This thesis investigates the factors that influence the migration of highly educated people between developed countries. A panel regression is run on the database consisting of the number of foreign born in the 28 selected OECD countries, specified by the country of birth and education level. The explanatory factors studied are indicators of average income, job security, income equality, specific government expenditures, knowledge intensity, taxes and other barriers. The results show that the highly educated migrants value job security and government expenditure on education. Governments should consider these factors when creating policies to attract foreign talent.

What drives knowledge migration

A study on factors influencing knowledge migration between developed countries

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Department of Economics

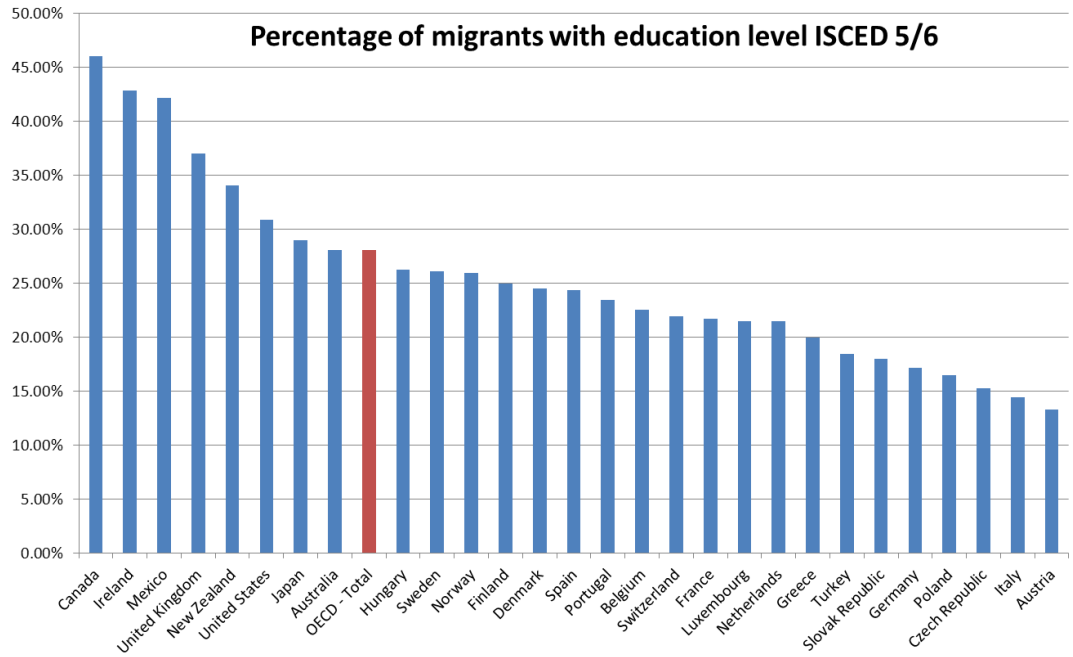
Supervisor: Dr. B.S.Y. Crutzen

Name: Charlotte Janssen
Exam number: 328380
E-mail adress: 328380cj@student.eur.nl

Contents

1. Introduction

On several occasions in the past it was witnessed how educated and/or skilled people fled developing countries for a better future in developed countries. This is also known as the socio-economic phenomenon: *'brain drain'*. These migration flows had negative consequences for the countries left behind, which meant that the gap with the developed countries remained as big as it was. Yet the emigration of highly skilled labor force is not merely a burden of the developing countries. In the current international competitive situation, human capital flows freely between the developed countries as well. The importance of attracting highly educated people has only recently been drawn attention to. In October 2012, The Boston Consulting Group published a report on the Dutch "earnings model" in the year 2030. It explains the importance of attracting foreign innovation talent in order to compete with other countries (Kurstjens, Maas, & Steffens, 2012). This type of talent is not necessarily linked to education level, but there does seem to be a connection between education level and the success of entrepreneurs (Van der Sluis, Van Praag, & Vijverberg, 2008). According to Liebig and Sousa-Poza, there is a competition for skilled migrants among OECD countries due to skills shortages (Liebig & Sousa-Poza, 2004).



Graph 1.1 – *Share of high educated migrants*
The number of migrants in a country with an education level of ISCED 5 and 6, divided by the total number of migrants living in that country. Migrants are defined as 'foreign- born'.
Source: OECD DIOC

From graph 1.1, it is clear that there is quite some difference in the relative amount of high educated migrants countries attract. This brings us to wonder why some countries are more successful at

attracting high educated migrants than others. The research question of this thesis therefore is: "*What factors influence the migration of highly educated people in OECD countries?*".

In order to answer the research question, the Database on Immigrants in OECD Countries (DIOC) is used. This database gives information about the countries of birth of the inhabitants of the OECD countries and their education level.

Knowledge migration is in this research defined as migration of people with an education level of International Standard Classification of Education (ISCED) levels 5 and 6. These are the highest two educational levels. In many countries, there is no clear distinction between the two, which is why both levels are included.

The outline of this thesis is as follows. In the next section, the main topics in the literature regarding this topic are discussed. In section 3, the data that is used for this research is discussed and in section 4 the main results are shown. In section 5, these results are discussed and a conclusion will be given in section 6.

2. Literature review

The most important factors to influence the migration of high educated people that can be found in literature are discussed in this chapter. The findings are divided in main topics which are most often addressed, being income, risk aversion, the availability of public services, networks of migrants, income (in)equality and other barriers and stimulants for migration. However, it is important to keep in mind that there are many types of high educated migrants and that each type is subject to its own push and pull factors (Mahroum, 2000). The choice for a country is a complex process (Geis, Uebelmesser, & Werding, 2008) and it is therefore difficult to draw conclusions for the entire group of migrants.

Before looking at the factors that influence migration, it is important to consider why the immigration of highly educated people is so important. This is only discussed to some extent in literature. The literature that is available, points towards its importance for the economy. According to the BCG report, immigration can help dealing with the consequences of an aging society (Kurstjens, Maas, & Steffens, 2012). Liebig and Sousa-Poza claim that many OECD countries have skills shortages, because of which the importance of highly skilled migration is increasing (Liebig & Sousa-Poza, 2004). This is also emphasized by the BCG report. The net import of talent should help us compete at an international level (Kurstjens, Maas, & Steffens, 2012). Although education level is not necessarily linked with skill or talent, it does seem to be related to the success of entrepreneurs (Van der Sluis, Van Praag, & Vijverberg, 2008).

2.1 Income

An important reason for immigrants (both high and lower educated) to move abroad is to find a better job. Migration is often regarded as an investment and is subject to economic incentives (Bowles, 1970). It is thus not surprising that wages have an effect on migration (Geis, Uebelmesser, & Werding, 2008). People are more likely to move to countries where they can earn more for the same job. Furthermore, Chiswick explained how migration incentives are a function of the ratio of wages in the country of destination to the origin (Chiswick, 1999). Since this research focusses on the share of high educated migrants in the total number of migrants, it is important to look specifically at the benefits for this group. Only when high educated people have more to gain from migration than lower educated people, their share in the total number of migrants will increase.

2.2 Risk aversion

Moving to another country can bring risks for a migrant. According to Heitmueller, risk averse individuals are less likely to migrate (Heitmueller, 2005). Unemployment in itself also has a negative effect on migration (Geis, Uebelmesser, & Werding, 2008). Job security might therefore be an

important criterion in choosing a foreign country to live in (Geis, Uebelmesser, & Werding, 2008). It is questionable, however, whether this would influence high educated migrants more than lower educated migrants. Perhaps it would even influence them even less. Therefore, the effects of unemployment and job security through legislation on the migration of higher educated people are investigated in this research.

2.3 Public services

Another important criterion can be the infrastructure and facilities of a country. For example, the availability of proper education and health care can be important (Geis, Uebelmesser, & Werding, 2008). People often prefer a country where the public services are good. High taxes, however, have a negative effect on the migration of high educated people, even though they generally facilitate better public services (Geis, Uebelmesser, & Werding, 2008). Both tax level and public expenditures on education are therefore included in this research.

2.4 Network

Another factor found in the literature is the function of existing networks of migrants. Especially scientists are known to migrate to centers of expertise (Mahroum, 2000), but it also applies to other groups of (high educated) migrants. Networks allow a potential migrant to receive detailed information about the country and provide a social network when he decides to move there (Geis, Uebelmesser, & Werding, 2008). Once there is a network, it will increase even more, for example through family re-unification (Geis, Uebelmesser, & Werding, 2008). Due to certain limitations, networking effects are beyond the scope of this research.

