit.edu/en/profile/country/cze/ [10/07/2016]

http://atlas.media.mit.edu/en/profile/country/hun/ [10/07/2016] learning algorithm, a hierarchical clustering is developed. That data analysis method is applied to numerous fields from astrology to medicine to economics. There are two basic approaches; or agglomerative where given a dataset, the elements provided based on similarity are grouped in clusters successively until all the elements form part of the same cluster ("at the bottom each cluster contain a single observation, at the highest there is only one cluster" (Friedman, Hastie and Tibshirani, 2001, p 520).), graphically being represented by a dendogram. Or divisive, where given an entire dataset, a sequential partitioning based on dissimilarity is performed.

The application of this method was inspired by the article published by the CNB (2015, p 24). In that case, it is observable the application of this methodology for 2007 (the one which is shown and which will be implemented, is an agglomerative version). Although the economic structure tends not to change in the short term, that method was decided to be applied once again, with more recent data. As during that period, Europe suffered the world financial crisis and the Euro crisis and several structural reforms were implemented by various governments (it is worth mentioning that in order to discern the effects of the reforms, short time periods do not tend to be appropriate). As it can be easily observed, the most similar country in socio-economic structural terms of the Czech Republic is Slovakia, just being followed by Hungary.

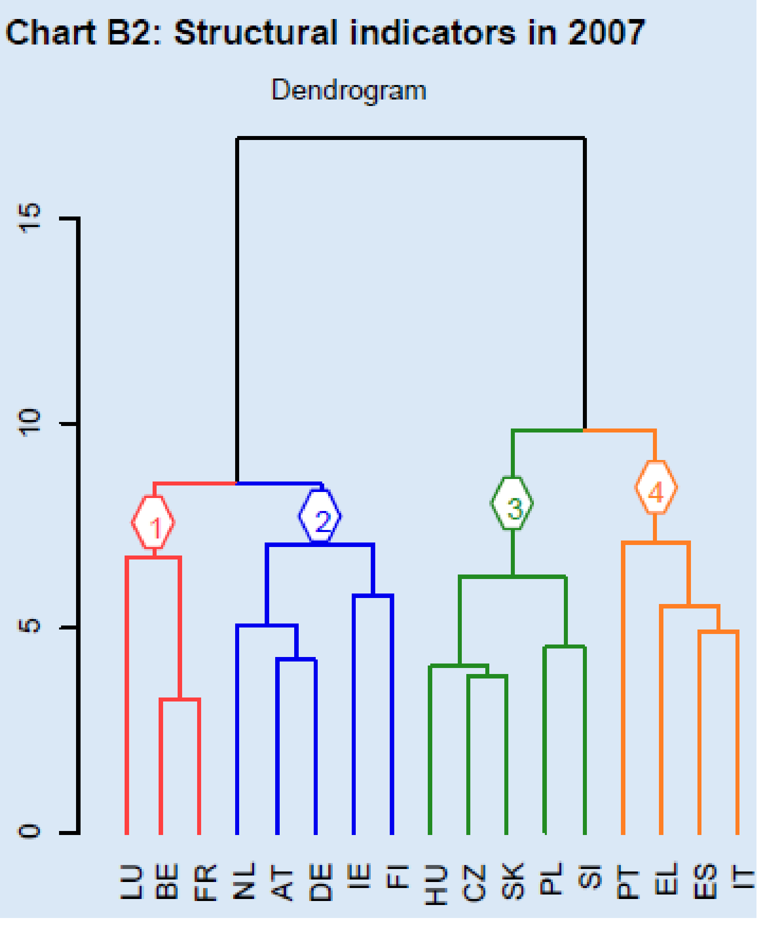

Figure 9: Cluster dendogram for 2007

Source: CNB (2015, p 24).

A crucial step is a correct selection of relevant attributes for each observation (in this case countries) in order to be able to proceed. Unfortunately, as it is not a supervised learning algorithm, there is no function which optimize so consequently no explicit rule to choose appropriate variables is at disposal. The chosen approach in this implementation is to use the same variables which implemented in his study Irac and Lopez (2015), considering their usefulness and applicability; being the same as the used by CNB (2015, p 24). As already clarified; this time using the most recent data sources as possible. Some studies have appeared in the literature in order to select appropriate features; their approach is to reduce the initially provided dataset while saving the most of the information. The reason of the reduction is that if some variables are excessively correlated then the information provided can be overrepresented so cluster are not formed correctly, a general cut off criterion is a correlation of 0, 9 or superior (Sarstedt and Mooi, 2014), see Appendix A for the variables used.

In order to implement that algorithm, a metric has to be provided and carefully chosen in order to be able to compare in terms of dissimilarity the observations of the dataset. There are numerous possibilities available, being up to the researcher to choose the one desired. The majority of the distance measures are derived from the generalization provided by the Minkowski distance (4). As, when p=2 then it is the Euclidean distance (5), when $p \to \infty$ it is the Maximum distance (6) and when p=1 then it is Manhattan distance (7). Within them, the Euclidean distance (5) remains as the most popular

$$D_k(X_i, X_j) = \left(\sum_{n=1}^{d} |Xin - Xjn|^k\right)^{1/k} \qquad (4)$$

Minkowski Distance

$$D_2(X_i, X_j) = \left(\sum_{n=1}^{d} (Xin - Xjn)^2\right)^{1/2} \qquad (5)$$

Euclidean Distance

$$D\infty(X_i, X_j) = \lim_{k \to \infty} \left(\sum_{n=1}^{d} |Xin - Xjn|^k\right)^{1/k} \qquad (6)$$

Maximum Distance

$$D_1(X_i, X_j) = \sum_{n=1}^{d} |Xin - Xjn|^1 \qquad (7)$$

Manhattan Distance

In this case, no guidance exists for the election of the appropriate distance metric. Nevertheless, the most widely used and preferred among researchers is the Euclidean distance and it is the one which will be implemented in that updated version of the hierarchical clustering.

A basic step generally recommended will be implemented; specifically the dataset will be standardized in order to limit the possible influence of outliers and achieve that each variable would have approximately the same weight. A basic recommendation for datasets where different variables have different units, as it is this case.

The formula used for standardization is:

$$\frac{\mathrm{xi} - \mathrm{mean(xi)}}{\mathrm{sd(xi)}} \qquad (8)$$

Standardization

Given the specific metric, a basic step remains before the implementation of the algorithm. Concretely, it must be specified how the distance among the clusters is computed in order to determine its dissimilarity. Four basic approaches are widely used (9, 10, 11 and 12), being their formulas extracted from the article written by Müllner (2013, p 4): single, complete, average and Ward method. In the notation provided, as an example the dissimilarity of cluster A with respect to cluster B is computed (as a remainder, in this case the Euclidean distance or D_2 metric is used). Being |A|, the cardinality of A or the number of elements contained in A; and ca being the centroid of a, or the element "in the middle of the cluster A" computed by the Euclidean distance. Equal notation is being used for set B

$$\min_{xi \in A\, xj \in B} D_2\,(xi, xj) \qquad (9)$$

Single Linkage

$$\max_{xi \in A\, xj \in B} D_2\,(xi, xj) \qquad (10)$$

Complete Linkage

$$\frac{1}{|A||B|} \sum_{xi \in A} \sum_{xj \in B} D2\,(xi, xj) \qquad (11)$$

Average Linkage

$$\sqrt{\frac{2|A||B|}{|A|+|B|}}\; D2\,(ca, cb) \qquad (12)$$

Ward Method

The single linkage computes similarity among clusters as the similarity among the closest pairs of the designed clusters, the complete linkage computes it through the farthest pair distance, the average linkage computes through the average and the Ward method at each step minimize the within cluster variance (Irac and Lopez , 2015, p. 6). The within cluster variance is understood as a measure of its heterogeneity, being determined by equation (13), assuming Euclidean metric (Friedman, Hastie and Tibshirani, 2001, p 509).

$$\sum_{xi \in B} D2\,(xi, cb)^2 \qquad (13)$$

Within cluster variance

The approach chosen is to compute the dissimilarities of the clusters with the complete linkage as the effect of possible outliers is reduced. Even though, no explicit rule exists; being equally in this case up to the researcher to choose the desired linkage, given a particular dataset. Finally, the result is the provided in figure 10. As a robustness assessment, it was computed this dendogram as well using the ward, average and single linkage, they can be observed in appendix C. The common pattern is that Czech Republic is encountered in the same cluster as Slovakia, Hungary, Spain and Italy. Being Italy and Spain much bigger economies and developed economies therefore slightly more different.

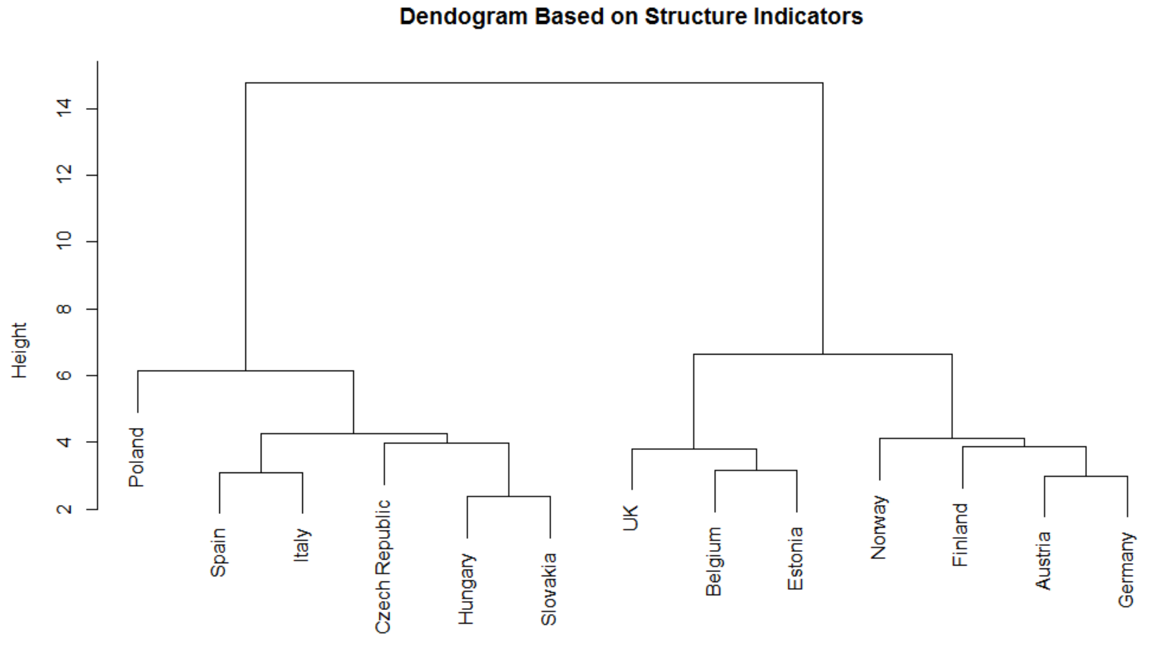

Figure 10: Cluster dendogram

Source: Author based on Müllner (2013)

As a remark, the observed hierarchical clustering is an example of the implementation of a market segmentation method (as well), considering markets as countries. Different companies for their implementations rather than use the variables provided, could use others more specific to their characteristics and apply it over a larger dataset (with more countries or regions).

As observed, consistently Hungary displays a tight structural similarity with respect to the Czech Republic (consistently much bigger than Germany). This model, as will be clear in subsequent sections will be supplied with data for 2010 and for 2014; given the provided analysis, at the moment there are not impediments in using exactly the same schemata for both time intervals.

With the introduction of Hungary, this indicator will better fulfil its purpose as described in Cuyvres and Viviers (2012, p.89), mapping the export penetration of similar[20] countries to the Czech Republic in foreign countries, inferring that if a market shows a high proportion of penetration by those countries, then no substantial barriers for Czech producers ought to exist. A possibility remains to include more countries similar to the Czech Republic such as Slovakia.

Germany will remain although being higher in the GVC not only thanks to their high ESI and thanks for being in the EU, but also thanks for being a major global commercial power. As such, it exports to a broad range of markets in which nowadays Czech Republic is not present. Therefore, those markets could be integrated into the model and evaluated as export opportunities or not.

[20] Cuyvres and Viviers (2012), use explicitly neighbours for their assessment. The understatement of that approach in that Thesis is that the most similar countries tend to be the neighbours.

As it is acknowledged, nowadays countries tend not to export but rather firms inside those countries. The approach of that thesis is that if two countries are similar, then firms inside those countries as well will tend to be similar as are influenced by a similar environment.

As a result, the new indicator used will take the form of:

$$MS\,(i,j) = \frac{\dfrac{Xg,i,j}{Xg,j} + \dfrac{Xh,i,j}{Xh,j}}{\dfrac{Xw,i,j}{Xw,j}} \qquad (14)$$

German and Hungarian Market Share

Where:

Xg,i,j : Germany export of product j to country i

Xg,j : Germany export of product j

Xg,i,j : Hungary export of product j to country i

Xg,j : Hungary export of product j

Xw,i,j : World export of the product j to country i

Xw,j : World export of product j

Another addition to the base model which will allow assessing better possible obstacles encountered is the LPI index (Arvis et al., 2014). As already mentioned, and as formulated by Anderson and Wincoop (2004), trade costs (the paper state this figures for developed countries; analogous parameters and proportions are believed be present in case of developing countries) reported in their ad valorem tax equivalent, are believed to be composed of: 21% transportation costs, 44% border related trade barriers and 55% retail and wholesale distribution costs.

The LPI index (Arvis et al., 2014), is an indicator conceived at the country level and assessing six different features: the efficiency of customs and border clearance, the quality of trade and transport infrastructure, the ease of arranging competitively priced shipments, the competence and quality of logistics services (trucking, forwarding, and customs brokerage), the ability to track and trace consignments and the frequency with which shipments reach consignees within scheduled or expected delivery times. As can be observed, those features relate tightly to the observable factors determining trade costs in addition to its purpose evaluating the quality of the infrastructure of a given territory (a fundamental factor as *per se* evaluating possible trade costs encountered). Therefore, it can be concluded that this measure is relevant and reliable as was developed by the World Bank. In the next table, we observe the disaggregated LPI score for the Czech Republic in 2014 by their respective components and the worldwide position that Czech Republic holds in it (LPI rank).

Country	Czech Republic
Year	2014
LPI Rank	32
LPI Score	3.49
Customs	3.24
Infrastructure	3.29
International shipments	3.59
Logistics competence	3.51
Tracking & tracing	3.56
Timeliness	3.73

Table 2: LPI of the Czech Republic in 2014

Source: Author based on Arvis et al. (2014)

The air distance used in previous implementations, which serves as a proxy for transport costs, will remain in the new model, being their use backed by their successful implementation and observed validity. It was argued that a better measure would be taken by the sea distance but given that the Czech Republic is a landlocked country, that preponderance in this case may not be so clear.

The major weakness of all those cited measures which will be incorporated is that they consider only the country level, not the market level; although a huge variance in the degree of protectionism inside countries for different products was observed. So a measure which would allow distinguishing the barriers at the product level seems to be necessary. In spite of being very rarely used by those models, the chosen approach is to use tariff and non-tariff barriers at the product level as an additional filter. On the one hand, it is true that barriers to trade have fallen in the last decades thanks to the numerous negotiations at multilateral and bilateral agreements, not being substantial for developed countries. However, their effects are still present in international trade and companies do take them into account as displayed by Wood and Robertson (2000) who conducted a survey of different companies by evaluating different indicators in order of importance. The legal measures (tariff, non-tariffs, and others such as visas patents or regulations); ranked consistently second in importance, just after indicators assessing the potential market demand. The study of Blonigen et al (2013) is another evidence since we can clearly observe in it that non-tariff barriers do induce market power for local producing firms, indicating that the higher they are, the greater the market power created. Finally, ongoing efforts for a further liberalization of world trade through the elimination of barriers to trade do signal that they are not negligible.

After the previous stage in which the trade barriers were cohesively evaluated, previous researchers included further filters; assessing more different features such as the Laffay index (a detailed and argued to be a robust measure of comparative advantage). Remembering the principles depicted by Papadopoulos et al (2002, p. 171), it seems rational and necessary to include more indicators in order to weed-out markets for which there could not be found good prospects. The best possible available framework is the work done by Urban et al (2014a; 2014b) as it is explicitly tailor made for the Czech Republic, being quite recent and so including all the past research. Therefore, the approach chosen is to include the previously used variables, considering their usefulness and applicability.

More specifically, the Laffay index is an important addition as it measures the comparative advantage directly. Actually, the unanimous support for international free trade among economists is derived from the understanding of the theory of comparative advantage, first formulated by David Ricardo (1772–1823). Although that theory is based on unreal assumptions, such labour and capital do not move among nations, “Samuelson calls it the only proposition in social science that is both true and non-trivial” (Schumacher, 2013, p 83). This theory states that each country should engage in the international trade, specializing in which it is best at (most relatively efficient), achieving as a consequence the highest possible gains for the overall society.

Therefore, that indicator provides guidance on the products that the Czech Republic focus on during their export ventures and should continue focusing on.

$$\text{LFI (i, j)} = \left[\frac{Mijt - Xijt}{Mijt + Xijt} - \frac{\sum_{j=1}^{k}(Mijt - Xijt)}{\sum_{j=1}^{k}(Mijt + Xijt)} \right] * \frac{Mijt + Xijt}{\sum_{j=1}^{k}(Mijt + Xijt)} \qquad (15)$$

Laffay Index

Where:

$X_{i,j,t}$: Export of product j to country i at time t

$M_{i,j,t}$: Import of product j to country i at time t

The interpretation of this measure can be understood as. “According to the index, the comparative advantage of country i in the production of item j is thus measured by the deviation of product j, normalised trade balance from the overall normalised trade balance, multiplied by the share of trade (imports plus exports) of product j on total trade” (Zaghini, 2003, p 11). As a result, this indicator allows us to control the intra-industry trade; re-export flows (already explained a very important factor especially for the Czech Republic, see figure 6) and cyclical short term components (Zaghini, 2003). Their value in absence of comparative advantage is 0, while the more positive its value for a product, the more specialized a country is in that product (the comparative advantage is stronger suggesting that the country is relatively efficient at producing that commodity). The way it is included in the filtering process is to compare for each product the value the Czech Republic has with respect to their trade partner[21] (as they do not face major transportation costs, potentially *ceteris paribus* it is likely that the highest competitor are the domestically producing firms), so favouring those markets where the prospective partner has a relative comparative disadvantage and the Czech Republic has an advantage (the difference is the highest). In addition, that indicator will favour those products for which there is already present production capacity.

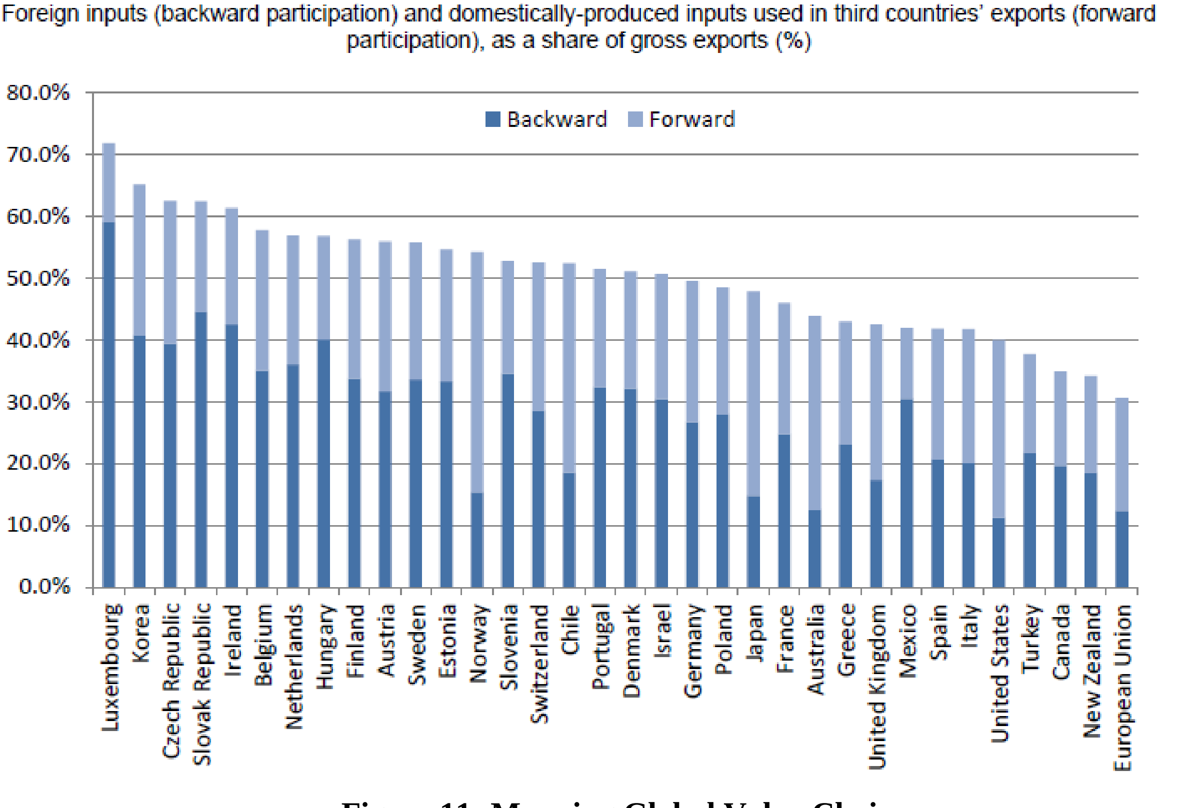

Figure 11: Mapping Global Value Chains

Source: Sebastien Miroudot (2012)

[21] LFI(CZ, product[i]) respect to LFI(country, product[i])

Progressing through the filtration process, two more indicators are assessed, on the one hand, the market importance index (16) (similar but not equal measure to the RCA) which takes that provided form (Urban et al, 2014a):

$$\text{MI (n, i, j)} = \frac{\frac{Xn,i,j}{Xw,i,j}}{\frac{Xn,j}{Xw,j}} \qquad (16)$$

Market Importance Index

Where:

Xn,i,j : Export of product j by the country n to country i

Xw,i,j : Total world export of product j to country i

Xn,j : Export of the product j by the country n

Xw,j : Total world export of product j

The reason for his inclusion is to favour those country-product combinations for which there exists a higher market share than it would be expected by the existing trade patterns, as it might signal that there exists some implied advantage although no explicit theoretical background is provided. Up to current evidence, its previous implementation was successful, so will remain in that model. The way to integrate it is by comparing its value with respect to the top 6 exporters to the given market.

Finally, the Prody index is included; its target is not to eliminate markets with bad prospects but to favour those markets which could potentially generate higher positive externalities for the Czech Republic (Bacchetta et al, 2014, p 27). The specific name of its indicator is the revealed technology content or Prody index and takes the form of:

$$\text{Prody}_k = \sum_i RCAik\ Yi \qquad (17)$$

Prody Index

Where Y is the GDP per capita, so it "describes the income level associated with a product" (Bacchetta et al, 2014, p 27).

Lastly as already discussed, in this version of the model it is encouraged to view the output as potential export opportunities. Therefore, the model took all world markets and selected a simple subsample. But, according to the needs, characteristics and prospects of the interested organization, this group should be further scrutinized to choose those markets with the highest probability of success and potentially most profitable.

The aim of this thesis is not only the construction of an EDSM, in addition it will be supplied with data for 2010 and for 2014, maintaining invariant the described model structure for both years. And, the hypothesis that for both time intervals the results will be equal is evaluated. Therefore, it will be determined if the model needs periodical updates or not.

H1: Retrieved country product combinations 2010 =

Retrieved country product combinations 2014

3.2 Threshold Selection

At this stage, the situation is similar as during the variable selection. As, for the variables considered, there does not exist at present a general consensus regarding which ones to use, how they should be weighted in order to reflect their relative importance and/or how to reflect their interaction effects (Papadopoulos, et al, 2002 p 167; Andersen and Strandskov, 1997; Gould, 2002).

Some researchers argue that the management board should be the ones who should set the appropriate weights according to their experience and knowledge. In other models, at the country level this approach is not feasible because it comprises excessive information, so some alternative procedures are carried out. In older models, such as in the portfolio techniques at the country level, the authors are the ones who set the corresponding weights. The EDSMs, as proposed by Cuyvers (1995, 2009, 2012), can be regarded as another example. In those models, some thresholds for individual filters are selected at the value where if surpassing, a considerable number of export opportunities fall apart.

In the light of this, the proposed methodology implemented by Urban et al (2014a, p 35) seems to be the most trustworthy, objective and relevant of all until now reviewed and, consequently it will be implemented in the current version (the implemented methodology will be a variation as no expert knowledge could have been provided; in the next sections, specific details are described). It consists of a first selection by the authors of 350 markets assessed explicitly as “verified” export opportunities through objective criteria, current successful destinations or explicitly reviewed by governmental organizations (Czech Trade and Czech Foreign Offices) as export opportunities. It is necessary to mention that those governmental organizations possess vast resources (which were used to select export opportunities, a scenario seldom achievable by any other organization). As we can see in Tlapa, Klepácek and Svoboda (2015)[22], a final categorization of export opportunities was done after an extensive research on the field and thanks to the collaboration of numerous business organizations. Given those selected markets, some parameters in various thresholds are adjusted so that the maximum amount of selected opportunities (those which are in the sample) is in fact assessed as potential export opportunities by the model. At other stages, a deterministic formula for the thresholds is deployed and depending on the nature of the data, some values which serve as thresholds are retrieved. Following the principles depicted by Papadopoulos et al (2002, p. 171), in this thesis it is chosen to continue with that methodology because of its relevance, until now evidence of success and its empirical base.

[22] Those authors did their research under the auspice of the Czech Ministry of Foreign Affairs.

The trustworthiness of that method is derived from their objectivity as it is based on empirical grounds. However, this method still implies some degree of subjectivity as a “selection” and some “criteria” for the selection are needed. For instance, the approach given by Tlapa, Klepácek and Svoboda (2015) in their identification of export opportunities is to select predominantly markets for which there is a low current amount of Czech exports to them, in contrast to what would be expected taking the current structure of the given market and the overall Czech exports structure.

The view chosen in this model is to continue with the successful approach of Urban et al (2014a). But, for the sample selection, the approach chosen is to not explicitly favour markets with abnormal low presence of Czech exports (as Tlapa, Klepácek and Svoboda (2015) did) or markets with high Czech exports. Rather, the critical view will be to select markets based on the relative comparative advantage, measured by the Laffay index of the Czech Republic and the sustainability and size of its growth (long term growth and short term growth being positive, while short term growth being higher than long term). It is necessary to highlight that in that case, there do not exist any scientific rule which would guide its selection process, therefore this approach is taken inspired by the article written by Tlapa, Klepácek and Svoboda (2015).

An important reason for the not selection of markets for the sample based on their current size and Czech exports to them is that this EDSM is intended to look at exports opportunities for all products worldwide and each product has an idiosyncratic potential volume (the export in monetary value can be equal to two products, but for one product with respect to its idiosyncratic characteristics it can be a low amount while for the other it can be a high amount). The reason for the disregard of the actual volume of exports, in addition to the fact that it is a factor dependent on cyclical elements is that the purpose of the model is to find export opportunities. Export opportunities are markets, defined as such that an exporter should be able to sell goods in them; given that, it is not necessary that the market has actually a low amount of Czech exports or a high one.

As a remainder, the sample is chosen in order to select the values of the thresholds in such a way that the output of the model during the successive filters, matches with the sample in the highest possible way. As described in Barreiro and Albandoz (2001, p.4), three different sampling frameworks exist aimed to achieve representativeness: probability sampling (the sample is chosen from a designed population aleatory), purposive sampling (the researcher tries to make it representative) and no-rule sampling. Within those broad frameworks, many different categories exist; their enumeration and description are not considered to be useful for that research. The technique chosen can be categorized as purposive sampling, as its core relies on a selection of a list; even though given that list the final sample is chosen aleatory. In statistics, some attempts have appeared in order to obtain theoretically appropriate sample sizes in order to achieve representativeness. Those studies can provide guidance, acknowledging that if the sample is very large it is not usually a problem (ideally, we would have already a complete list of opportunities so that the matching during the calibration process would be guaranteed to be flawless). In any case, the intention is that the sample achieves representativeness of the world export opportunities at the given time interval.