2.5 Equality

Income (in)equality has also been identified as a factor of migration. One can expect this to especially influence lower educated people, since they are often in the lower income scales and have more to benefit from migration to countries with more income equality. According to Liebig and Sousa-Poza however, highly educated people are generally more inclined to migrate, due to which they still will migrate more if the income inequality in their country of origin is high (Liebig & Sousa-Poza, 2004). Due to the possible difference in effects between lower en higher educated migrants, it is especially interesting to include this factor in the research. The proxy's for income equality are the income share of the richest 10% and the Gini index, as discussed in the next chapter.

2.6 Barriers and stimulants

Besides these factors that influence migration flows, there can also be some factors that additionally stimulate or hold back people to migrate.

Studying abroad is known to be a stimulant for moving abroad. It allows students to get used to the foreign country. It can be expected that between 15 and 35% of students stay in the foreign country they studied in (OECD, 2009). International student mobility contributes to human development and global understanding (Kehm, 2005). This can in turn also improve professional mobility. Factors that influence international student mobility flows are for example cost of living differences, quality rankings of universities, distance between countries and the behavior of other students (Gonzalez, Mesanza, & Mariel, 2011). However, the limitations of this research do not allow including this as a factor in explaining high educated migration.

On the other hand, the immigration policy of a country can form a barrier. According to Mahroum, Europe is regarded as bureaucratic and cumbersome (Mahroum, 2000). On the other hand, the influence of the immigration policy might not be that important to the migrants personally, since it is often taken care of by their employer (Berkhout, Smid, & Volkerink, 2010). This could be especially true for high educated people.

Another barrier can be formed by language differences and the distance between countries. One can expect people to migrate more easily to countries where the same language is spoken as in their country of origin. If not the same language is spoken in two countries, this might form a barrier for migration streams between those countries. Countries with a language that is spoken widely outside the national borders, like English-speaking countries, have an advantage in attracting migrants (OECD, 2009). It is however unclear if this effect differs for high educated people compared to lower educated people. The distance between countries can also form such a barrier. One can imagine it to be easier to migrate to a neighbor country than to a country on the other side of the world. It is difficult to say if this effect differs for high educated people. Perhaps they care less about distance than others, but this can't be said with certainty. The impact of the language and distance barriers on migration is included in this study, while the immigration policy of a given country is not.

2.7 Summary

In this chapter, we have seen that already various ideas exist on what factors influence migration and the migration of high educated people. However, the literature does not consider the effect of different factors on the number of high educated migrants compared to the total number of migrants. This is therefore what will be subject of this thesis. In the context of this research it is important to keep in mind the difference between high and low educated people, but also that there are many different types of high educated people, each of which subject to its own push and pull factors (Mahroum, 2000). In the next chapter the variables that are used in this research will be specified.

3. Data

In this chapter, the data that is used for the research is discussed. First, the dataset used for the construction of the dependent variable is discussed. Next, all the independent variables (x-variables) are discussed by what they mean, why they were chosen, how they were changed and what a statistically significant outcome for this variable would mean. In the last part of this chapter, the methodology for this research will be given.

3.1 Main dataset and the dependent variable[1]

The main dataset is the Database on Immigrants in OECD Countries (DIOC). This database gives information on the immigrants in OECD countries: their country of birth, their education level and their age. The OECD is an international organization of cooperation between countries on topics of economic growth and financial stability. The member states represent the majority of the developed countries, therefore the migration flows between these countries are chosen as the subject of this research. A selection of 28 OECD countries was used (table 3.1), since only for these OECD countries data on education level was available.

Table 3.1

Australia	France	Luxembourg	Slovak republic
Austria	Germany	Mexico	Spain
Belgium	Greece	Netherlands	Sweden
Canada	Hungary	New Zealand	Switzerland
Czech republic	Ireland	Norway	Turkey
Denmark	Italy	Poland	United Kingdom
Finland	Japan	Portugal	United States

Table 3.1 – Countries included in the research
These are the 28 countries that are included in the research. They are the OECD countries for which data on countries of birth and education level is available.

All the data is from around the year 2000, when an extensive round of censuses was held. For this research, the data on foreign-born of age 25-64 was used, since at that age, most have reached their final education level. The research looks at all the immigrants that were born in one of these countries and are currently a resident of another one of these countries. That way, there are 28 x 27 possible combinations. Each of these combinations 'from – to' has a total number of migrants and a number of migrants with a high education level. The share of high educated migrants as a percentage of the total number of migrants is used as the dependent variable. The variable is called '$\%HEM_{i,j}$', which stands for the percentage high educated migrants to country of residence i, from country of birth j.

$$\%HEM_{i,j} = \frac{migrants\ from\ j\ to\ i\ with\ a\ high\ education\ level}{total\ number\ of\ migrants\ from\ j\ to\ i} \qquad \textbf{Equation 1}$$

[1] The statistics of the dependent variable are stated in Appendix B2

3.2 Independent variables[2]

In this section, the variables which are used in the regressions are discussed. For each variable it is explained why it is chosen, what it means and if necessary how it is adapted to be able to work with that variable, as well as what a significant outcome would mean.

For certain factors, the following formula is used:

$$relative\ value = \frac{value\ country\ of\ residence - value\ country\ of\ birth}{value\ country\ of\ birth} \quad \textbf{Equation 2}$$

In this way, a value for each combination of countries is obtained. This value than indicates a 'percentage improvement' for moving from the country of birth to the country of residence.

In other cases, the value in the country of birth is simply subtracted from the value for the country of residence:

$$relative\ value = value\ country\ of\ residence - value\ country\ of\ birth \quad \textbf{Equation 3}$$

This formula is for example used when the values for each country are already percentages.

Furthermore, an original value is in certain cases divided by the GDP of the country. This is done in order to take the size of the country into account. This is done to correct for the fact that larger countries generally have higher output of for example scientific journal articles.

In the following sub-sections, the independent variables are explained. For each variable, the meaning, the reason why and its implications are given. Since the dependent variable is from the year 2000, so are the data for each independent variable.

GDP per capita

The first factor which can be seen as an important factor for migration is the urge of the immigrant to increase his welfare. The most common proxy for the welfare level of a country is the GDP per capita. This variable gives the gross domestic product divided by the midyear population in the year 2000 in current U.S. dollars[3]. The data is transformed to a relative difference in GDP per capita between the country of birth and the country of residence by the following calculation:

Equation 4

$$Relative\ GDP\ per\ capita = \frac{GDP\ per\ capita\ country\ of\ residence - GDP\ per\ capita\ country\ of\ birth}{GDP\ per\ capita\ country\ of\ birth}$$

[2] The raw data and statistics of the relative data for all the countries are in Appendix B1 and B2
[3] Data.worldbank.org

Or in short:

$$GDP_{i,j} = \frac{GDP_i - GDP_j}{GDP_j} \qquad \textbf{Equation 5}$$

A significant positive effect would indicate that highly educated people in particular are drawn to countries with a high GDP per capita. GDP per capita can be seen as an indicator of average wealth.

Unemployment

As seen in the literature, migrants can be risk averse. One of the risks they face is unemployment. This variable is the unemployment as a percentage of the total workforce. Unemployment is in this case defined as the share of the labor force that is without work but available for and seeking employment, although the definitions of labor force and unemployment differ per country[4]. According to Geis et al., unemployment has a negative effect on migration (Geis, Uebelmesser, & Werding, 2008), which is why it is used as a variable in this research.

In order to compare the unemployment rates between two countries, the relative difference of the unemployment rates between the two countries is taken:

$$unemployment_{i,j} = \frac{unemployment_i - unemployment_j}{unemployment_j} \qquad \textbf{Equation 6}$$

A significant negative effect would indicate that people with a high education level are relatively less attracted to countries with a high unemployment rate than people with a low education.

Employment Protection Legislation (EPL) indicator

The OECD indicators of employment protection measure the procedures and costs involved in dismissing individuals or groups of workers and the procedures involved in hiring workers on fixed-term or temporary work agency contracts[5]. It ranks the legal requirements for dismissals on a scale from 0 to 6. A higher value indicates stricter regulation and therefor more job security. Of these countries, people can most easily be fired in the US, whereas the employment protection legislation is the strictest in Turkey[6]. According to Geis et al., this indicator is a good measure for job security (Geis, Uebelmesser, & Werding, 2008).

[4] Data.worldbank.org
[5] http://www.oecd.org/employment/emp/oecdindicatorsofemploymentprotection.htm
[6] For the EPL-indicator, no data for Luxembourg is available.

In order to compare two countries, we only look at the difference between those countries. In order to do so, the value for the country of birth is subtracted from the value for the country of residence. This variable will be denoted '$EPL_{i,j}$'.

$$EPL_{i,j} = EPL_i - EPL_j \qquad \textbf{Equation 7}$$

A significant positive effect would indicate that highly educated people are, more than people with a low education, drawn to countries with a high EPL indicator, which means they care more about job security than lower educated people.

Expenditure on education

When high educated people migrate to another country, they are likely to bring a family with them. Therefore, they might be interested by the amount that is spent on education in that country. This variable gives the total public expenditure on education expressed as a percentage of the GDP. Public expenditure on education includes government spending on educational institutions, education administration, and the transfers/subsidies for private entities (students/households and other private entities).[7]

To compare a set of countries, the percentage in the country of birth is subtracted from the percentage in the country of residence. The variable will be called '$education_{i,j}$'.

$$education_{i,j} = education_i - education_j \qquad \textbf{Equation 8}$$

A significant positive effect would indicate that high educated persons care more about government expenditure than lower educated persons.

Expenditure on research and development

Another factor that could attract highly educated people is the expenditure on research and development (R&D)[8], since it could be an indicator for the knowledge intensity of a country (Berkhout, Smid, & Volkerink, 2010). This variable gives the expenditure on R&D as a percentage of GDP. Expenditures for R&D are current and capital expenditures (both public and private) on creative work undertaken systematically to increase knowledge, including knowledge of humanity, culture, and society, and the use of knowledge for new applications. R&D covers basic research,

[7] Data.worldbank.org
[8] For the expenditures on GDP, the data for Denmark, Greece, New Zealand, Norway and Sweden is not available

applied research, and experimental development.[9] In this case as well, the percentage in the country of birth is subtracted from the percentage in the country of residence.

$$R\&D_{i,j} = R\&D_i - R\&D_j \qquad \textbf{Equation 9}$$

The variable will be called '$R\&D_{i,j}$' and a significant positive effect would indicate that people with a high education are more drawn to countries with higher expenditures on research and development than people with a low education level.

Patent applications

The amount of patent applications is an indicator of the level of innovation in that country. The variable gives the number of patent applications in the inventor's country of residence to the European Patent Office (EPO) divided by the GDP. The relative difference between a set of countries is taken by using the following formula:

$$patents_{i,j} = \frac{patents_i - patents_j}{patents_j} \qquad \textbf{Equation 10}$$

A significant positive effect would mean that highly educated people are more inclined to migrate to countries with relatively many patent applications (and therefore perhaps a higher level of innovation) than lower educated people.

Journal articles

Another factor which indicates the level of innovation is the number of published journal articles in a country. This variable refers to the number of scientific and engineering articles published in the following fields: physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.[10] To use these data, the number of journal articles is divided by the GDP in order to get a relative number. Next, the data has been transformed to the relative difference in relatively published articles between the country of birth and the country of origin:

$$articles_{i,j} = \frac{articles_i - articles_j}{articles_j} \qquad \textbf{Equation 11}$$

A significant positive effect would show that people with a high education generally migrate more than lower educated people to countries where relatively many scientific articles are published.

[9] Data.worldbank.org
[10] Data.worldbank.org

High technology exports

A more specific indicator of innovation is the amount of high technology exports. This variable gives the export of products with high R&D intensity (such as in aerospace, computers, pharmaceuticals, scientific instruments, and electrical machinery) as a percentage of the total of manufactured exports.[11] The percentage in the country of birth is subtracted from the percentage in the country of residence, to be able to see if the difference has an effect. This variable will be denoted as 'hightech$_{i,j}$'.

$$hightech_{i,j} = hightech_i - hightech_j \qquad \textbf{Equation 12}$$

A significant positive effect would again indicate that highly educated people are more than lower educated people drawn to countries where high R&D intensive exports are a large portion of the total manufactured exports of that country.

Language

The existence of language barriers can also influence migration flows. In order to take this into account, a variable has been created to indicate whether the same language is spoken in a set of countries[12]. If multiple languages are spoken in a country, the two main most important languages are taken into account. When the variable, named 'language$_{i,j}$', is equal to (0), this means there is a language barrier. When it is equal to (1), the same language is spoken and there is no language barrier.

$$language_{i,j} = 0 \; for \; language_{1,2;i} \neq language_{1,2;j} \qquad \textbf{Equation 13.1}$$

$$language_{i,j} = 1 \; for \; language_{1,2;i} = language_{1,2;j} \qquad \textbf{Equation 13.2}$$

One would expect more people to migrate to countries where the same language is spoken as in their country of birth, since this would benefit their integration in the country. It is however unclear what the effect is on the part of high educated migrants in the total number of migrants from one country to another. Perhaps language is more important to lower educated people, which would lead to a decreasing share of high educated migrants when there is no language barrier.

[11] Data.worldbank.org
[12] http://www.internetworldstats.com/languages.htm

Distance

Another interesting variable could be the distance between countries. In this research, the great circle distance between the capital cities of two countries in kilometers, denoted 'distance$_{i,j}$', is used[13].

Equation 14

$$distance_{i,j} = great\ circle\ distance\ between\ the\ capital\ cities\ of\ countries\ i\ and\ j$$

This could influence a migrant's decision to move to one country instead of the other. But just as in the case of language, distance could possibly be more important to lower educated migrants than higher educated migrants, which would cause the share of high educated migrants to increase as distance decreases.

Income share of the richest part of the population

This variable gives the share of the total income that is earned by the richest part of the population[14]. The highly educated are most likely to become rich. Therefore, it might play a role if they are rewarded more in a different country than their country of birth. In this case, we look at the income share of the richest 10% of the population[15]. The data is transformed to make it is possible to compare a set of countries:

$$incomeshare_{i,j} = \frac{incomeshare_i - incomeshare_j}{incomeshare_j} \qquad \textbf{Equation 15}$$

If the sign of this variable is positive, this indicates that highly educated people are drawn to countries where the richest people earn the most.

Gini index

As discussed in the literature section, Liebig and Sousa-Poza address the importance of income equality in explaining migration flows (Liebig & Sousa-Poza, 2004). The Gini index is in indicator for the (in)equality of income distribution in a country. It measures the extent to which the distribution of income deviates from a perfectly equal distribution. A Gini index of 0 represents perfect equality,

[13] http://privatewww.essex.ac.uk/~ksg/data-5.html
[14] For the income share of the richest part of the population, the data for Australia, Czech Republic, Denmark, France, Japan, Netherlands, New Zealand, Portugal, Slovak Republic, Turkey and the United Kingdom is not available.
[15] Data.worldbank.org

while an index of 100 implies perfect inequality.[16] In order to compare a set of countries, the Gini-index of the country of birth is subtracted from the Gini-index of the country of residence. This variable is denoted '$gini_{i,j}$'.

$$gini_{i,j} = gini_i - gini_j \qquad \textbf{Equation 16}$$

A significant positive effect would indicate people with a high education are drawn to countries with higher income inequality than their country of birth.

Taxes on income, profits & capital gain

This variable gives the tax revenue on income, profits and capital gain as a percentage of the GDP of a country[17]. According to Geis et al., high taxes are regarded as negative by high educated people (Geis, Uebelmesser, & Werding, 2008). To be able to compare two countries, this percentage in the country of birth is subtracted from the country of residence.

$$incometax_{i,j} = incometax_i - incometax_j \qquad \textbf{Equation 17}$$

This variable, named '$incometax_{i,j}$', could tell us whether the high educated care more about the amount of personal taxes than the lower educated.

Total tax revenue

Another way to address the tax-issue is to look at the total tax revenue per capita in current USD[18]. This gives us an idea on the average amount of (direct and indirect) taxes a resident of that country would need to pay. To compare the country of birth and the country of origin, a calculation for the relative difference has been made:

$$totaltax_{i,j} = \frac{totaltax_i - totaltax_j}{totaltax_j} \qquad \textbf{Equation 18}$$

This variable could tell us whether people with a high education care more about the tax revenue than lower educated people. If this variable has a significant negative effect, high educated people prefer to live in countries with lower taxes.