Krejcie's and Morgan's work (1970) is a classic paper, where a deterministic formula was derived (18) to determine the sample size:

$$n = \frac{X^2 N\, P\,(1-P)}{d^2(N-1) + X^2 P(1-P)} \qquad (18)$$

Sample Size Determination

Where:

X^2: Table value of chi-square for one degree of freedom at desired confidence level

N: Population size

d^2: Degree of accuracy expressed as proportion

P: Population proportion

In this case, the population size (N) is the list of all opportunities; ex ante it is unknown as is the purpose of the model to find it. As an approximation, the number of export opportunities retrieved by Urban et al (2014a) can be used, which were approximately 20000. Therefore, it can be amount approximately in that case to 25000. A 95% confidence interval was chosen as well as a 3, 5% margin error (d), in addition it is assumed as recommended by the author to use P = 0, 5 (that is an estimate of the heterogeneity of the population, in our case it is believed that that measure would be lower but in that way the maximum sample size is obtained which is not detrimental but rather beneficial). As a result, the recommended sample size and used sample size is 760[23].

Taking into account the previously depicted facts, it is considered that an estimation is being performed (first, a group of opportunities have been chosen explicitly believing in its reliability in order to gather appropriately all the remaining ones) because, no verification of the export opportunities is yet possible (the only possible way to verify all the opportunities empirically, would be to start an export venture to each of them and evaluate its success by a representative corporation). As indicated above, the primordial objective criterion for the selection of the markets in the samples is that they are growing and not decelerating (short term growth not lower than long term, being both positive). The logic behind, is that if the market is growing, that shows that at the moment, there is a sustained increase in demand inferring that there should persist that tendency at least in the near term consequently, or new firms will enter (for instance, realizing export opportunities) or current companies will increase its sales (always assuming effective competition). The selection will be performed in conjunction, with a positive value for the Laffay index, the larger the better (as already indicated, the Laffay index measures the competitive advantage). It reveals those products for which the Czech Republic is relatively specialized in. Therefore, those markets should be theoretically export opportunities. As explained in previous sections, ideally we would have many more measures in order to choose export opportunities. After the initial scrutiny based on both criteria

[23] That sample size, will be used in both implementations (data from 2010 and data from 2014). Further details regarding the datasets are encountered in section 3.

(those markets with negative short term or long term growth or negative LFI or higher long term growth than short term are eliminated), the finally chosen 760 markets are collected. Therefore, based on those which have the highest values (highest positive growth and comparative advantage, setting equal value for both variables) a ranking of the markets is performed (first, given a standardization of their values; formula 8 is used); given them, randomly are selected from the highest 1000 the finally used 760 (the implementation has been carried out through R Studio software, setting as seed, 760; equal procedure has been carried out for 2010 and for 2014). The initial intention was to include some markets into the sample from the list provided by Tlapa, Klepácek and Svoboda (2015), but this model works at HS6 and the list provided is at HS4. Further explicit selection was not implemented, in order to minimize as much as possible biases considering that at the moment, no expert opinion could have been provided.

Once the sample of markets has been obtained, for the gathering of the exact parameters in each filter, three criterions have been established. First, the maximum possible of in the sample markets had to remain. Second, the thresholds had to remain sequential such that it is possible to label the markets if they had higher or lower values (not allowing for the case that the threshold to label a market 2 was lower than to market it as 1). And third, if it was not possible to filter some markets without sacrificing in-the sample markets, then the reference values were those which had been established by Urban et al (2014a), as it is the most similar up to now constructed model being the objective for both of them the same, and using fairly the same datasets. Stressing that at the moment, it is not possible to determine the values of the different thresholds unambiguously. Finally, for the LPI indicator, its value was obtained tacking as inspiration the article written by Cuyvers (2009) setting the parameter at the point where if surpassing, a major number of opportunities fall apart. And, the market share value was obtained in a similar manner, applying the “rule of thumb” proposed by Cuyvers et al (2009), assessing if at least one country or together have a market importance of at least 0, 95 for any given evaluated market.

By the other hand, the model does not only eliminate markets based on thresholds. As such, in order to be able to process with the filtration process, for all markets there had to be enough data. Therefore, a great number of markets were eliminated because no data was available to them for a given indicator. For instance, all markets to which neither Hungary neither Germany export, were eliminated.

For next models, ideally the data calibration would be achieved by collecting a large amount of historical data by a highly heterogeneous portfolio of different firms from different sectors. Past successful historical export ventures initializing a trade contact, would be assessed and they would serve as the basis for the setting of the different parameters (they will form the sample). The restriction would be on the initialization by new firms into the market. Thus, a proper calibration procedure ("the basic idea of calibration is to choose parameter values on the basis of microeconomic evidence and then to compare to the model's predictions" (Romer, p 217)) would be performed. Although being the method just described even more objective than the previously reviewed ones as a purposive sampling is not carried out, concerns are raised regarding the validity of the sample. Especially, if the historical dataset comprises an extensive period, as the international economic environment is in constant transformation.

It is also important to point out that as was the case with papers written by Cuyvers (1995, 2009, 2012, 2016), this EDSM applied to the Czech Republic as already anticipated does not take a pure filtering method (complete elimination from the model of markets without appropriate parameters at a given stage). Instead, it allows as addition, for trade-offs among different filters. Therefore the filtration happens primarily when at each stage the value of a filter is 0 but in addition, some markets are eliminated when filter 1, 3, 4 and 5 sum less than 7. This process of explicit trade-offs among filters is necessary and will be included in the final model as the risk of eliminating prospective markets is reduced because it allows to evaluate them correctly and explicitly from different angles and allow to provide explicit interactions among various different parameters, see appendix C for the illustration of the whole filtration process.

The explicit trade-offs and interactions are achieved in various ways. As partially explained, the first filter based on the growth rates and sizes of the markets which labels markets from 0 to 2 (parameters previously obtained delimit each threshold inside each filter, as a result depending on its value a market can be labelled differently), serve as a base for the filter based on the HHI during the assessment of different markets (figure 5). In that filter, three are possible values, depending how the market was labelled previously; a higher or lower threshold is applied, finally markets are marked either as 0 or 1. For the filter based on distance, two thresholds are available consequently markets are labelled from 0 to 2 depending on the surpassed threshold. The LFI, Market importance and Prody filters applies a similar approach, three different thresholds being established, so markets are labelled from 0 to 3 (integer values).

4 Data

The model will use not only the same data as Urban et al. (2014a) and Urban et al. (2014b) did (for those common filters); in addition, it will be supplied with a more recent dataset, updated up to 2014. As such, it will be clear if there exists some sort of continuity in the outputted markets labelled as opportunities as for both implementations, the model would have the same structure. So, it would be possible to confirm or to discard the stated hypothesis (H1) that for both years, the outputted markets are equal.

In the light of this, an initial caution remark has to be stated regarding the provided hypothesis. In the previous version values were obtained for 2010, at that time the world economy was in a tumultuous situation. A generally accepted fact is that the Global Financial Crisis which had his peak in 2008 and 2009 has enduring consequences still today. During that period, the global trade collapsed a 12% YoY (Year on Year) basis principally as the paper written by Chor and Manova (2012) explains because of the consumer gloomy outlooks and a sharp reduction in available credit facility. While the world trade in 2010 experienced a sharp rebound (surged a 14, 5%), “the highest in record going back to 1950” (WTO, 2011, p. 20) but not achieving the previous trend. That rebound being much more pronounced in the developing economies than in the developed ones, China rose markedly. Correspondingly, this model takes into consideration parameters dependent on short term market developments. As such, some biases could have been added through those unprecedented events, rather than disclose the underlying structure of the markets of interest and their true market potential.

In addition, it is believed that the devaluation of the Czech Koruna the 7 of November 2013 has repercussions into the model results. The model focus basically on to the international environment, at the same time tacking into consideration the Czech economic structure which for the short run, is believed that is invariant. Nevertheless, the considerable gain in competitiveness achieved through that devaluation could affect determined parameters of the model, concretely the Market Importance (MI), the Laffay index (LFI) and the RCA. And, as explained, the world economy is in constant transformation.

Therefore, as the world economy with comparison to 2010 in 2014 was more stable with less economic uncertainties and with a more robust and liquid financial system (it is necessary to acknowledge the tendency of exporters to rely on trade credit during their ventures), the Czech Republic had a different exchange rate regime (without the asymmetric commitment) and generally, the world economy was in a different phase of their business cycle. It is expected that hypothesis H1 will not hold, empirical verification will be provided.

The principal source of data as for previous versions is the UN COMTRADE database[24] (United Nations Commodity Trade Statistics Database). The advantages of this database are its extensive coverage (it is the largest depositary of trade data, containing over 1 billion records[25]), its joint collection of data, its joint standards and its joint processing; therefore full harmonization is achieved. On the other hand, this dataset is not perfect as it is acknowledged that it contains statistical discrepancies as the mirror statistics confirm (exports globally are not equal to imports for any given year, and for many countries exports reported by one country to another do not match with the reported imports from the mentioned exporting country). The main cause of

[24] http://comtrade.un.org/db/help/uReadMeFirst.aspx [21/04/2016]

[25] Source: http://comtrade.un.org/db/help/uReadMeFirst.aspx [21/04/2016]

those discrepancies is that export data tend to be less monitored than the import data by the customs organizations (Bacchetta et al, 2014). As such, the "Rotterdam effect" is identified in which countries with coast serves as pure transits for many products. For them, a higher amount of trade volume than real is accounted. Correspondingly, mirror statistics for countries without coast such as the Czech Republic show in particular many discrepancies (statistically they tend to be erroneously recorded as final destination the country with a coast which serves for mere transition purposes) (Wlazel, 2012). As a final remark, it is necessary to mention that this database does not currently show all the world trade value despite its purpose and as it would be desirable (in 2015, the data showed only the 53, 37% of world trade value[26]). The assumption of the model is that the dataset provided, reflects the entire world trade but no such dataset exists at the moment (consequently, regrettably some markets during the filtration process are eliminated because of the inexistence of sufficient data, not because of their *a priori* adverse prospects).

Acknowledging discrepancies in trade statistics, some researchers have tried to estimate the real flows behind. An example is the research carried out by Wlazel (2012) in which it is argued that the current amount of Czech exports directed towards Germany is overrepresented. In fact, it is argued that many goods are directly re-exported without modifications to diverse destinations such as China. In that paper, it is estimated that in the period from 2004 to 2007 one fifth of reported exports to Germany where in fact re-exported. For 2007, the value of the estimated re-exported goods via Germany totalled 6085 million euros.

That fact would provide support for the sole inclusion of Germany in the revealed absence barriers to trade parameter discussed in the previous section instead as discussed, with the inclusion of Hungary. However, that research (Wlazel, 2012), is not supported by any empirical evidence as it is based on not proven hypothesis and quite strong assumptions. As, the proportion that re-exports a country X from another, say A is equal to the proportion of reported imports coming from A (Wlazel,

---

[26] Source: http://comtrade.un.org/db/default.aspx [27/04/2016]

2012 p 19); no verification was yet provided for that assumption. Therefore, it is preferred to include Hungary as discussed while; Germany still being included as well.

Another source of data for this model is the World Bank database, more in particular the World Development Indicators. Within this source, the GDP per capita at constant prices in PPP dollars of 2011 is extracted in order to construct the Prody index correctly. The LPI index was downloaded from its corresponding server[27], the FVA was retrieved thanks to the OECD database[28].

In the Tables 2 and 3, the COMTRADE dataset for 2014 can be observed. As summary statistics, the value of the reported trade flow is provided. Considering each flow, it is determined by a commodity of interest, a reporter (the country from which the flow is initialized) and a partner (the country which receives the flow). As anticipated, the commodities are coded using the HS6 and the countries are coded by the iso3 nomenclature.

Variable	Observations	Mean	St. Deviation	Minimum	Maximum
Value	7779587	2334251	9,12E+07	1	8,56E+10

Table 3: Summary statistics of imports 2014

Source: Author based on COMTRADE 2014

Variable	Observations	Mean	St. Deviation	Minimum	Maximum
Value	7276805	2466917	1,01E+08	1	1,11E+11

Table 4: Summary statistics of exports 2014

Source: Author based on COMTRADE 2014

[27] http://lpi.worldbank.org/ [21/04/2016]

[28] https://data.oecd.org/ [21/04/2016]

Given that dataset, each indicator described in the previous section is computed thanks to R software based on: Urban (2015a; 2015b), Hadley (2011), Grothendieck (2014), Dowle et al (2015), R Core Team (2015), and Eduardo and Maechler (2015) as well the filtering process and the calibration computations. When necessary; other years are used, not only data for 2014.

Regarding non-tariff barriers, it was indicated that they do affect potential prospects for companies starting to export to a given destination, so they should be included during the filtration process. But in practical terms; ad valorem non-tariff barriers most recent data is provided for 2005[29] at the world level. A possibility would be to use that source and assume that major changes did not happen during the last 10 years, but in practical terms, that assumption is not sustainable. Just an example, we see the following free trade agreements (include deals concerning tariff and non-tariff measures) implemented by the EU since 2005. It is shown that the number of signed agreements is not negligible, consequently the number of countries and markets involved.

29 //http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:21085330~menuPK:64214916~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html [21/04/2016]

Year	Partner	Form
2005	Algeria	Association Agreement
2005	Chile	Association Agreement and Additional Protocol
2008	Bosnia and Herzegovina	Stabilisation and Association Agreement
2009	Albania	Stabilisation and Association Agreement
2009	Madagascar, Mauritius, the Seychelles, and Zimbabwe	Economic Partnership Agreement
2009	Cameroon	Economic Partnership Agreement
2010	Montenegro	Stabilisation and Association Agreement
2010	Serbia	Interim Agreement on trade and trade related matters
2010	Korea	New Generation Free Trade Agreement
2011	Papua New Guinea and Fiji	Interim Partnership Agreement
2012	Iraq	Partnership and Cooperation Agreement
2012	Colombia and Peru	Trade Agreement
2012	Central America	Association Agreement with a strong trade component

Table 5: EU signed agreements in force since 2005

Source: Author based on EC (2016)

In addition, under way[30] are negotiations regarding the TTIP (Transatlantic Trade and Investment Partnership) with the US and the CETA (Comprehensive Economic and Trade Agreement) with Canada; major advances being expected (Pelkmans et al. 2015). Therefore, non-tariff barriers are advised to be evaluated in the third phase, during the in-depth examination of the potential export opportunities outputted after the filtration process (among other variables), if updated and verified data are found. In this work, only a retrieve of potential export opportunities will be performed.

For ad valorem tariffs, no data was found with sufficient detail neither at the country neither at the product level worldwide; therefore those variables can neither be included into the model during the filtration process. However, remembering their importance, it is advisable that they should be included during the in-depth examination phase.

[30] As present day, this article takes reference the 21/04/16.