[16] Data.worldbank.org
[17] Stats.oecd.org
[18] Stats.oecd.org

3.3 Methodology

In this section, the methodology of the research is explained. In order to discover which factors influence the migration of high educated people, a regression is run on the dependent variable and the independent variables. The data is processed as a panel data set. Normally this would mean that the data has both a cross-sectional and a time-series component, but in this case it has two cross-sectional components: the country of residence and the country of birth. The panel data set has fixed effects and is controlled for heteroskedasticity.

The process is as follows. First, a very simple regression is done, with just the factors of GDP, unemployment and an indicator of job security (the EPL-indicator):

$$\%\, HEM_{i,j} = b_0 + b_1 GDP_{i,j} + b_2 unemployment_{i,j} + b_3 EPL_{i,j} \qquad \textbf{Regression 1}$$

Since these factors are the most basic indicators, they are used as a base for the other regressions. This gives a starting point to check for additional explanatory value for the other variables. The following regressions each add a group of variables to this basic regression, sorted by topic: government expenditure, knowledge intensity, other barriers, income (in)equality and taxes.

The second regression adds some specific variables of government expenditure, namely expenditure on education and on research and development:

Regression 2

$$\%\, HEM_{i,j} = b_0 + b_1 GDP_{i,j} + b_2 unemployment\%_{i,j} + b_3 EPL_{i,j} + b_4 education_{i,j} + b_5 R\&D_{i,j}$$

Another category added is the knowledge intensity of a country. In the next regression, the patent applications, the journal articles published and the percentage high technology exports are added. These variables constitute the indicators for the knowledge intensity in a country. Countries with many patent applications, many journal articles published and a high percentage high technology exports are presumably countries with a high innovation level and where many scientific activities are going on.

Regression 3

$$\%\, HEM_{i,j} = b_0 + b_1 GDP_{i,j} + b_2 unemployment\%_{i,j} + b_3 EPL_{i,j} + b_4 education_{i,j} + b_5 R\&D_{i,j} + b_6 patents_{i,j} + b_7 articles_{i,j} + b_8 hightech_{i,j}$$

Next, the factors that can't be influenced at all are considered. The following regression again looks at GDP, unemployment and EPL-indicator, but adds the barriers of language and distance:

Regression 4

$$\%\, HEM_{i,j} = b_0 + b_1 GDP_{i,j} + b_2 unemployment\%_{i,j} + b_3 EPL_{i,j} + b_4 language_{i,j} + b_5 distance_{i,j}$$

The distribution of income might also influence the migration of highly educated people. Regression 5 therefore adds factors of income (in)equality to regression 1. The Gini-index and the income share earned by the richest 10% of the population are indicators of the way income is distributed:

Regression 5

$$\%\, HEM_{i,j} = b_0 + b_1 GDP_{i,j} + b_2 unemployment\%_{i,j} + b_3 EPL_{i,j} + b_4 incomeshare_{i,j} + b_5 gini_{i,j}$$

Finally, the level of taxes might influence the migration of highly educated people. The following regression considers the variables on taxes:

Regression 6

$$\%\, HEM_{i,j} = b_0 + b_1 GDP_{i,j} + b_2 unemployment\%_{i,j} + b_3 EPL_{i,j} + b_4 incometax_{i,j} + b_5 totaltax_{i,j}$$

The last regression combines all the variables that together are most capable of explaining the y-variable. It is derived by creating an equation that contains all the variables and then crossing out the variables that are not significant. This leaves the following regression:

Regression 7

$$\%\, HEM_{i,j} = b_0 + b_1 GDP_{i,j} + b_2 EPL_{i,j} + b_3 education_{i,j} + b_4 R\&D_{i,j} + b_5 articles_{i,j} + b_6 distance_{i,j} + b_7 incomeshare_{i,j} + b_8 gini_{i,j}$$

3.4 Conclusion

For this research, data from the OECD DIOC is used for the dependent variable and several other sources are used for the independent variables. The raw data for the independent variables can be found in table B1 in Appendix B. This section also shows how the data has been adapted in order to make it useful for the research, either by taking the absolute difference or the percentage improvement. The statistics of the relative data can be found in Appendix B, table B2. The relative data is used to run the panel regressions. The first regression is done to test the significance of the controlling variables. Next, other variables are added in groups, sorted by the categories government expenditure, knowledge intensity, other barriers, income equality and taxes. Finally a large regression is constructed of the most significant variables for purposes of gaining the highest explanatory power.

4. Results

In this chapter, the results of the regressions will be explained. The outcomes of the regressions 1 to 7 are given in table 4.1 and the original Stata output can be found in Appendix A. The table gives the coefficients of the variables and their significance level. Also, the number of observations and the R-square of each regression is given.

General indicators and risk aversion

Based on regression 1, GDP, unemployment and the EPL indicator all have a significant effect on the share of highly educated migrants in the total number of migrants. GDP has a negative effect. This indicates that when the average income in a country is higher than the average income in the country of birth, the share of migrants with a high education in the total number of migrants to that country becomes lower. This might mean that people with a lower education find the average income more important than people with a higher education. Unemployment and the EPL-indicator have a positive effect. As the unemployment rate in one country is higher than in another, more people with a high education than with a lower education will move to the country with the higher employment rate. This could indicate that people with a high education find unemployment in the destination country less negative than people with a lower education. When job security is better in one country than another, more people with a high education relative to the total number of migrants will migrate to the country with the better job security. This indicates that people with a high education value job security relatively more than people with a lower education.

Specific government expenditure

In the second regression, the effect of GDP on the share of high educated migrants is again negative, but not significant. It does not seem to have any additional explanatory power anymore. The other variables are significant. Unemployment and the EPL-indicator are still positive and their value is more or less the same. From this regression, we can see that the expenditures on both education and R&D have additional explanatory power. The effect of education is positive, which means that when a larger portion of GDP is spent on education in a country than in the country of birth, relatively more people with a high education compared to the total number of migrants will migrate to that country. This could indicate that people with a high education value the expenditures on education in a country more than people with a lower education. On the other hand, the effect of expenditures on R&D is negative. When a larger portion of GDP in a country is spent on R&D than in the country of birth, relatively more people with a high education compared to the total number of migrants will migrate to that country. This would indicate that people with a high education value the expenditures on R&D less than people with a lower education.

Table 4.1

Regression	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Constant	39.462*** *(162.82)*	38.431*** *(103.45)*	39.167*** *(93.50)*	36.056*** *(48.03)*	40.620*** *(84.33)*	39.517*** *(141.23)*	33.263*** *(20.74)*
$GDP_{i,j}$	-1.461*** *(-5.59)*	-0.94 *(-0.21)*	-0.299 *(-0.48)*	-1.515*** *(-5.04)*	-2.364*** *(-6.66)*	-4.120*** *(-5.05)*	-1.875** *(-2.98)*
$Unemployment_{i,j}$	2.477*** *(4.71)*	2.907*** (4.66)	1.872*** *(2.88)*	2.281*** *(4.50)*	-0.286 *(-0.31)*	1.834*** *(3.23)*	
$EPL_{i,j}$	5.894** *(7.75)*	5.813** *(6.43)*	6.266** *(6.39)*	5.46** *(7.79)*	6.430*** *(4.54)*	6.346*** *(8.14)*	7.325*** *(4.61)*
$Education_{i,j}$		2.093*** *(2.10)*	2.198*** *(2.19)*				5.962** *(2.28)*
$R\&D_{i,j}$		-4.585*** *(-4.80)*	-4.344*** *(-4.21)*				-3.241** *(-2.97)*
$Patents_{i,j}$			0.129** *(2.35)*				
$Articles_{i,j}$			-0.129*** *(-3.33)*				-0.174*** *(-4.86)*
$Hightech_{i,j}$			-0.753* *(-1.85)*				
$Language_{i,j}$				-3.205 *(-1.42)*			
$Distance_{i,j}$				0.001*** *(4.94)*			0.002*** *(4.89)*
$Incomeshare_{i,j}$					-3.273*** *(-4.08)*		-5.741*** *(4.94)*
$Gini_{i,j}$					1.762*** *(3.06)*		4.152*** *(4.48)*
$Incometax_{i,j}$						0.03 *(0.03)*	
$Totaltax_{i,j}$						1.890*** *(3.41)*	
R-squared (within)	20.1%	25.2%	28.1%	25.1%	27.08%	21.6%	42.0%
No. observations	652	426	426	652	233	652	151

*T-values are reported in parentheses. *, **, *** indicates significance at the 90%, 95% and 99% level, respectively*

Knowledge intensity

In the third regression, GDP is again not significant, but the other factors are. The coefficients for unemployment, the EPL-indicator, expenditure on education and R&D are quite the same as they were. The three new variables are significant. The relative number of patents filed has a significant and positive effect on the share of high educated immigrants. This could indicate that when a country has a relatively large amount of patents filed, a larger share of the migrants to that country will be highly educated. The coefficients for the number of journal articles and the amount of high technology exports are on the other hand negative. If the relative (to GDP) number of scientific journal articles published is higher, then the share of migrants with a high education will be lower. This indicates that the number of journal articles published does not stimulate knowledge migration a country. The relative amount of high technology exports of a country also has a negative effect on the share of immigrants in that country with a high education. The coefficient for high technology exports is however only significant at a 10% level, due to which its effect could be questioned.

Other barriers

The results of regression 4 show that if the same language is spoken in two countries (language = 1), the share of migrants with a high education of the total migrants between those countries will be lower. This could mean that people with a low education are more drawn to countries with the same language than people with a high education, i.e. a language barrier is more difficult for them than for high educated people to overcome. However, the coefficient is not significant, so it is dangerous to draw conclusions from it. The distance between the countries however does have a significant effect. If the distance between two countries is greater, the share of high educated people in the total number of migrants will be larger. This could indicate that people with a high education are less held back by distance to migrate than people with a low education.

Income equality

In regression 5, where the factors of income equality are added, unemployment is not significant and does not have any additional explanatory power to the other variables in this regression. The new variables of the income share of the richest 10% of the population and the Gini index both have a significant effect. If the richest 10% of the population in one country earn a relatively larger share of the income than in another country, the share of migrants that have a high education in the total number of migrants to that country with a higher income share will be lower. This could indicate that highly educated people prefer equality over the prospect of earning much. On the other hand, it could mean they generally do not expect to become part in the richest 10% of the population when

they migrate. The effect of the Gini-index points us to the second conclusion. When the Gini-index in one country is higher (i.e. more inequality), the share of highly educated migrants in the total number of migrants to that country will be higher. This indicates that highly educated people prefer inequality more than or care less about equality than people with a lower education.

Taxes

The results for regression 6 show that relative income tax does not have any additional explanatory power, since its coefficient is not significant. The coefficient for the total amount of taxes is positive and significant. This means that when the total taxes are higher in one country than another, the share of migrants with a high education in the total number of migrants to the country with the high taxes will be higher.

Final equation

In the final equation, all variables have a significant effect on the explanatory variable. The signs of these variables have all remained the same. It is therefore safe to say that these eight variables have significant explanatory power on the share of highly educated migrants in the total number of migrants.

5. Discussion

The net import of high skilled migrants is considered to be positive for the economy of a country. Therefore, this research aimed to find the factors that influence the migration of highly educated people. In this chapter, the implications of the research will be discussed. Next, the results will be applied to the case of the Netherlands. The chapter will end with a section on limitations and suggestions for further research.

5.1 Policy implications

If the aim of a country is to attract relatively more migrants with a high education, the most general policy would be to make the factors that pull migrants more attractive for highly educated people than for lower educated people. This research has given some ideas on what to focus on.

Government policy

From the research, three factors can be derived that can and should be influenced by the government in order to attract highly educated people: job security, education and income equality.

The EPL-indicator had a significant and positive effect on the share of high educated migrants. This means that when the legislation on employment protection, a proxy for job security, is better, relatively more highly educated in the total number of migrants will be drawn to that country. Apparently, highly educated people value job security. This might be because they migrate for job opportunities more often than lower educated people. Governments could therefore improve job security through legislation in order to be an attractive country for high educated migrants.

Highly educated migrants also seem to value the governmental expenditures on education. An explanation could be that highly educated migrants want their children to grow up with the opportunities of good education. Or perhaps a country where a lot is spent on education has the right environment, with more people with a high education, for highly educated people to live in. Either way, it can be useful for governments to invest in education in order to attract highly educated migrants.

Finally, highly educated people prefer more income inequality. Taking the Gini-index as a proxy, the share of migrants with a high education increases as the division of income is more unequal. Generally people with a higher education end up in a higher income scale, which can explain why they would like the income in higher income scales to be even higher. However, the share of high educated migrants decreases when the richest 10% earn a larger portion of the income. This suggests they do not expect to reach the highest income scale. Governments should therefore seek

to reward jobs that require a higher education level, but should also not be hesitant to burden the richest people in the county.

Of course, attracting highly educated immigrants can't be seen as the only or most important objective of the measures discussed in this section. Yet it can prove to be an additional argument to implement certain policy and should be taken into account in the decision making process.

Unexpected outcomes

The research also gave a few unexpected outcomes, namely the results for expenditures on R&D, the amount of published scientific articles and the total taxes per capita.

The total (public and private) expenditures on R&D as a percentage of GDP were used as a proxy for the knowledge intensity of a country (Berkhout, Smid, & Volkerink, 2010). However, it appeared to have a negative effect on the share of highly educated migrants in the total number of migrants. An explanation could be that expenditures on R&D only attract specific types of migrants, e.g. researchers and technically skilled people, but not necessarily highly educated migrants in general.

The relative number of scientific journal articles published also appeared to have a negative effect. The problem with this factor could lie in the fact that the number of journal articles published was divided by the GDP of a country. This could prove to be a distorting factor in the comparability of the data between the countries.

The total taxes per capita in a country would be expected to have a negative effect on the immigration of highly educated people. However, the results came out positive. This could be linked to the idea that higher taxes generally implicate better facilities, which in turn can attract high educated migrants.

Factors not influenced by government policy

Of course there are also factor that can´t or shouldn't be influenced by the government. These include the GDP per capita, unemployment and distance.

The GDP per capita is an indicator of the wealth of a country and can't be influenced directly by the government. Even if it could, a higher share of highly educated migrants is attracted when the GDP per capita is lower and it is undesirable to lower the GDP per capita.

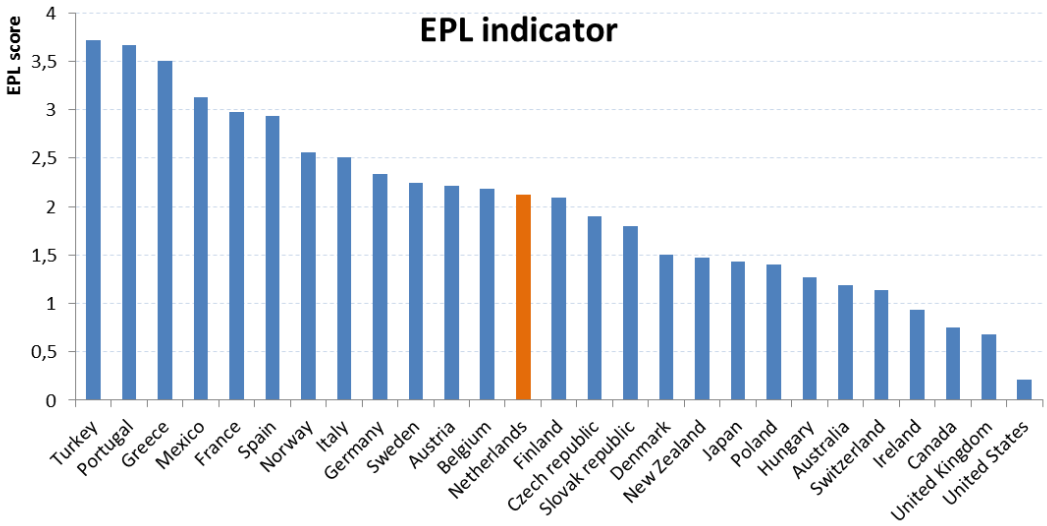
The unemployment rate should also not be influenced by the government in order to attract highly educated immigrants, since a higher unemployment rate yields a higher share of highly educated migrants. Increasing unemployment could however have other (negative) consequences, which are

beyond the scope of this research. These consequences are presumably more important than attracting highly educated immigrants. Furthermore, it does not mean that highly educated people prefer a higher unemployment rate, but it could also just indicate that lower educated people are generally more discouraged by it.