5 Model Results

After the gathering all the needed indicators from the dataset, a filtering process was established. Consequently, a list of opportunities for 2010 and for 2014 was finally obtained. For 2010, the EDSM outputted 2300 market-product combinations; the list of the countries with the highest number of them is the following:

Countries	Number of Opportunities
Russian Federation	343
Poland	216
Slovakia	205
Ukraine	142
Germany	139
Turkey	134
India	130
Brazil	125
China	111
Hungary	90
Italy	83
South Africa	77
Israel	76
Switzerland	76
United Kingdom	76
Sweden	75
Rep. of Korea	73
Georgia	71
Netherlands	70
Thailand	69

Table 6: Number of opportunities by country for 2010

Source: Author based on COMTRADE

Broadly speaking, the list includes a mix of developed and developing countries. In appendix C, two figures (C1 and C2) in which all countries for which export opportunities exist for 2010 and for 2014 are shown.

In the Appendix B, the opportunities are displayed, classified by commodities for 2010 and for 2014 at the HS2 level, results are available at the HS6 level. Not surprisingly, the highest number of them corresponds to products related to machinery, rather than to agriculture, highlighting the predominance of the industry in the Czech economic structure.

In the next table, the sample output is displayed, this time for the model with data for 2014:

Countries	Number of Opportunities
Poland	380
Germany	340
Slovakia	318
Russian Federation	312
United Kingdom	304
Hungary	260
USA	244
France	226
Italy	208
Mexico	191
Netherlands	171
China	169
Spain	168
Turkey	163
Bulgaria	151
Lithuania	137
Canada	127
Croatia	127
Belgium	117
India	115

Table 7: Number of opportunities by country for 2014

Source: Author based on COMTRADE

Approximately, the same set of countries is included as for the model outputted for 2010, but this time in different order and quantity. Therefore, it can be concluded that the stated hypothesis does not hold because for 2010, 2300 opportunities were retrieved while for 2014, this number amounted to 5876.

It is worth mentioning that, it could be the case that both models would output the same number of opportunities but the hypothesis still would not be validated. The hypothesis stated was that both models would output identical country-product combinations.

In biology and in other scientific branches, scientific algorithms have been developed which compute retrieving an index, the similarity among samples. There are different and various, the most appropriate for this particular case being, the binary similarity indices. For their calculation, they frequently use the following table for the comparison of the two samples.

		Sample A	
		Present	Absent
Sample B	Present	a	b
	Absent	c	d

Table 8: Sample Combinations

Source: Author

Where a is the number of elements present in both samples and in the same position, b those present in sample A but not in B, c those present in sample B but not in A and d those which are neither present at A neither at B. So, this principle is next illustrated as, proposed by two samples. In the case used for the assessment of the model outputs, each position of the vector would represent a possible country-product[31] combination, being in each vector all possible country-product combinations and for each case 1 would represent that that market is an opportunity and 0 that it is not.

[31] In that case, the evaluation was performed at the level HS6 the outputted results are at level HS2 for practicality.

Sample A	1	0	1	0
Sample B	1	1	0	0
	a=1	b=1	c=1	d=1

Table 9: Parameters for the index

Source: Author

The most widely used binary similarity coefficient, is the Jaccard index; which is calculated taking into account the following:

$$\text{Jaccard} = \frac{a}{a+b+c} \qquad (19)$$

It takes values between 1 and 0, being 1 the case when both samples are equal.

Actually, there are multitudes of other indices developed by various researchers which basically display different combinations of the parameters (a, b, c and d).

$$\text{Sorensen} = \frac{a}{a+0.5(b+c)} \qquad (20)$$

$$\text{Kulczynski} = 0.5\left[\frac{a}{a+b} + \frac{a}{a+c}\right] \qquad (21)$$

In fact, there are many others more but according to the study conducted by Hubalek (1982), those cited are the preferred ones as they retain advisable properties. Those properties are: symmetry (the index would be the same regarding if it would be computed regarding the sample A with respect to B or sample B with respect to A), null association occur when a = 0 and they take values in the range 0 to 1.

Similarity 2010 respect to 2014	
Jaccard	0,0650
Sorenson	0,1221
Kulczynski	0,1509

Table 10: Similarity among the opportunities[32]

Source: Author based on Vavrek and Matew (2011)

Similarity 2010 respect to 2014	
a	499
b	5377
c	1811

Table 11: Similarity among the opportunities with coefficients

Source: Author

Empirically, in the table provided it can be easily observed that their similarity is reduced, as the values are far nearer for 0 than for 1. Even though this measured dissimilarity, we encounter for both years a common pattern, regarding the absolute preponderance in terms of quantity of export opportunities at the level HS2, “nuclear reactors, boilers, machinery and mechanical appliances parts thereof”.

Consequently, it can be deducted that this model should be actualized frequently and supplied whenever possible with the most actual datasets in order to produce meaningful and real results, as only the 21, 7% of the retrieved opportunities

[32] Computed based on R Core Team (2015) and Vavrek and Matthew (2011); as a robustness check there were also computed using the vegan package and the results were equal.

for 2010 were still opportunities retrieved by the model for 2014. Nevertheless, this time instability could be driven by the calibration methodology and the variables chosen. As such, there is the reasonable hypothesis that the model described in Urban et al (2014a) would not share that time-instability. Thus, the generalization of that model-time-instability to other models is not guaranteed, yet. But if demonstrated, then export promotion agencies should not rely on them for the design of their long term export promotion strategies as, always there would be uncertainty regarding if the results provided are still valid. And, remembering that data collection always takes time, so even provided the most up to date datasets, the results not necessary could be real and valid. As a result, the recommendation drawn in this research is that public agencies should not rely exclusively on those models, rather they should provide as much as possible information regarding foreign markets, and promote networking abroad, as indicated by various researchers, those are the major drivers allowing business to discover export opportunities (Chandra et al, 2009). Private enterprises such as EEIP can and do provide related services in those areas.

6 Robust Assessment

6.1 Sensitivity Analysis

During the model construction, the critical thresholds and the methodology selected were already discussed in previous sections. In this phase, an analysis is carried out evaluating the sensitivity of the outputted results with regard to variations in the thresholds. An exploration of the sensitivities of the different parameters is implemented. That procedure, also called sensitivity analysis is highly recommended by various institutions around the world as it provides transparency to the model (Borgonovo and Plischke, 2016) and helps to verify and validate it (ensuring the overall quality in the modelling process).

Consequently, the sensitivity analysis is performed by evaluating minor changes in the input vector of the system. In that thesis, still within the scope of the sensibility analysis framework rather than an evaluation given changes in the input parameters, the focus will be on the proper structural elements of the model specifically on the various thresholds given by the various parameters. As the major uncertainty concerning the model, is the selection of the appropriate parameters and thresholds.

In some applications, the target of the sensitivity analysis is to categorize the most important parameters of the model, those which dominate the model behaviour (Gan et al, 2014); those parameters for which if uncertainty regarding their values is eliminated, the model improvement will be maximal. As such, a ranking in terms of importance for the different thresholds will be provided.

Generally, within the scope of the sensitivity analysis, we can differentiate two basic methodologies; on the one hand local sensitivity and on the other, global sensitivity analysis (Saltelli, 2006, p 11). The purpose of the local sensitivity methods is to analyse the response of the system to changes in parameters independently. The common approach is to vary a parameter of interest deterministically around some predefined value by a random range. This method tends not to be suitable if there exist substantial interaction effects among parameters. An example of implementation of this method is a tornado diagram where the output of the model response to marginal changes for each independently analysed parameter is displayed. Furthermore, global sensitivity measures analyse the parameters of interest given changes in the total parameter space. In this way, the different interaction effects within parameters can be observed, therefore this method is more informative and recommended.

Given those two broad frameworks, we can identify a multitude of different methodologies tailor made to different circumstances. Cohesive reviews with detailed descriptions have been recently published by various renowned researchers. Borgonovo and Plischke (2016), Saltelli et al (2008) and Gan et al (2014) are recommended for further information.

The methodologies selected are; firstly the construction of sensitivity indices for all interested variables. It would allow to make *a priori* ranking in terms of importance, acknowledging that the interpretation has to be cautious as it is a filtering model and per se there are present non-linarites through the trade-offs implemented. After that, a graphical method will be performed; concretely, a tornado diagram. That would allow for a quick inspection of the internal behaviour of the model, in a comfortable manner.

Global sensitivity measures are advisable. But, at the moment it would not be feasible to performs such analysis as it would be necessary to check (2^17 – 18), 131054 possible combinations as there are 17 parameters in the model. That evaluation is advised to be performed by future research.

- ## *Sensitivity Indices:*

Mathematically, the dependence of a parameter on an input at a given point is performed through a partial derivative at the given point. In that case, a function which we could derive is not available. Therefore, a simplification proposed by Lenhart et al (2002, p 646) will be applied; being the sensitivity index at each calibrated value given by the following formula (22), where, y is the number of export opportunities outputted by the model and x is a value inside the range of variation of the threshold:

$$I_s = \frac{(y2 - y1)\ x0}{2\ \Delta x\ y0} \qquad (22)$$

Sensitivity Indice[33]

[33] $\Delta x = x2 - x1$

For each standardized sensitivity index, Lenhart et al (2002, p 646) provides the following interpretation.

Class	Index	Sensitivity
I	0 <= I_s < 0,05	Small to negligible
II	0,05 <= I_s < 0,2	Medium
III	0,2 <= I_s < 1	High
IV	I_s >= 1	Very high

Table 12: Interpretation of Sensitivity Indices

Source: Lenhart et al (2002)

Consequently, the complete list for each filter and the parameter of that filter is finally outputted. As it can be observed (table 13), the most sensitive parameter is the last filter. The cause of that is the fact that the last filter encompasses remaining ones. Just as a remainder, each filter labelled from 0 to 1 or 2 or 3 successive markets. So, the last filter eliminate based on the non-zero values which the markets were labelled previously. All other parameters are highly sensitive, at least a threshold within them. Finally, the least sensitive parameters are the: RMS, LPI and KM, the reason for this is, as observed in appendix C, those filters not eliminate markets individually, but through a sum.

Parameters	Sensitivity Index 2010	Sensitivity Index 2014	Sensitivity
HHI 1	0,000	-0,019	Small
HHI 2	-0,534	-0,353	High
HHI 3	0,000	-0,165	Small/Medium
KM 1	-0,003	-0,004	Small
KM 2	0,000	0,000	Negligible
LPI	0,012	0,021	Small
RMS	-0,058	0,000	Small
LFI 1	0,000	0,000	Negligible
LFI 2	0,541	0,375	High
LFI 3	0,000	0,000	Negligible
MI 1	0,000	-0,281	Small/High
MI 2	-0,219	-0,315	High
MI 3	-0,045	-0,051	Small
Prody 1	0,032	0,088	Small/Medium
Prody 2	0,552	0,468	High
Prody 3	0,443	0,307	High
Final Threshold	6,388	4,801	Very High

Table 13: Sensitivity indices of the parameters of the model

Source: Author based on Lenhart et al (2002)

- *Tornado Diagram:*

It is a graphical representation of the quantitative change from a base scenario of the output of a model (y) to changes in input parameters (x_s). The range of applied values in x_s for the documentation of the behaviour of the model is arbitrary. Nevertheless, an increase and a decrease of 10% tends to be applied, a range of variation of the 100% is applied in the provided examples, except for the final threshold were the range of variation is 14, 29%. This graphical local sensitivity method, which documents the individual effects of each variable x_s is usually constructed vertically. Specifically, each variable x_s and its range of change is situated in one column, so immediately it can be observed which one has the major local effect. At the provided examples for both years, the initial position is the number of opportunities without any modification in the thresholds. Only those which present major ranges of variation are displayed.

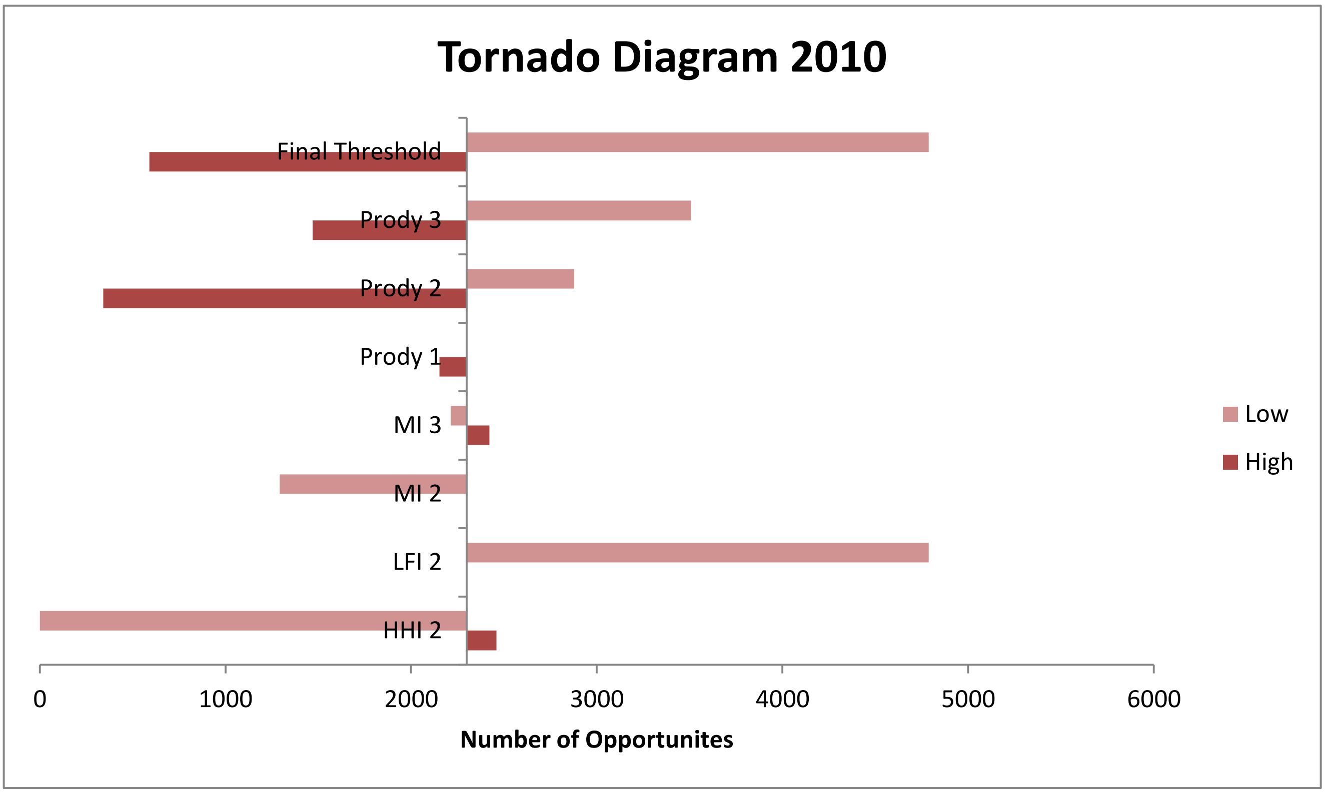

Figure 12: Tornado diagram for 2010

Source: Author

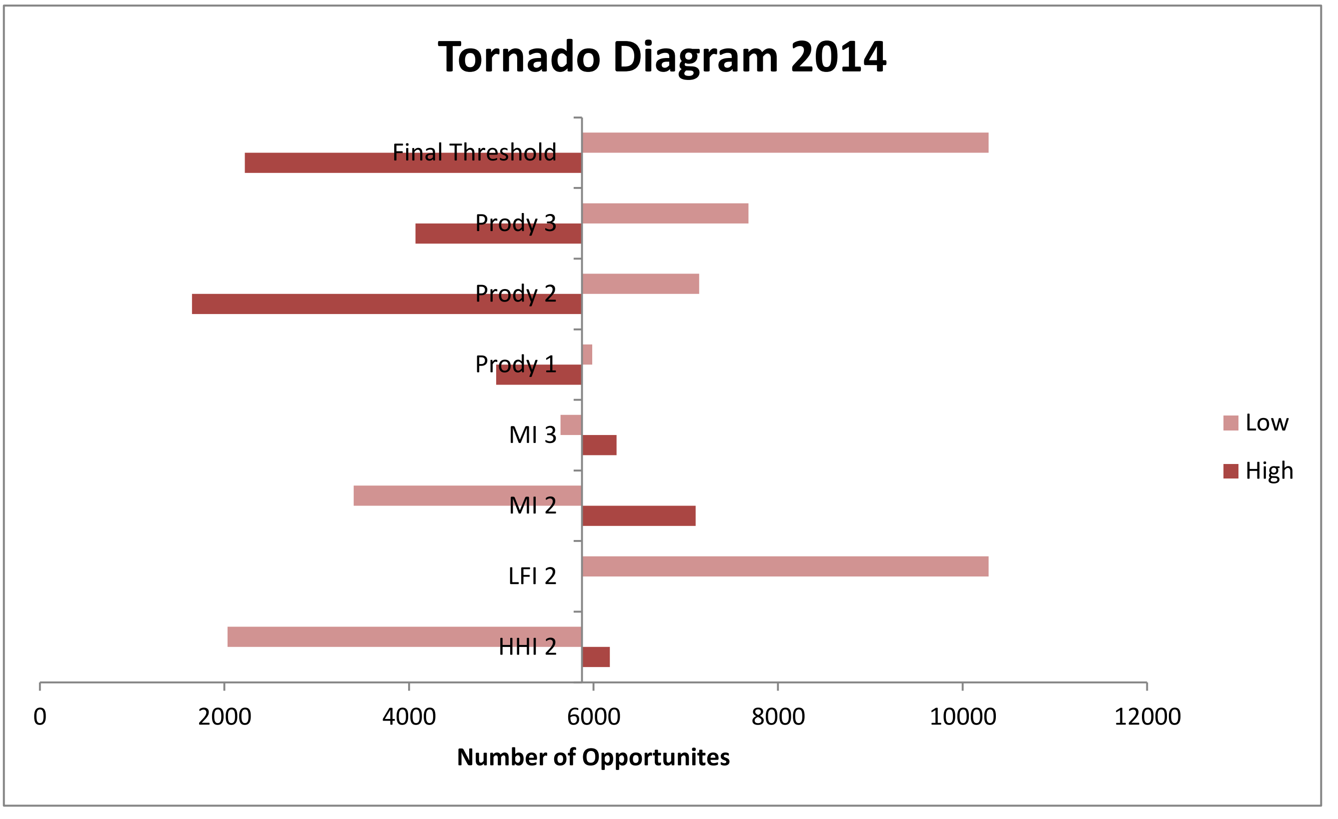

Figure 13: Tornado diagram for 2014

Source: Author

Again, it is observed that the final thresholds are the most influential while the effect for the different parameters is not symmetric.

6.2 Robustness of Additions

During the model construction, previous studies conducted by researchers in the field (being the major inspiration, the study performed by Urban et al (2014a), as encompasses all previous research) were considered as the starting point; some additions were then incorporated considering their usefulness and necessity. Using the same framework as for the evaluation of the initially provided hypothesis, at that phase it is evaluated if the additions which have been done, convey large variations in the model results (as a base model in this case, the one performed by Urban et al (2014a), is used). It is necessary to mention that the model is extremely dependent on the provided dataset, therefore for a thorough evaluation it would be convenient to perform that displayed analysis taking into consideration, many different datasets for different years.

Similarity 2010 with/without adds	
Jaccard	0,9007293
Sorenson	NAN
Kulczinski	NAN

Table 15: Similarity 2010[34]

Source: Author

Similarity 2014 with/without adds	
Jaccard	0,8600755
Sorenson	0,9247749
Kulczinski	0,926085

Table 16: Similarity 2014[35]

Source: Author

[34] Computed based on R Core Team (2015) and Vavrek and Matthew (2011)

[35] Computed based on R Core Team (2015) and Vavrek and Matthew (2011)

Similarity 2010 with/without adds	
a	2223
b	168
c	77

Table 17: Similarity 2010 with coefficients [36]

Source: Author

Similarity 2014 with/without adds	
a	5237
b	213
c	639

Table 18: Similarity 2014 with coefficients[37]

Source: Author

As can be observed in the tables above and, given the provided results, the transformations do not convey strong impact on the model results. Therefore, if there is no available enough data for those filters, a model without those parameters will tend still to produce meaningful and valid results.

[36] Computed based on R Core Team (2015) and Vavrek and Matthew (2011)

[37] Computed based on R Core Team (2015) and Vavrek and Matthew (2011)

7 Conclusion

The aim of this thesis was the development of an EDSM taking into account all the previous knowledge in the field. In addition, the hypothesis that the model results would be equal given two different time intervals was formulated. The aim of this testing was to evaluate if there is a necessity for actualizations. As anticipated, that hypothesis was not credible *a priori* because the model takes into consideration short term economic fluctuations. This was then finally tested and refuted. Therefore, it is recommended to supply the model whenever possible with the most updated datasets, and export promotion agencies should not rely on it during their long term strategies. That result could be generalized to other market selection models although a proper empirical approach is advisable before any conclusion is drawn. It is worth to mentioning, that it would be necessary to apply this test to many different datasets in order to gather conclusive results. As consequence, up to provided evidence, export promotion agencies should rather than display lists of export opportunities; focus on providing all the possible relevant information of foreign markets and enabling business to expand their ties abroad. As a consequence, corporations will be better equipped to discover export opportunities by themselves, according to their idiosyncratic characteristics and the current international economic environment.

In addition, an overall robustness assessment to the model was implemented, uncovering their behaviour, a novel approach not yet implemented in an EDSM targeted to the Czech Republic. Finally, it was concluded that the last filters have the highest importance to the model; therefore explicit research targeting their usefulness should be carried out. Furthermore, the interaction effects among the parameters were not evaluated at the moment, that task is currently unfeasible.

As explained, the importance of those decision support systems is immense as they allow different companies to select worth foreign markets while reducing uncertainties, allowing them to grow. And, as it is demonstrated in the current economic literature, exporting companies who use them display overall higher rates of success during their initialized export ventures. Nevertheless, no scientific rule exists at the moment for the construction of those models as all of them, incorporate a subjective component. As a result, in this thesis there was performed an extensive literature review to gather all relevant information for their construction; selecting the most relevant studies and driving conclusion from them.

Bibliography

Andersen, O., & Buvik, A. (2002). Firms' internationalization and alternative approaches to the international customer/market selection. *International Business Review*, 11(3), 347-363.

Andersen, P. H., & Strandskov, J. (1997). International market selection: A cognitive mapping perspective. *Journal of Global Marketing*, 11(3), 65-84.

Anderson, J. E., & Van Wincoop, E. (2004). Trade costs. *Journal of Economic literature*, 42(3), 691-751.

Armstrong, J. S. (1970). An application of econometric models to international marketing. *Journal of Marketing Research,* 190-198.

Arvis, J. F., Saslavsky, D., Ojala, L., Shepherd, B., Busch, C., & Raj, A. (2014). *Connecting to Compete 2014: Trade Logistics in the Global Economy--The Logistics Performance Index and Its Indicators*. Washington: The World Bank
Source:
http://lpi.worldbank.org/ [14/04/2016]

Audretsch, D. B., Keilbach, M. C., & Lehmann, E. E. (2006). *Entrepreneurship and economic growth*. Oxford: Oxford University Press.

Bacchetta M., Beverelli C., Cadot Marco Fugazza O., Grether Matthias J., Alessandro H. Roberta Piermartini, N. (2014) *A practical guide for Trade Policy Analysis* Geneva: WTO

Barreiro, P. L., & Albandoz, J. P. (2001). Population and sample. Sampling techniques. *Management Mathematics for European Schools MaMaEusch* (994342-CP-1-2001-1-DECOMENIUS-C21.

Bilkey, W. J. (1978). An attempted integration of the literature on the export behavior of firms. *Journal of International Business Studies*, 9(1), 33–46.

Blonigen, B. A., Liebman, B. H., Pierce, J. R., & Wilson, W. W. (2013). Are all trade protection policies created equal? empirical evidence for non equivalent market power effects of tariffs and quotas. *Journal of International Economics*, 89(2), 369-378

Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research,* 248(3), 869-887.

Brouthers, L. E., & Nakos, G. (2005). The role of systematic international market selection on small firms' export performance. *Journal of Small Business Management*, 43(4), 363-381.

Brouthers, L. E., Mukhopadhyay, S., Wilkinson, T. J., & Brouthers, K. D. (2009). International market selection and subsidiary performance: A neural network approach. *Journal of World Business*, 44(3), 262-273.

Cavusgil, S. T. (1985). Guidelines for export market research. *Business Horizons*, 28(6), 27-33.

Chandra, Y., Styles, C., & Wilkinson, I. (2009). The recognition of first time international entrepreneurial opportunities: Evidence from firms in knowledge-based industries. *International Marketing Review*, 26(1), 30-61.

Chor, D., & Manova, K. (2012). Off the cliff and back? Credit conditions and international trade during the global financial crisis. *Journal of international economics*, 87(1), 117-133

CNB (2015): *Analyses of the Czech Republic's current economic alignment with the euro area*, Praha: CNB

Cuyvers, P. De Pelsmacker, G. Rayp and I. Roozen (1995): "A Decision Support Model for the Planning and Assessment of Export Promotion Activities by Government Promotion Institutions: the Belgian Case", *International Journal of Research in Marketing*,.Vol.12, No.2, pp.173-186.,

Cuyvers, L.; Rossouw, R.; Steenkamp, E.; Viviers, (2009). Export Market Selection Methods and the Identification of Realistic Export Opportunities for South Africa Using an Export Decision Support. Arcadia: Trade & Industrial Policy Strategies Working Paper Series 3

Cuyvers, L., & Viviers, W. (Eds.). (2012). *Export Promotion-A Decision Support Model Approach*. Johannesburg: Sun Media Metro.

Cuyvers, L.,Viviers, W & Ermie A. Steenkamp. (Eds.). (2016) *The Utilization of export opportunities in the world market: A comparison of Belgium, South Africa and Thailand* North-West University. South Africa.
Source:
https://www.researchgate.net/publication/294888087_The_utilisation_of_export_opportunities_in_the_world_market_A_comparison_of_Belgium_South_Africa_and_Thailand [04/04/2016]

Dikmen, I., & Birgonul, M. T. (2004). Neural network model to support international market entry decisions. *Journal of Construction Engineering and Management*, 130(1), 59-66.

Dimitratos, P., Plakoyiannaki, E., Pitsoulaki, A., & Tüselmann, H. J. (2010). The global smaller firm in international entrepreneurship. *International Business Review,* 19(6), 589-606.

Dowle M., Srinivasan A., Short T, Lianoglou S with contributions from Saporta R an d Antonyan E (2015). data.table: Extension of Data.frame. R package version 1.9.6

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, 35(5), 352-359.

EC (2016) Agreements
Source:
http://ec.europa.eu/trade/policy/countries-and-regions/agreements/#_customs-unions [21/04/2016]

ECB (2013) The external environment of the euro area Monthly Bulletin
Source:
https://www.ecb.europa.eu/pub/pdf/other/mb201306_focus01.en.pdf
[21/4/2016]

Eduardo L. T. Conceicao and Maechler M. (2015). DEoptimR: Differential Evolution Optimization in Pure R． R package version 1.0-4

Estrategia, S. G. D. A. (2005). Metodología de la Secretaría de Estado de Turismo y Comercio para la selección de mercados prioritarios. *Boletín ICE Económico: Información Comercial Española,* (2836), 3-15.

Finger, J. M., & Kreinin, M. E. (1979). A Measure of Export Similarity and Its Possible Uses. *The Economic Journal*, 89(356), 905-912.

Forsgren, M. (2002). The concept of learning in the Uppsala internationalization process model: a critical review. *International business review*, 11(3), 257-277.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Berlin: Springer series in statistics.

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W. & Di, Z. (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software*, 51, 269-285.

Gereffi, G., & Fernandez-Stark, K. (2011). Global value chain analysis: a primer. Center on Globalization, Governance & Competitiveness (CGGC), Duke University, North Carolina, USA.

Gould, R.R. (2002) *International markets selection screening technique: Replacing intuition with a multidimensional Framework to select a short--list of countries*. (Doctoral dissertation), RMIT University, Melbourne.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3, 1157-1182.

Grodzicki, M.J. (2014). Global Value Chain and Competitiveness of V4 Economies (chapter 1). In: Kiendl-Wendner, D. & Wach, K. (eds), *International Competitiveness in Visegrad Countries: Macro and Micro Perspectives.* Graz: Fachhochschule Joanneum, pp. 13-31.

Grothendieck G. (2014). sqldf: Perform SQL Selects on R Data Frames. R package version 0.4-10

Hadley W. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1), 1-29

Hindle, T. (2003). *Guide to Management Ideas* (2 ed.) London: Profile Books

Hubalek, Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. *Biological Reviews*, 57(4), 669-689.