Another factor that can't be influenced by the government is the distance between countries. However, it should be taken into consideration when assessing the competitive position of the country in its ability to attract highly educated immigrants. This should prove useful in deciding how much effort should be put in the improvement of this competitive position.

5.2 The Netherlands

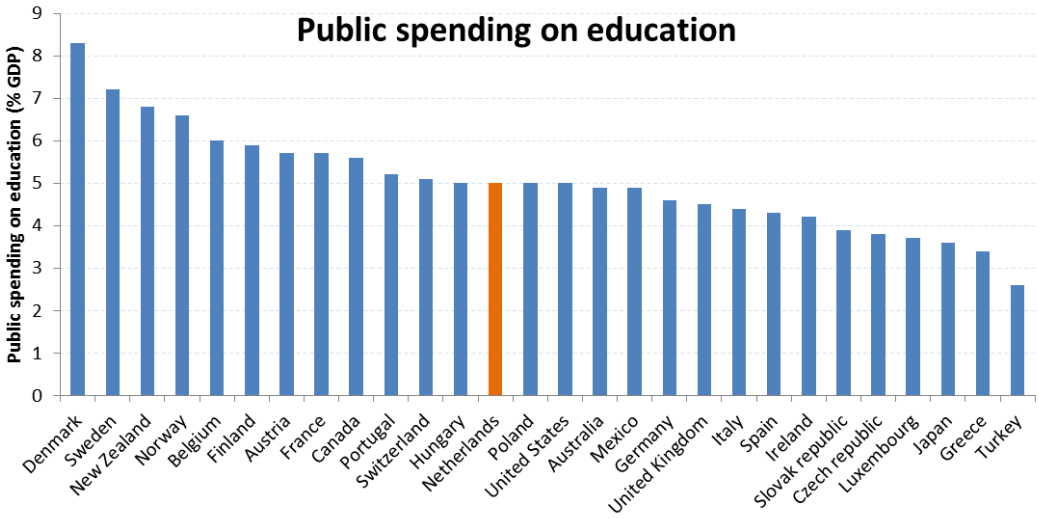
The policy implications described above can hold for any of the OECD countries. So what does this mean for the Netherlands? As seen from graph 1.1 in the introduction, the Netherlands are not escalating at attracting highly educated migrants.



Graph 5.1 – EPL indicator for OECD countries
This graph shows the indicator for Employment Protection Legislation. The values are on a scale of 0 to 6, where a higher value indicates stricter employment protection legislation and therefore more job security.
Source: stats.oecd.org

As can be seen in graph 5.1 and 5.2, the Netherlands are doing average when it comes to the employment protection legislation and public spending on education. Perhaps the Dutch government could expand its expenditures on education in order to create a more desired environment for foreign highly educated people. They could also make the employment protection legislation more strict, in order to improve job security. This can however be regarded as negative by the private sector, since it will make companies more careful and discourage them to hire people for the long term. The BCG report also suggests adapting the Dutch immigration policy (Kurstjens, Maas,

& Steffens, 2012). And besides attracting foreign talent, the Netherlands should also make more of an effort to keep them here (Kurstjens, Maas, & Steffens, 2012).



Graph 5.2 – Public spending on education in OECD countries
This graph shows the public expenditures on education as a percentage of the GDP for the OECD countries studied in this thesis.
Source: stats.oecd.org

5.3 Limitations and suggestions for further research

Despite the significance of the results, the interpretations of the research should be taken with caution due to the limitations of the research.

First of all, the research only looks at stock quantities, namely the countries of birth of the inhabitants of the countries concerned, and not at migration flows. This poses problems when linking those stock quantities to other variables. Because the data consists of stock quantities, the other data all had to be taken from the same year as the stock quantities of the dependent variable, but this does not necessarily truly explain the reasons for the foreign born to migrate in the years before. This problem is caused by the fact that the data for the dependent variable was only available for one year. It makes it impossible to include networking effects in the research and makes it difficult to find true causal relations.

Furthermore, this research only looks at the share of high educated migrants, not at the total number of migrants. This poses problems, because the share of high educated migrants can increase, even when the total number of highly educated migrants decreases. An increase in the share of high educated migrants is therefore not necessarily an improvement. This could be overcome by a similar research which includes the total number of migrants as a dependent variable.

Another limitation is caused by the fact that for some variables, the data for some countries is not available. For example, for the variable of the income share of the richest 10% of the population, the data for up to eleven out of the twenty-eight countries are missing. Some data is also not available for several other variables, as is given in the chapter on 'data'. Therefore, care should be taken in interpreting the results for these specific variables.

Finally, this research only looks at the migration between 28 of the OECD countries. It does not look at the migrants in those countries that were born in other than those countries and at the migrants in other countries from those countries. One should therefore be careful in applying these results to all the countries in the world.

Further research could overcome these limitations if measurements were taken of the flow of migrants all over the world, each year. That way, a time series could be done in order to better investigate the cause-and-effect relations. However, this requires intensive data collection by all countries, which is almost impossible to accomplish. A start would therefore be to at least look at all the countries for which data is available and not just at these 28 x 27 combinations of countries. The research could also be improved by adding a dependent variable which gives the total number of migrants, so that the actual impact of the independent variables on migration can be estimated.

6. Conclusion

In this thesis, the factors that influence knowledge migration between developed countries have been investigated. For this purpose, a panel dataset was used. The dataset is composed out of the number of immigrants living in one of the 28 selected OECD countries in the year 2000 and originating from one of the selected OECD. Next, the share of those immigrants with a high education level (ISCED 5/6) is taken. The resulting figure is used as the dependent variable, having a different observation for each combination of country of birth and country of residence. Factors which are expected to influence the migration of the highly educated are used as independent variables, taken relatively between the country of birth and the country of origin, to determine which ones have a (economically and statistically) significant influence. The factors include variables on average income, unemployment, job security, income equality, government expenditures, knowledge intensity, taxes and other barriers.

The results showed that a stricter legislation on employment protection would lead a higher share of high educated migrants in the total number of migrants. This could indicate that highly educated people are risk averse and want risks to be minimal. It could also imply that highly educated people are more likely to migrate for job reasons and therefore value employment protection legislation more than lower educated people. Highly educated people are also drawn to countries where a larger percentage of the GDP is spent on education than in their country of birth and to countries where the income inequality is greater.

The key idea in this research is that in order to attract highly educated immigrants, the benefits for this particular group should be high. This includes higher incomes for the sectors they will work in and special facilities, including job security but also immigration policy, to make immigration to a country more attractive. By improving on these factors, a developed country will be able to compete for international talent. It remains important to keep in mind that every person, thus every migrant, is different and subject to different push and pull factors: policy should be adapted to the different types of migrants (Mahroum, 2000).

Finally, it should be noted that this research is limited due to the type of data that is used and the number of countries that are subject. The results should therefore be interpreted with care.

Bibliography

Berkhout, E., Smid, T., & Volkerink, M. (2010). *Wat beweegt kennismigranten?* Amsterdam: SEO Economisch Onderzoek.

Bowles, S. (1970, November). Migration as Investment: Empirical Tests of the Human Investment Approach to Geographical Mobility. *The Review of Economics and Statistics, 52*(4), 356-362.

Chiswick, B. R. (1999). Are Immigrants Favorably Self-Selected? *The American Economic Review, 89*(2), 181-185.

Geis, W., Uebelmesser, S., & Werding, M. (2008). *How do migrants choose their destination country? An analysis of institutional determinants.* Munich: CESifo Group.

Gonzalez, C. R., Mesanza, R. B., & Mariel, P. (2011). The determinants of international student mobility flows: an empirical study on the Erasmus programme. *High Educ, 62*, 413-430.

Heitmueller, A. (2005). Unemployment benefits, risk aversion, and migration incentives. *Journal of Population Economics*, 93-112.

Kehm, B. M. (2005). The Contribution of International Student Mobility to Human Development and Global Understanding. *US-China Education Review, 2*(1), 18-24.

Kurstjens, H., Maas, J. W., & Steffens, T. (2012). *NL 2030: Contouren van een nieuw Nederlands verdienmodel.* Amsterdam: The Boston Consulting Group.