Irac, D., & Lopez, J. (2015). Euro area structural convergence? A multi-criterion cluster analysis. *International Economics*, 143, 1-22.

Jeannet, J. P., & Hennessey, H. D. (2005). *Global marketing strategies* Boston: Dreamtech Press.

Keegan, W. J & Schlegelmilch, B. B. (2001). *Global marketing management: A European perspective*., Harlow: Pearson Education.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30, 607-610

Lenhart, T., Eckhardt, K., Fohrer, N., & Frede, H. G. (2002). Comparison of two different approaches of sensitivity analysis. *Physics and Chemistry of the Earth, Parts A/B/C*, 27(9), 645-654.

Leromain, E., & Orefice, G. (2014). New revealed comparative advantage index: Dataset and empirical distribution. *International Economics*, *139*, 48-70.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica,* 23(2), 143-149.

Miroudot, S., & De Backer, K. (2013). Mapping Global Value Chains. Working Paper (forthcoming), OECD, Paris.

Müllner, D. (2013). fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software*, 53(9), 1-18.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S.,(2005). *Tools for Composite Indicators Building*. Ispra: European Commission

OECD (2015) Trade in Value Added: Czech Republic
Source:
https://www.oecd.org/sti/ind/tiva/CN_2015_CzechRepublic.pdf [14/04/2016]

Onkvisit, S., & Shaw, J. J. (2004). *International marketing: Analysis and strategy.* New York: Routledge.

Ozorhon, B., Dikmen, I. and Birgonul, M. T. (2006) "Case-Based Reasoning Model for International Market Selection", *Journal of Construction Engineering and Management,* 940-947, September

Papadopoulos, N., & Denis, J. E. (1988). Inventory, taxonomy and assessment of methods for international market selection. *International Marketing Review*, 5(3), 38-51.

Papadopoulos, N., Chen, H., & Thomas, D. R. (2002). Toward a tradeoff model for international market selection. *International Business Review,* 11(2) 165-192.

Patel, V. M., Gopalan, R., Li, R., & Chellappa, R. (2015). Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine*, 32(3), 53-69.

Pelkmans, J., Lee, T. M., Mustilli, F., Cernat, L., Norman-López, A., Tucci, A., ... & Matthes, J. (2015). TTIP: Political and Economic Rationale and Implications. *Intereconomics,* 50(6), 312-343.

Peterson, R. T. (1990). Screening is the First Step in Evaluating Foreign Markets. *Marketing News,* 24(14), 9

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Romer D. (2011). *Advanced Macroeconomics.* New York: McGraw-Hill/Irwin

Sakarya, S., Eckman, M., & Hyllegard, K. H. (2007). Market selection for international expansion: Assessing opportunities in emerging markets. *International Marketing Review*, 24(2), 208-238.

Saltelli, A. (2006). *The critique of modelling and sensitivity analysis in the scientific discourse. an overview of good practices*. Institute for the Protection and Security of the Citizen, European Commission, Luxemburg, ISSN, 1018-5593.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. New York: John Wiley & Sons.

Sarstedt, M. & Mooi, E. A. (2014). *A concise guide to market research. The process, data, and methods using IBM SPSS Statistics*. Berlin: Springer

Schumacher, R. (2013). Deconstructing the theory of comparative advantage. *World Economic Review*, 2, 83-105.

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of management review*, 25(1), 217-226.

Steenkamp, J. B. E., & Ter Hofstede, F. (2002). International market segmentation: issues and perspectives. International *Journal of Research in Marketing*, 19(3), 185-213.

Steenkamp, E. A. (2011). *The identification of export opportunities for South African products with special reference to Africa* (Doctoral dissertation), North-West University.

Suddaby, R., Bruton, G. D., & Si, S. X. (2015). Entrepreneurship through a qualitative lens: Insights on the construction and/or discovery of entrepreneurial opportunity. *Journal of Business Venturing*, 30(1), 1-10.

Tlapa M., Klepácek R. & Svoboda M. (2015) *Mapa Globalních oborových příležitostí* [Global map of sectorial opportunities] MZV [Ministry of Foreign Affairs]: Prague

Urban, M., Mejstřík, M., & Chvalkovská, J. G. (2014a) Application of the decision support model to Czech exports. *Acta Oeconomica Pragensia*, 2014(2), 33-47.

Urban M., Gutierrez Chvalkovská J., Mejstřík M., (2014b) The Discriminate Analysis and Relevant Model for Mapping Export Potential *World Academy of Science, Engineering and Technology International Journal of Social, Education, Economics and Management Engineering* Vol:8, No:10, 2014

Urban M. (2015a) SQL Commands for EDSM [SQL]

Urban M. (2015b) STATA Commands for EDSM [STATA]

Vavrek, Matthew J. (2011). fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, 14:1T

Verikas, A., & Bacauskiene, M. (2002). Feature selection with neural networks. *Pattern Recognition Letters*, 23(11), 1323-1335.

Vlčková J., De Castro T & Antal J. (2015) Upgrading in the Global Value Chains: The Case of the Czech Republic Conference Paper VŠE

Wilke, A., & Mata, R. (2012). *Cognitive bias. In V. S. Ramachandran (Ed.),* Encyclopedia of human behavior (2nd edition). Maryland Heights, MD: Elsevier

WLAZEL, M (2012) *Reexporty a nepřímé exporty visegrádských zemí přes Německo do Číny.* Praha, 58 s. Bakalářská práce (Bc.) Univerzita Karlova, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí diplomové práce Mgr. Ing. Vilém Semerák, Ph.D.

Wood, V. R., & Robertson, K. R. (2000). Evaluating international markets: The importance of information by industry, by country of destination, and by type of export transaction. *International Marketing Review*, 17(1), 34-55.

World Economic Situation and Prospects (2012)
Source:
https://www.google.cz/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=0CCkQFjACahUKEwiwyIz7q8TIAhWEliwKHXENBPg&url=http%3A%2F%2Fwww.un.org%2Fen%2Fdevelopment%2Fdesa%2Fpolicy%2Fwesp%2Fwesp_archive%2F2012chap2.pdf&usg=AFQjCNH47sfRPmZ4zy4SsADHbtFCgf-b0Q&sig2=_4447mUQG3zBh7k5u5lcWA&cad=rja [10/12/2015]

WTO (2011) World Trade in 2010
Source:
https://www.wto.org/english/res_e/booksp_e/anrep_e/wtr11-1_e.pdf
[21/02/2016]

Yale (2015) Grad Trade notes
Source:
www.econ.yale.edu/~ka265/teaching/GradTrade/notes/ClassNotes.pdf
[01/09/2015]

Zaghini, A. (2003). Trade advantages and specialization converge in accession countries. ECB Working Paper Series working paper no.249

Appendix A

- *Variables used for the Hierarchical clustering, figure 10 based on Irac and Lopez (2015),*

Control of Corruption (cont.corrp): Indicator assessing the degree of corruption control, latest available data is for 2014 Source[38]

Employment Rates (Employ.rate): Latest available data is for 2014. Source[39]

Expenditure on Education (exp.educ): Expense on education as a percent of GDP, latest available data is for 2011. Source[40]

Government Effectiveness (gov.effect): Indicator assessing the degree of government efffectivness, latest available data is for 2014 Source[41]

Long Term Unemployment (L.Unemp): Long term unemployment as percent of total unemployment, latest available data is for 2014. Source[42]

Perception of Corruption (Percp.Corrupt): Index measuring how corrupt sector are seen to be. Constructed by Transparency International, latest data is for 2015. Source[43]

Patents (patents): Number of triadic patents per million of inhabitants, latest available data is for 2011. Source[44]

---

[38] http://info.worldbank.org/governance/wgi/index.aspx#reports [28/06/2016]

[39] http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Employment_rate,_age_group_15%E2%80%9364,_2004%E2%80%9314_(%25)_YB16.png [28/06/2016]

[40] http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc510&plugin=1 [28/06/2016]

[41] http://info.worldbank.org/governance/wgi/index.aspx#reports [28/06/2016]

[42] http://data.worldbank.org/indicator/SL.UEM.LTRM.ZS [28/06/2016]

[43] http://www.transparency.org/research/cpi/overview [28/06/2016]

[44] http://www.oecd-ilibrary.org/economics/oecd-factbook-2014/triadic-patent-families-number-per-million-inhabitants-2011_factbook-2014-graph144-en [28/06/2016]

PISA (Pisa): Dataset evaluating school performance, latest available dataset is for 2012. Source[45]

Product Market Regulation (Prod.reg): "Measure the degree to which policies promote or inhibit competition in areas of the product market where competition is viable", latest available source is 2013. Source[46]

Political Stability and Absence of violence (Pol.Stab): Index, latest available for 2014 Source[47]

R&D Expenditure (RDExp): Amount in percent of GDP allocated to R&D, latest available dataset is for 2012. Source[48]

Regulatory Quality (Reg.Q): Index, latest available for 2014. Source[49]

Rule of Law (Rule.Law): Indicator assessing the degree of the rule of law, latest available data is for 2014 Source[50]

Strictness of Employment Protection (StrictEMpl): Index measuring the strictness of protection of employment legislation, latest data is for 2014. Source[51]

Woman Employment Rate (Wom.Emprate): Employment rate of woman, latest data is for 2013. Source[52]

Temporary Unemployment (Temp.Unemp): Employment with a defined termination as percent of total population. Latest available data is for 2013. Source[53]

Working Age Population (Work.Age.Pop): Percent of population in age of working, latest data is for 2014. Source[54]

[45] http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm [28/06/2016]

[46] https://stats.oecd.org/index.aspx?DataSetCode=PMR [28/06/2016]

[47] http://info.worldbank.org/governance/wgi/index.aspx#home [28/06/2016]

[48] http://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS [28/06/2016]

[49] http://info.worldbank.org/governance/wgi/index.aspx#home [28/06/2016]

[50] http://info.worldbank.org/governance/wgi/index.aspx#reports [28/06/2016]

[51] https://stats.oecd.org/Index.aspx?DataSetCode=EPL_OV [28/06/2016]

[52] http://www.oecd-ilibrary.org/employment/employment-rate-of-women_20752342-table5 [28/06/2016]

[53] https://data.oecd.org/emp/temporary-employment.htm [28/06/2016]

- *The next table provides the correlation table given the entire standardized dataset for the 13 countries selected:*

	cont.corrp	Employ.rate	exp.educ	gov.effect	L.Unemp	Percp.Corrup	patents	Pisa	Prod.reg	RDExp	Pol.Stab	Reg.Q	Rule.Law	Wom.Emprate	StrictEMpl	Temp.Unemp	Work.Age.Pop
cont.corrp	1,00	0,76	0,86	0,94	-0,78	0,99	0,75	0,59	-0,13	0,69	0,41	0,91	0,96	0,91	-0,11	-0,23	-0,50
Employ.rate	0,76	1,00	0,48	0,73	-0,67	0,76	0,53	0,42	-0,31	0,56	0,55	0,85	0,82	0,94	0,05	-0,48	-0,21
exp.educ	0,86	0,48	1,00	0,79	-0,76	0,87	0,59	0,52	-0,05	0,58	0,39	0,72	0,80	0,68	-0,20	-0,18	-0,50
gov.effect	0,94	0,73	0,79	1,00	-0,73	0,95	0,78	0,46	-0,19	0,74	0,47	0,86	0,96	0,87	-0,02	-0,17	-0,50
L.Unemp	-0,78	-0,67	-0,76	-0,73	1,00	-0,79	-0,46	-0,47	-0,10	-0,50	-0,53	-0,68	-0,80	-0,80	-0,09	0,03	0,13
Percp.Corrup	0,99	0,76	0,87	0,95	-0,79	1,00	0,75	0,61	-0,12	0,71	0,44	0,92	0,97	0,90	-0,12	-0,20	-0,46
patents	0,75	0,53	0,59	0,78	-0,46	0,75	1,00	0,46	-0,30	0,88	0,44	0,67	0,75	0,65	0,19	-0,10	-0,58
Pisa	0,59	0,42	0,52	0,46	-0,47	0,61	0,46	1,00	0,14	0,56	0,25	0,66	0,56	0,48	0,09	0,15	-0,13
Prod.reg	-0,13	-0,31	-0,05	-0,19	-0,10	-0,12	-0,30	0,14	1,00	-0,37	0,02	-0,33	-0,25	-0,26	0,23	0,70	0,61
RDExp	0,69	0,56	0,58	0,74	-0,50	0,71	0,88	0,56	-0,37	1,00	0,54	0,66	0,77	0,66	0,25	-0,24	-0,55
Pol.Stab	0,41	0,55	0,39	0,47	-0,53	0,44	0,44	0,25	0,02	0,54	1,00	0,43	0,46	0,57	0,38	-0,22	-0,02
Reg.Q	0,91	0,85	0,72	0,86	-0,68	0,92	0,67	0,66	-0,33	0,66	0,43	1,00	0,91	0,93	-0,16	-0,34	-0,47
Rule.Law	0,96	0,82	0,80	0,96	-0,80	0,97	0,75	0,56	-0,25	0,77	0,46	0,91	1,00	0,93	-0,01	-0,28	-0,46
Wom.Emprate	0,91	0,94	0,68	0,87	-0,80	0,90	0,65	0,48	-0,26	0,66	0,57	0,93	0,93	1,00	0,02	-0,41	-0,39
StrictEMpl	-0,11	0,05	-0,20	-0,02	-0,09	-0,12	0,19	0,09	0,23	0,25	0,38	-0,16	-0,01	0,02	1,00	0,23	0,08
Temp.Unemp	-0,23	-0,48	-0,18	-0,17	0,03	-0,20	-0,10	0,15	0,70	-0,24	-0,22	-0,34	-0,28	-0,41	0,23	1,00	0,44
Work.Age.Pop	-0,50	-0,21	-0,50	-0,50	0,13	-0,46	-0,58	-0,13	0,61	-0,55	-0,02	-0,47	-0,46	-0,39	0,08	0,44	1,00

Table A1: Correlation of standardized dataset

Source: Author

As can be easily observable, some variables are correlated in excess (>= 0, 9). Therefore, for the cluster analysis the variables perception of corruption, rule of law, woman employment rate and control of corruption were eliminated as they provide redundant information following the recommendations of Sarsted and Mooi, (2014); as a result, a set of relevant and non-redundant variables, appropriate for hierarchical clustering was obtained.

[54] https://data.oecd.org/pop/working-age-population.htm [28/06/2016]

Appendix B

Commodities	Number of Opportunities
"Nuclear reactors, boilers, machinery and mechanical appliances; parts thereof"	735
"Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments and apparatus; parts and accessories thereof"	187
"Electrical machinery and equipment and parts thereof; sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles,"	159
"Tools, implements, cutlery, spoons and forks, of base metal; parts thereof of base metal."	111
"Articles of iron or steel."	99
"Plastics and articles thereof"	68
"Glass and glassware."	62
"Organic chemicals."	52
"Miscellaneous articles of base metal."	50
"Furniture; bedding, mattresses, mattress supports, cushions and similar stuffed furnishings; lamps and lighting fittings, not elsewhere specified or included; illuminated signs, illuminated name-plates and the like; prefabricated buildings."	45
"Vehicles other than railway or tramway rolling-stock, and parts and accessories thereof."	45
"Iron and steel."	43
"Paper and paperboard; articles of paper pulp, of paper or of paperboard."	42
"Articles of stone, plaster, cement, asbestos, mica or similar materials."	37
"Toys, games and sports requisites; parts and accessories thereof"	33
"Rubber and articles thereof"	32
"Essential oils and resinoids; perfumery, cosmetic or toilet preparations."	29
"Inorganic chemicals; organic or inorganic compounds of precious metals, of rare-earth metals, of radioactive elements or of isotopes."	28
"Miscellaneous manufactured articles."	27
"Musical instruments; parts and accessories of such articles."	25
"Miscellaneous chemical products."	22
"Arms and ammunition; parts and accessories thereof."	18
"Printed books, newspapers, pictures and other products of the printing industry; manuscripts, typescripts and plans."	17
"Dairy produce; birds' eggs; natural honey; edible products of animal origin, not elsewhere specified or included."	16

Table B1: Number of opportunities by commodities 2010 - 1

Source: Author

Commodities	Number of Opportunities
"Impregnated, coated, covered or laminated textile fabrics; textile articles of a kind suitable for industrial use."	16
"Photographic or cinematographic goods."	16
"Man-made staple fibres."	14
"Tanning or dyeing extracts; tannins and their derivatives; dyes, pigments and other colouring matter; paints and varnishes; putty and other mastics; inks."	14
"Articles of leather; saddlery and harness; travel goods, handbags and similar containers; articles of animal gut (other than silk-worm gut)."	13
"Oil seeds and oleaginous fruits; miscellaneous grains, seeds and fruit; industrial or medicinal plants; straw and fodder."	13
"Aluminium and articles thereof"	10
"Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal and articles thereof; imitation, jewellery; coin."	10
"Railway or tramway locomotives, rolling-stock and parts thereat railway or tramway track fixtures and fittings and parts thereof; mechanical (including electro-mechanical) traffic signalling equipment of all kinds."	10
"Special woven fabrics; tufted textile fabrics; lace; tapestries; trimmings; embroidery."	10
"Ceramic products."	9
"Clocks and watches and parts thereof."	9
"Man-made filaments."	9
"Other base metals; cermets; articles thereof."	9
"Soap, organic surface-active agents, washing preparations, lubricating preparations, artificial waxes, prepared waxes, polishing or scouring preparations, candles and similar articles, modelling pastes, \"dental waxes\" and dental preparations with a basis of plaster."	9
"Aircraft, spacecraft, and parts thereof."	8
"Pharmaceutical products."	8
"Works of art, collectors' pieces and antiques."	8
"Articles of apparel and clothing accessories, not knitted or crocheted."	7

Table B2: Number of opportunities by commodities 2010 - 2

Source: Author

Commodities	Number of Opportunities
"Cotton,"	7
"Lac; gums, resins and other vegetable saps and extracts."	7
"Products of the milling industry; malt; starches; inulin; wheat gluten."	7
"Wool, fine or coarse animal hair; horsehair yarn and woven fabric."	7
"Edible vegetables and certain roots and tubers."	6
"Preparations of vegetables, fruit, nuts or other parts of plants."	5
"Products of animal origin, not elsewhere specified or included"	5
"Wood and articles of wood; wood charcoal,"	5
"Articles of apparel and clothing accessories, knitted or crocheted."	4
"Carpets and other textile floor coverings."	4
"Copper and articles thereof"	4
"Mineral fuels, mineral oils and products of their distillation; bituminous substances; mineral waxes."	4
"Preparations of meat, of fish or of crustaceans, molluscs or other aquatic invertebrates."	4
"Wadding, felt and nonwovens; special yarns; twine, cordage, ropes and cables and articles thereof"	4
"Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes."	3
"Beverages, spirits and vinegar."	3
"Furskins and artificial fur; manufactures thereof."	3
"Meat and edible meat offal."	3
"Miscellaneous edible preparations."	3
"Other vegetable textile fibres; paper yarn and woven fabrics of paper yarn."	3
"Salt; sulphur; earths and stone; plastering materials, lime and cement."	3
"Cereals."	2
"Fertilisers."	2
"Headgear and parts thereof"	2
"Knitted or crocheted fabrics."	2
"Live animals."	2
"Pulp of wood or of other fibrous cellulosic material; recovered (waste and scrap) paper or paperboard."	2
"Tobacco and manufactured tobacco substitutes."	2
"Albuminoidal substances; modified starches; glues; enzymes."	1
"Cork and articles of cork."	1
"Nickel and articles thereof."	1
"Other made up textile articles; sets; worn clothing and worn textile articles; rags."	1
"Preparations of cereals, flour, starch or milk; pastrycooks' products."	1
"Residues and waste from the food industries; prepared animal fodder."	1
"Ships, boats and floating structures."	1
"Zinc and articles thereof."	1

Table B3: Number of opportunities by commodities 2010 - 3

Source: Author

Commodities	Number of Opportunities
"Nuclear reactors, boilers, machinery and mechanical appliances; parts thereof"	1557
"Electrical machinery and equipment and parts thereof; sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles,"	498
"Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments and apparatus; parts and accessories thereof"	480
"Plastics and articles thereof"	238
"Articles of iron or steel."	228
"Tools, implements, cutlery, spoons and forks, of base metal; parts thereof of base metal."	207
"Organic chemicals."	182
"Glass and glassware."	164
"Vehicles other than railway or tramway rolling-stock, and parts and accessories thereof."	162
"Paper and paperboard; articles of paper pulp, of paper or of paperboard."	139
"Miscellaneous articles of base metal."	129
"Furniture; bedding, mattresses, mattress supports, cushions and similar stuffed furnishings; lamps and lighting fittings, not elsewhere specified or included; illuminated signs, illuminated name-plates and the like; prefabricated buildings."	127
"Iron and steel."	116
"Articles of stone, plaster, cement, asbestos, mica or similar materials."	89
"Rubber and articles thereof"	78
"Miscellaneous manufactured articles."	75
"Toys, games and sports requisites; parts and accessories thereof"	73
"Inorganic chemicals; organic or inorganic compounds of precious metals, of rare-earth metals, of radioactive elements or of isotopes."	71
"Clocks and watches and parts thereof."	70
"Impregnated, coated, covered or laminated textile fabrics; textile articles of a kind suitable for industrial use."	68
"Essential oils and resinoids; perfumery, cosmetic or toilet preparations."	66
"Miscellaneous chemical products."	61

Table B4: Number of opportunities by commodities 2014 - 1

Source: Author

Commodities	Number of Opportunities
"Tanning or dyeing extracts; tannins and their derivatives; dyes, pigments and other colouring matter; paints and varnishes; putty and other mastics; inks."	60
"Dairy produce; birds' eggs; natural honey; edible products of animal origin, not elsewhere specified or included."	59
"Articles of leather; saddlery and harness; travel goods, handbags and similar containers; articles of animal gut (other than silk-worm gut)."	54
"Ceramic products."	49
"Aluminium and articles thereof"	48
"Railway or tramway locomotives, rolling-stock and parts thereat railway or tramway track fixtures and fittings and parts thereof; mechanical (including electro-mechanical) traffic signalling equipment of all kinds."	47
"Oil seeds and oleaginous fruits; miscellaneous grains, seeds and fruit; industrial or medicinal plants; straw and fodder."	33
"Soap, organic surface-active agents, washing preparations, lubricating preparations, artificial waxes, prepared waxes, polishing or scouring preparations, candles and similar articles, modelling pastes, \"dental waxes\" and dental preparations with a basis of plaster."	32
"Arms and ammunition; parts and accessories thereof."	31
"Articles of apparel and clothing accessories, not knitted or crocheted."	31
"Pharmaceutical products."	31
"Wood and articles of wood; wood charcoal,"	31
"Copper and articles thereof"	29
"Man-made filaments."	28
"Photographic or cinematographic goods."	26
"Printed books, newspapers, pictures and other products of the printing industry; manuscripts, typescripts and plans."	25
"Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal and articles thereof; imitation, jewellery; coin."	24

Table B5: Number of opportunities by commodities 2014 - 2

Source: Author

Commodities	Number of Opportunities
"Products of the milling industry; malt; starches; inulin; wheat gluten."	21
"Man-made staple fibres."	19
"Footwear, gaiters and the like; parts of such articles,"	18
"Musical instruments; parts and accessories of such articles."	18
"Preparations of vegetables, fruit, nuts or other parts of plants."	16
"Articles of apparel and clothing accessories, knitted or crocheted."	14
"Nickel and articles thereof."	14
"Special woven fabrics; tufted textile fabrics; lace; tapestries; trimmings; embroidery."	13
"Umbrellas, sun umbrellas, walking-sticks, seat-sticks, whips, riding-crops and parts thereof"	13
"Cotton,"	12
"Preparations of meat, of fish or of crustaceans, molluscs or other aquatic invertebrates."	12
"Preparations of cereals, flour, starch or milk; pastrycooks' products."	10
"Wool, fine or coarse animal hair; horsehair yarn and woven fabric."	10
"Meat and edible meat offal."	9
"Other base metals; cermets; articles thereof."	9
"Wadding, felt and nonwovens; special yarns; twine, cordage, ropes and cables and articles thereof"	9
"Miscellaneous edible preparations."	8
"Salt; sulphur; earths and stone; plastering materials, lime and cement."	8
"Albuminoidal substances; modified starches; glues; enzymes."	7
"Carpets and other textile floor coverings."	7
"Edible vegetables and certain roots and tubers."	7
"Fertilisers."	7
"Lac; gums, resins and other vegetable saps and extracts."	7
"Mineral fuels, mineral oils and products of their distillation; bituminous substances; mineral waxes."	7
"Residues and waste from the food industries; prepared animal fodder."	7

Table B6: Number of opportunities by commodities 2014 - 3

Source: Author

Commodities	Number of Opportunities
"Beverages, spirits and vinegar."	6
"Live animals."	6
"Other made up textile articles; sets; worn clothing and worn textile articles; rags."	6
"Ships, boats and floating structures."	6
"Coffee, tea, mate and spices."	5
"Explosives; pyrotechnic products; matches; pyrophoric alloys; certain combustible preparations."	5
"Other vegetable textile fibres; paper yarn and woven fabrics of paper yarn."	5
"Works of art, collectors' pieces and antiques."	5
"Aircraft, spacecraft, and parts thereof."	4
"Cocoa and cocoa preparations."	4
"Edible fruit and nuts; peel of citrus fruit or melons."	4
"Furskins and artificial fur; manufactures thereof."	4
"Fish and crustaceans, molluscs and other aquatic invertebrates."	3
"Cork and articles of cork."	2
"Headgear and parts thereof"	2
"Products of animal origin, not elsewhere specified or included"	2
"Pulp of wood or of other fibrous cellulosic material; recovered (waste and scrap) paper or paperboard."	2
"Sugars and sugar confectionery."	2
"Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes."	1
"Prepared feathers and down and articles made of feathers or of down; artificial flowers; articles of human hair."	1
"Silk."	1
"Vegetable plaiting materials; vegetable products not elsewhere specified or included."	1
"Zinc and articles thereof."	1

Table B7: Number of opportunities by commodities 2014 - 4

Source: Author

Appendix C

Filter 1		
Measures	Thresholds	Label
Market Size	Data	1/0[55]
Short Term Growth	Data	1/0
Long Term Growth	Data	1/0
F1 [Size + Short Growth + Long Growth]	Figure 7 [white part out]	F1

Table C1: Filter 1 2010

Source: Author

Filter 2		
Measures	Thresholds	Label
HHI	If F1 == 3 0.7584567	1/0
	If F1 == 2 0.5997948	1/0
	If F1 == 1 0.2426495	1/0
RMS	0,95	1/0
KM	5654	2/-
	9864	1/0
LPI	2,65	1/0

Table C2: Filter 2 2010

Source: Author

[55] Through this notation was it is understood is that if the market surpass the threshold, then it is labelled as 1, otherwise is labelled as 0.

Filter 3		
Measures	Thresholds	Label
LFI CZ - LFI Partner	lower than 0	0
	lower than 0,023	1
	lower than 0,21	2
	else	3

Table C3: Filter 3 2010

Source: Author

Filter 4		
Measures	Thresholds	Label
MI top 6 countries - MI CZ	higher than 2.9647229	0
	higher than 1.4975073	1
	higher than 0.3038163	2
	else	3

Table C4: Filter 4 2010

Source: Author

Filter 5		
Measures	Thresholds	Label
Prody	lower than 10173 lower than 18500 lower than 26970 else	0 1 2 3
F1 or HHI or LFI or [RMS + KM + LPI]	0	opportunity / not opportunity
Total F1 + LFI + MI + Prody	7	opportunity / not opportunity

Table C5: Filter 5 2010

Source: Author

Filter 1		
Measures	Thresholds	Label
Market Size	Data	1/0
Short Term Growth	Data	1/0
Long Term Growth	Data	1/0
F1 [Size + Short Growth + Long Growth]	Figure 7 [white part out]	F1

Table C6: Filter 1 2014

Source: Author

Filter 2		
Measures	Thresholds	Label
HHI	If F1 == 3 0.633016	1/0
	If F1 == 2 0.5981624	1/0
	If F1 == 1 0.2635241	1/0
RMS	0,95	1/0
KM	5654	2/-
	9864	1/0
LPI	2,65	1/0

Table C7: Filter 2 2014

Source: Author

Filter 3		
Measures	Thresholds	Label
LFI CZ - LFI Partner	lower than 0	0
	lower than 0,023	1
	lower than 0,19	2
	else	3

Table C8: Filter 3 2014

Source: Author

Filter 4		
Measures	Thresholds	Label
MI top 6 countries - MI CZ	higher than 2.7536558	0
	higher than 1.6841612	1
	higher than 0.4581677	2
	else	3

Table C9: Filter 4 2014

Source: Author

Filter 5		
Measures	Thresholds	Label
Prody	lower than 10847	0
	lower than 18464	1
	lower than 26479	2
	else	3
F1 or HHI or LFI or [RMS + KM + LPI]	0	opportunity / not opportunity
Total F1 + LFI + MI + Prody	7	opportunity / not opportunity

Table C10: Filter 5 2014

Source: Author

- *The data dependent thresholds take the following form :*

For short term and long term threshold (G):

$$G = \begin{cases} G_{w,j} * s_j \text{ if } G_{w,j} > 0 \\ G_{w,j} / s_j \text{ if } G_{w,j} =< 0 \end{cases}$$

$$s_j = \frac{1}{(RCAj+0.85)e^{(RCA-0.01)}}$$

$$RCAj = \frac{\frac{Xcz,j\ (1-FVAcz,j)}{Xw,j}}{\frac{Xcz\ (1-FVAcz)}{Xw}}$$

For market size (Sj):

$$Sj = \begin{cases} 0,02*M_{w,j} & \text{if } RCA_j > 0 \\ \frac{3-RCAj}{100} * M_{w,j} & \text{if } RCA_j <= 0 \end{cases}$$

G_w = Short or long term world growth export of product j

$X_{cz,\,j}$ = Czech export of product j

$X_{cz,}$ = Total Czech export

$X_{w,\,j}$ = World export of product j

X_w = Total world export

$M_{w,\,j}$ = World import of product j

$FVA_{cz,\,j}$ = Average import content of exports in the industry j

FVA_j = Average import content of Czech exports

Source: Urban et al (2014a), Cuyvers et al (2009) and, Cuyvers and Viviers (2012).