Liebig, T., & Sousa-Poza, A. (2004). Migration, Self-Selection and Income Inequality: An International Analysis. *KYKLOS, 57*, 125-146.

Mahroum, S. (2000). Highly skilled globetrotters: mapping the international migration of human capital. *R&D Management, 30*(1), 23-32.

OECD. (2009). Specifities in managing highly-skilled migration. *International Migration Outlook: SOPEMI 2009*, 162-176.

The World Bank. (n.d.). Retrieved from http://data.worldbank.org

Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2008). Education and entrepreneurship selection and performance: a review of the empirical literature. *Journal of Economic Surveys, 22*(5), 795-841.

Appendix A

Stata output for:

$$\%\, HED_{i,j} = b_0 + b_1 GDP + b_2 unemployment\% + b_3 EPL$$ **Regression 1**

```
Fixed-effects (within) regression               Number of obs      =       652
Group variable: country                         Number of groups   =        27

R-sq:  within  = 0.2058                         Obs per group: min =        10
       between = 0.0228                                        avg =      24.1
       overall = 0.0448                                        max =        26

                                                F(3,26)            =     39.45
corr(u_i, Xb)  = -0.4612                        Prob > F           =    0.0000

                                  (Std. Err. adjusted for 27 clusters in country)
------------------------------------------------------------------------------
                             Robust
higheducatedi~2564 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
     relativegdppc |  -1.461273    .261176    -5.59   0.000    -1.998128    -.924418
unemploymentchange |   2.476516   .5259121     4.71   0.000     1.395488    3.557543
            epldif |   5.893984   .7602929     7.75   0.000      4.33118    7.456789
             _cons |   39.46152   .2423564   162.82   0.000     38.96335    39.95969
-------------------+----------------------------------------------------------------
           sigma_u |  14.059438
           sigma_e |  13.714891
               rho |  .51240332   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Regression 2

$$\%\, HED_{i,j} = b_0 + b_1 GDP + b_2 unemployment\% + b_3 EPL + b_4 education + b_5 R\&D$$

```
Fixed-effects (within) regression               Number of obs      =       426
Group variable: country                         Number of groups   =        22

R-sq:  within  = 0.2529                         Obs per group: min =         9
       between = 0.0224                                        avg =      19.4
       overall = 0.0661                                        max =        21

                                                F(5,21)            =     18.79
corr(u_i, Xb)  = -0.4851                        Prob > F           =    0.0000

                                  (Std. Err. adjusted for 22 clusters in country)
------------------------------------------------------------------------------
                             Robust
higheducatedi~2564 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
     relativegdppc |  -.0941887   .4437874    -0.21   0.834    -1.017095    .8287177
unemploymentchange |   2.907242   .6245159     4.66   0.000     1.608491    4.205994
            epldif |   5.813497   .9040699     6.43   0.000     3.933381    7.693613
         eduexpdif |   2.092691   .9960888     2.10   0.048     .0212111    4.164171
             rddif |  -4.585063   .9552128    -4.80   0.000    -6.571536   -2.598589
             _cons |   38.43072   .3714839   103.45   0.000     37.65818    39.20327
-------------------+----------------------------------------------------------------
           sigma_u |  15.929489
           sigma_e |  14.497716
               rho |  .54695171   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Regression 3

$$\%\,HED_{i,j} = b_0 + b_1 GDP + b_2 unemployment\% + b_3 EPL + b_4 education + b_5 R\&D + b_6 patents + b_7 articles + b_8 hightech$$

```
Fixed-effects (within) regression               Number of obs      =       426
Group variable: country                         Number of groups   =        22

R-sq:  within  = 0.2808                         Obs per group: min =         9
       between = 0.0298                                        avg =      19.4
       overall = 0.0626                                        max =        21

                                                F(8,21)            =     23.85
corr(u_i, Xb)  = -0.5259                        Prob > F           =    0.0000

                                    (Std. Err. adjusted for 22 clusters in country)
------------------------------------------------------------------------------
                             Robust
higheducatedimmigr~2564      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------
         relativegdppc   -.2986981   .6193225    -0.48   0.635    -1.58665    .9892537
     unemploymentchange    1.872162   .6496404     2.88   0.009     .5211613    3.223163
                epldif    6.266434   .9811717     6.39   0.000     4.225976    8.306892
             eduexpdif    2.197839   1.004396     2.19   0.040     .1090836    4.286594
                rddif   -4.343835   1.032559    -4.21   0.000    -6.49116   -2.196511
          relativeepo2    .1286859   .0547123     2.35   0.029     .0149054    .2424663
        relativejournal   -.1289034   .0387436    -3.33   0.003    -.2094752   -.0483316
relativehightechexports   -.7533715   .4067292    -1.85   0.078    -1.599211    .0924681
                _cons    39.16705   .4189065    93.50   0.000     38.29588    40.03821
-------------------------+----------------------------------------------------
              sigma_u  16.607323
              sigma_e  14.277961
                  rho  .57499322   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Regression 4

$$\%\,HED_{i,j} = b_0 + b_1 GDP + b_2 unemployment\% + b_3 EPL + b_4 language + b_5 distance$$

```
Fixed-effects (within) regression               Number of obs      =       652
Group variable: country                         Number of groups   =        27

R-sq:  within  = 0.2514                         Obs per group: min =        10
       between = 0.0029                                        avg =      24.1
       overall = 0.0800                                        max =        26

                                                F(5,26)            =     26.29
corr(u_i, Xb)  = -0.4007                        Prob > F           =    0.0000

                                    (Std. Err. adjusted for 27 clusters in country)
------------------------------------------------------------------------------
                             Robust
higheducatedimmi~2564        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------
         relativegdppc   -1.515336    .300887    -5.04   0.000    -2.133818   -.8968541
     unemploymentchange    2.281182   .5073227     4.50   0.000     1.238366    3.323999
                epldif    5.446392   .6987537     7.79   0.000     4.010084    6.882701
          samelanguage   -3.205302    2.25186    -1.42   0.167    -7.834067    1.423464
distancetofromx1000km    .0007337   .0001486     4.94   0.000     .0004282    .0010393
                _cons    36.05615   .7506902    48.03   0.000     34.51309    37.59922
-------------------------+----------------------------------------------------
              sigma_u  13.331555
              sigma_e  13.336485
                  rho  .49981513   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Regression 5

$$\%\,HED_{i,j} = b_0 + b_1 GDP + b_2 unemployment\% + b_3 EPL + b_4 incomeshare + b_5 gini$$

```
Fixed-effects (within) regression               Number of obs      =       233
Group variable: country                         Number of groups   =        16

R-sq:  within  = 0.2708                         Obs per group: min =         8
       between = 0.4225                                        avg =      14.6
       overall = 0.0000                                        max =        15

                                                F(5,15)            =     15.37
corr(u_i, Xb)  = -0.6348                        Prob > F           =    0.0000

                                    (Std. Err. adjusted for 16 clusters in country)
------------------------------------------------------------------------------
             |               Robust
higheducatedi~2564 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     relativegdppc |  -2.364271   .3547547    -6.66   0.000    -3.120413   -1.608129
unemploymentchange |  -.2862113    .938048    -0.31   0.764    -2.285613    1.713191
           epldif |   6.429765   1.416413     4.54   0.000     3.410751    9.448778
  incomeshare10dif |  -3.273101   .8028226    -4.08   0.001    -4.984276   -1.561925
          ginidif |    1.76202   .5752877     3.06   0.008     .5358238    2.988217
            _cons |   40.62024   .4817091    84.33   0.000      39.5935    41.64698
-------------+----------------------------------------------------------------
          sigma_u |  15.896085
          sigma_e |  11.480553
              rho |  .65719896   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Regression 6

$$\%\,HED_{i,j} = b_0 + b_1 GDP + b_2 unemployment\% + b_3 EPL + b_4 incometax + b_5 totaltax$$

```
Fixed-effects (within) regression               Number of obs      =       652
Group variable: country                         Number of groups   =        27

R-sq:  within  = 0.2158                         Obs per group: min =        10
       between = 0.0341                                        avg =      24.1
       overall = 0.0383                                        max =        26

                                                F(5,26)            =     31.20
corr(u_i, Xb)  = -0.5043                        Prob > F           =    0.0000