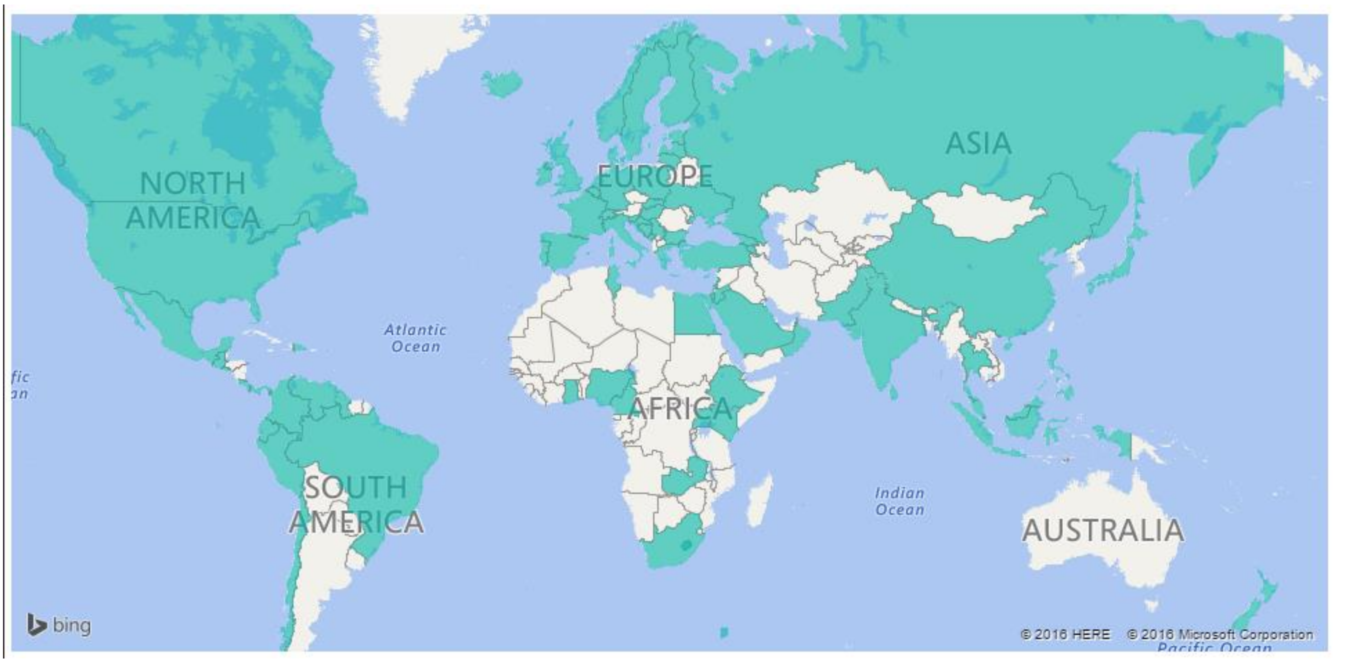

Figure C1: Countries with opportunities 2010

Source: Author based on Power BI

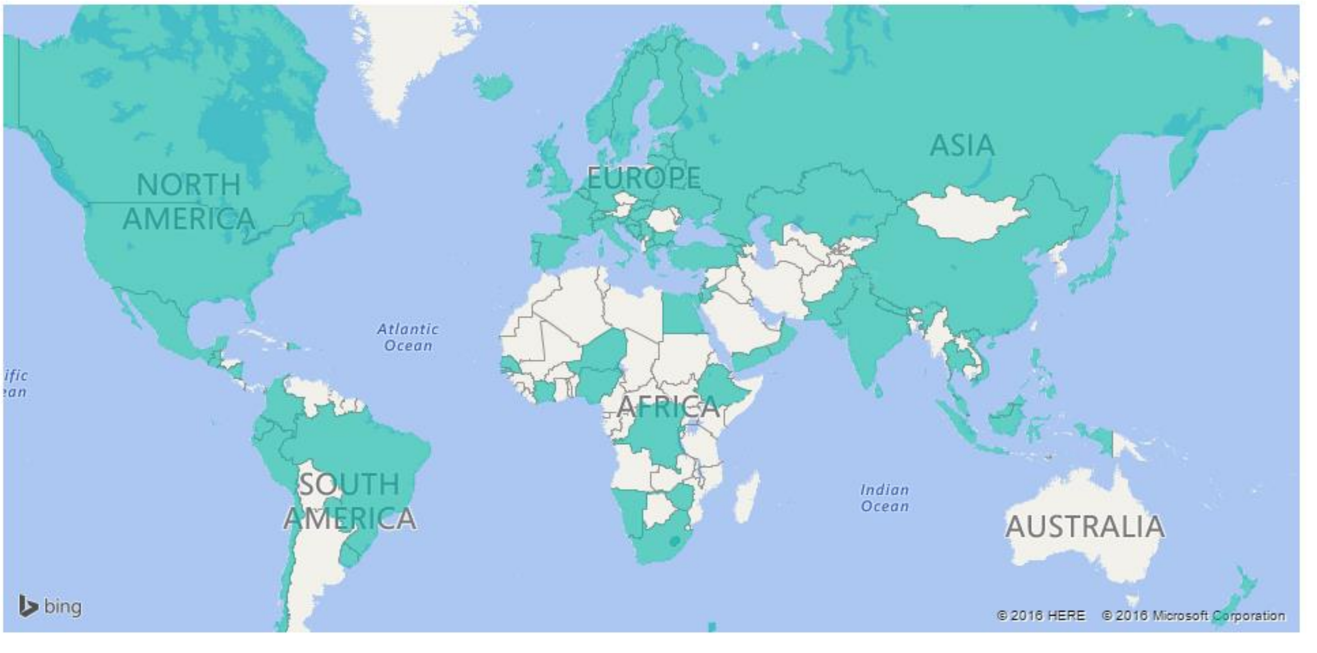

Figure C2: Countries with opportunities 2014

Source: Author based on Power BI

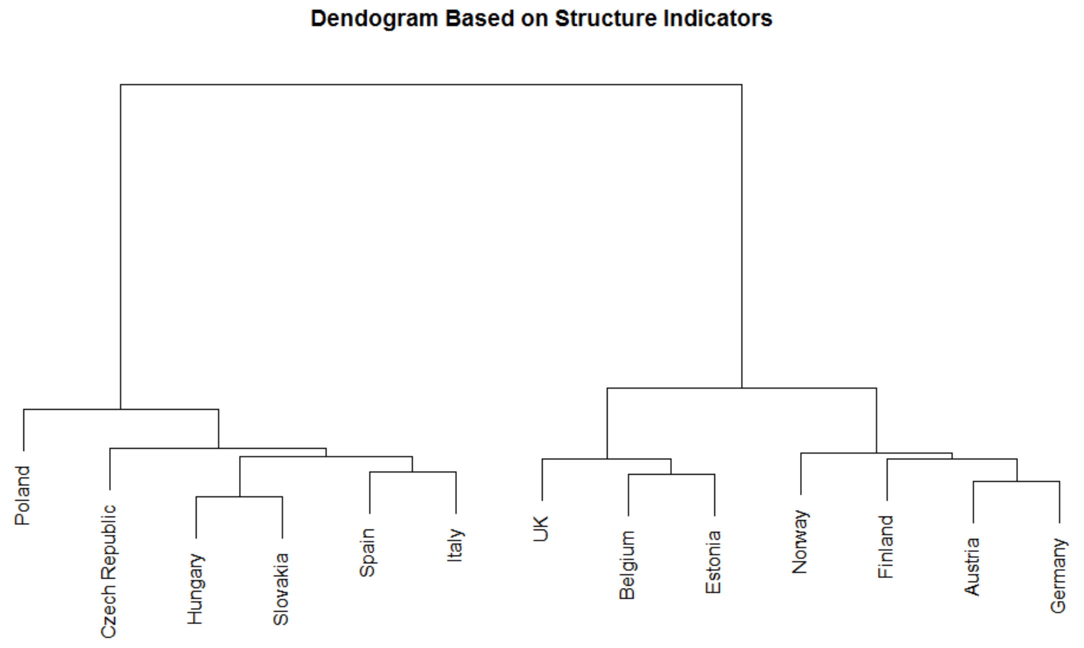

Figure C3: Cluster dendogram ward linkage

Source: Author based on Müllner (2013)

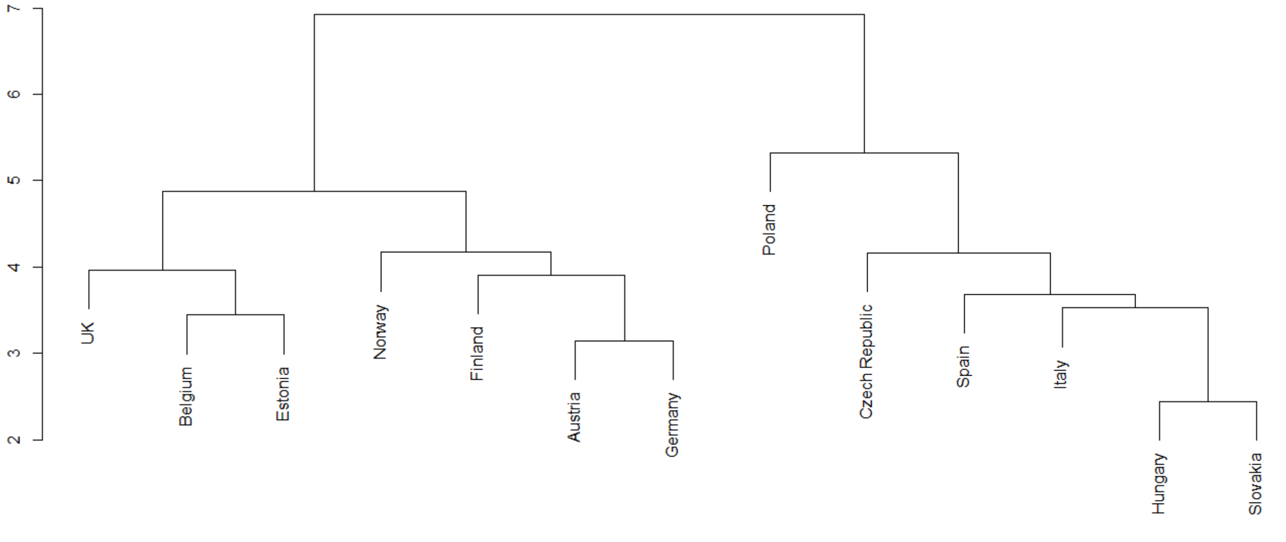

Figure C4: Cluster dendogram average linkage

Source: Author based on Müllner (2013)

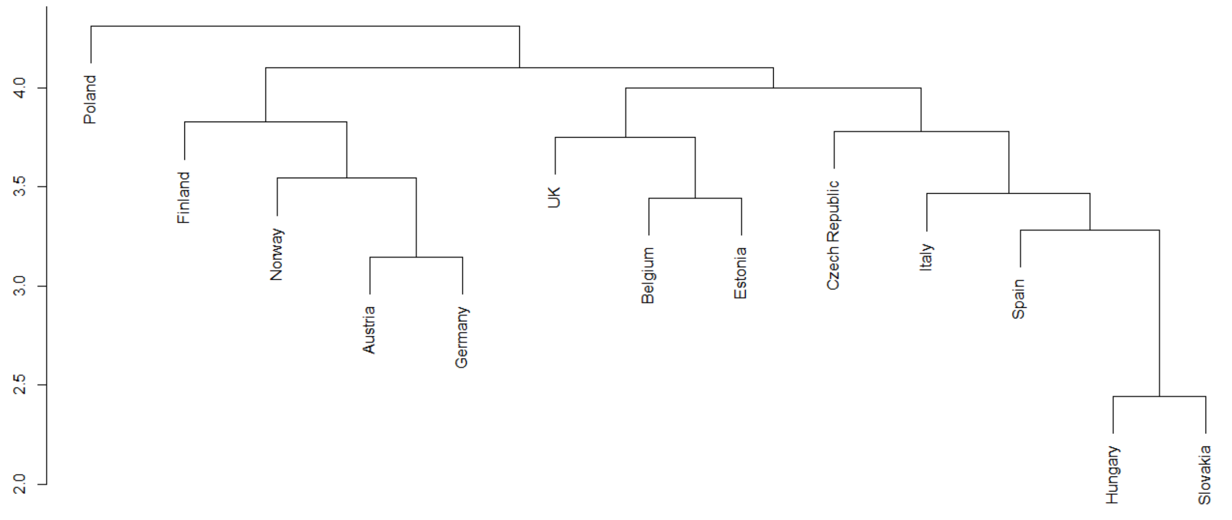

Figure C5: Cluster dendogram single linkage

Source: Author based on Müllner (2013)