                                    (Std. Err. adjusted for 27 clusters in country)
------------------------------------------------------------------------------
             |               Robust
higheducatedi~2564 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     relativegdppc |  -4.119879   .8152139    -5.05   0.000    -5.795575   -2.444182
unemploymentchange |   1.834296   .5672828     3.23   0.003      .668229    3.000362
           epldif |   6.345806   .7793935     8.14   0.000      4.74374    7.947872
      incometaxdif |   .0031172   .1239485     0.03   0.980    -.2516626     .257897
  totaltaxpcchange |   1.890135   .5547437     3.41   0.002     .7498425    3.030426
            _cons |   39.51669   .2798114   141.23   0.000     38.94153    40.09185
-------------+----------------------------------------------------------------
          sigma_u |  14.656174
          sigma_e |  13.650295
              rho |  .53549046   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Regression 7

$$\%\, HED_{i,j} = b_0 + b_1 GDP + b_2 EPL + b_3 education + b_4 R\&D + b_5 articles + b_6 distance + b_7 incomeshare + b_8 gini$$

```
Fixed-effects (within) regression               Number of obs      =       151
Group variable: country                         Number of groups   =        13

R-sq:  within  = 0.4200                         Obs per group: min =         7
       between = 0.0021                                        avg =      11.6
       overall = 0.1098                                        max =        12

                                                F(8,12)            =     12.09
corr(u_i, Xb)  = -0.5038                        Prob > F           =    0.0001

                                    (Std. Err. adjusted for 13 clusters in country)
```

higheducatedimmi~2564	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
relativegdppc	-1.875117	.6296607	-2.98	0.012	-3.24703	-.5032045
epldif	7.32465	1.589479	4.61	0.001	3.861474	10.78783
eduexpdif	5.961825	2.612961	2.28	0.042	.2686721	11.65498
rddif	-3.240622	1.092462	-2.97	0.012	-5.620894	-.8603514
relativejournal	-.174387	.0359005	-4.86	0.000	-.2526074	-.0961666
distancetofromx1000km	.0023663	.0004839	4.89	0.000	.001312	.0034207
incomeshare10dif	-5.741248	1.162235	-4.94	0.000	-8.273541	-3.208955
ginidif	4.151987	.9267661	4.48	0.001	2.132737	6.171237
_cons	33.26254	1.603642	20.74	0.000	29.76851	36.75658
sigma_u	14.731855					
sigma_e	11.359756					
rho	.62711749	(fraction of variance due to u_i)				

Appendix B

Table B1

	GDP_i	$Unemployment_i$	EPL_i	$Education_i$	$R\&D_i$	$Patents_i$	$Articles_i$	$Hightech_i$	$Incomeshare_i$	$Gini_i$	$Incometax_i$	$Totaltax_i$
Australia	$21.708	6,3%	1,2	4,9%	1,6%	1029	14341	15%			17,6%	$6.491
Austria	$23.974	3,5%	2,2	5,7%	1,9%	1186	4257	15%	23,1%	29,2	12,2%	$10.322
Belgium	$22.697	6,6%	2,2	6,0%	2,0%	1320	5735	10%	28,1%	33,0	17,2%	$10.138
Canada	$23.560	6,8%	0,8	5,6%	1,9%	1697	22701	19%	24,8%	32,6	17,8%	$8.563
Czech republic	$5.725	8,8%	1,9	3,8%	1,2%	66	2481	8%			7,7%	$1.944
Denmark	$29.980	4,5%	1,5	8,3%		978	4883	21%			29,8%	$14.790
Finland	$23.530	9,7%	2,1	5,9%	3,3%	1434	4844	27%	22,6%	26,9	20,4%	$11.113
France	$21.775	10,2%	3,0	5,7%	2,2%	7307	31427	25%			11,1%	$9.974
Germany	$22.946	7,7%	2,3	4,6%	2,5%	22178	43509	19%	22,1%	28,3	11,3%	$8.599
Greece	$11.396	11,1%	3,5	3,4%		56	2976	14%	26,0%	34,3	9,4%	$3.915
Hungary	$4.543	6,4%	1,3	5,0%	0,8%	120	2358	27%	22,9%	27,3	9,5%	$1.783
Ireland	$25.630	4,3%	0,9	4,2%	1,1%	208	1581	48%	27,2%	34,3	13,1%	$7.944
Italy	$19.388	10,8%	2,5	4,4%	1,0%	4004	21409	9%	26,8%	36,0	13,9%	$8.142
Japan	$37.292	4,8%	1,4	3,6%	3,0%	22230	57101	29%			9,3%	$9.945
Luxembourg	$46.453	2,3%		3,7%	1,7%	82	40	17%	23,8%	30,8	14,1%	$18.186
Mexico	$5.817	2,6%	3,1	4,9%	0,4%	30	2971	22%	41,4%	51,9	4,6%	$1.090
Netherlands	$24.180	2,7%	2,1	5,0%	1,9%	3474	12341	36%			10,0%	$9.573
New Zealand	$13.376	6,2%	1,5	6,8%		177	2851	10%			19,9%	$4.675
Norway	$37.473	3,4%	2,6	6,6%		406	3136	17%	23,4%	25,8	19,2%	$15.987
Poland	$4.454	16,1%	1,4	5,0%	0,6%	43	5506	3%	26,1%	32,9	6,8%	$1.466
Portugal	$11.471	3,9%	3,7	5,2%	0,7%	42	1880	6%			9,2%	$3.540
Slovak republic	$5.330	18,8%	1,8	3,9%	0,6%	11	978	4%			7,0%	$1.817
Spain	$14.414	13,9%	2,9	4,3%	0,9%	805	14795	8%	26,6%	34,7	9,8%	$4.946
Sweden	$27.869	5,8%	2,2	7,2%		2301	9883	23%	22,2%	25,0	21,0%	$14.338
Switzerland	$35.639	2,7%	1,1	5,1%	2,5%	2738	8504	22%	25,9%	33,7	13,0%	$10.458
Turkey	$4.189	6,5%	3,7	2,6%	0,5%	45	3484	5%			7,1%	$1.004
United Kingdom	$25.058	5,5%	0,7	4,5%	1,8%	6091	48216	32%			14,2%	$9.120
United States	$35.082	4,0%	0,2	5,0%	2,7%	32065	192743	34%	29,9%	40,8	14,9%	$10.354
Average	*$20.891*	*7,0%*	*2,0*	*5,0%*	*1,6%*	*4004,4*	*18819*	*18,8%*	*26,0%*	*32,8*	*13,3%*	*$7.865*
Median	*$22.821*	*6,3%*	*2,1*	*5,0%*	*1,7%*	*1003,2*	*5195*	*18,0%*	*25,9%*	*32,9*	*12,6%*	*$8.581*
Std. Deviation	*$11.380*	*4,1%*	*0,9*	*1,2%*	*0,8%*	*7809,5*	*36652*	*10,8%*	*4,4%*	*6,2*	*5,5%*	*$4.663*

Table B1 – Independent variable data per country

This table gives the raw data for each independent variable per country. The source and definition of each variable is explained in section 3.2. The data for the language barrier and the distance between countries are not available, since these are relative data per definition.

Table B2

	Total immigrants (age 25-64)	Immigrants with ISCED 5/6 (age 25-64)	$\%HED_{ij}$
Average	33034	7858	40%
Median	2272	879	38%
Std. Deviation	241820	29190	18%

	GDP_{ij}	$Unemployment_{ij}$	EPL_{ij}	$Education_{ij}$	$R\&D_{ij}$	$Patents_{ij}$	$Articles_{ij}$	$Hightech_{ij}$	$Incomeshare_{ij}$	$Gini_{ij}$	$Incometax_{ij}$	$Totaltax_{ij}$
Average (abs)	114,5%	79,6%	1,0	1,3	0,9	621%	111,6%	12,0	15,8%	6,7	6,0	149,3%
Median (abs)	58,4%	53,5%	0,9	1,1	0,9	88%	43,8%	10,0	12,0%	5,4	4,9	64,5%
Std. Deviation	187,4%	116,8%	1,3	1,7	1,2	2122%	286,1%	15,2	22,7%	9,1	7,7	267,4%

Table B2 – Statistics relative data

This table gives the average, median and standard deviation for the relative data (per combination of countries i and j).The construction of the variables is explained in section 3.2.