What is the Role of Cultural Factors in the Gravity model of FDI?

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Department of Economics
Supervisor: Dr. E.M. Bosker

Name: Diederick de Ruijter
Exam Number: 297106
E-mail address: 297106dr@student.eur.nl

Acknowledgements

I greatly appreciate the help of my supervisor Dr. E.M. Bosker. When I got stuck we discussed the possibilities to make this master thesis a success. After each meeting I started again with renewed energy and new insights. I also would like to thank my colleagues who where there for me if I needed help. Last but not least I am grateful to my parents and friends who supported me throughout my study. Whenever I doubted which route to take, they advised me or had faith in my choices which eventually led to the end of my study.

Table of Contents

1 Introduction 3
2 Theoretic background for the gravity model 5
3 FDI 8
4 Gravity model and FDI 11
5 Research done so far 13
6 Explanation of the variables 14
7 Empirical section 19
7.1 The model 19
7.2 Expectations of the variables 19
7.3 Results 23
7.3.1 Dataset 1985-2006 23
7.3.2 Dataset 2000-2006 30
7.3.3 The final model 34
8 Conclusions and Summary 39
References 41

1 Introduction

Multinationals are everywhere. You can enjoy your morning coffee at Starbucks in Mumbai, fly with Air Emirates to Bangkok and finish the day with a Big Mac at McDonalds. Some people see the positive effects of those multinationals. For example, jobs will be created and there will be positive spillover for the host country. Others see the threats of multinationals, small local firms cannot compete with multinationals and may disappear. As well as the pollution that multinationals cause in the host country. What nobody can deny is that multinationals have a huge impact in today's economy. In this paper I will try to explain which variables are important for the multinationals when choosing their location. The model that I will use for this research is the gravity equation. Most people know the gravity equation from the Newton's law of universal gravitation, but this model can also be used to explain other flows. Other examples are the gravity models of trade, immigration, and foreign direct investment (FDI).

There have been various other papers that have focused on the gravity model of FDI. However, not all authors found the same variables significant. Besides home and host country Gross Domestic Product (GDP) and the distance between the countries which are the three variables that form the basis of the gravity model, I will expand upon this by including the following; the sizes of the population in the home and host country, contiguity, official and spoken languages and two different variables concerning colonization. My primary goal is to prove that cultural similarities, in my study this concerns language and colonization, have an effect on foreign direct investment.

The four variables are:

- Home country and host country have a common official language
- Home country and host country have a common spoken language
- Home country and host country ever had a colonial relationship
- Home country and host country had a colonial relationship after 1945

Besides those four variables it will be interesting to see what the effect of the variables concerning population and contiguity will be. In section 2, I will provide a theoretical background for the gravity model. Section 3 explains what FDI is and shows its increased importance in the economy. Section 4 shows the relationship between trade and FDI and thus provides a reason why the gravity model can be used for FDI. Section 5 is a section where previous papers regarding the gravity model of FDI are discussed. In section 6, I will describe the variables that I will use and the source of the data. Section 7 is the empirical section of this paper, where I will present a series of test models. As a baseline model, I will not add any fixed effects. Additionally, I will include a model with fixed effects for the home country, the host country and for each year. For both models, I will use OLS, a method to achieve robust standard errors and clustering. I will use three clusters, where I will cluster my data on the home country, the host country and the home-host country combination. From all those models, the fixed effects model and data clustered on the home country is the best model for my study. I will discuss the results from that model and compare the results with three related papers in section 7.3.3. In section 7.3.2, I will compare the roles of all the variables; comparing later years of my dataset to the earlier years, to find which variables became more/less important over time. In the last section, I will summarize my paper, draw conclusions and give suggestions for further research.

2 Theoretic background for the gravity model

In 1962, Tinbergen was one of the first economists who researched international trade flows. Since then lots of research has been done to find out if trade can be explained via the gravity model of trade. The name for this model can be derived from Newton's law of universal gravitation. The formula for Newton's law of universal gravitation is

(1) $F_{ij= A}\frac{M_i M_j}{{D_{ij}}^2}$ (Head 2003)

Where the attractive force F_{ij} between i and j is determined by the masses of i (M_i) and j (M_j) and the distance between i and j (D_{ij}).

The most basic form of the gravity model of trade looks like:

(2) $T_{ij} = A\frac{Y_i^{\beta_1} Y_j^{\beta_2}}{D_{ij}^{\beta_3}} + \varepsilon_{ij}$

Where T_{ij} represents the value of exports from country i (home) to country j (host). Y_i is the national income of country i and Y_j is the national income of country j. D_{ij} is the distance between country I and country j. ε_{ij} is the error term. And A is a constant. The similarities between formula 1 and 2 are obvious. As can be seen from (2), on average trade is higher when incomes in the home and host country are high and distance between countries is small.

The statistical way to calculate the value of the exports is to take logs of both sides. This gives:

(3) $lnT_{ij} = \beta_0 + \beta_1 lnY_1 + \beta_2 lnY_2 - \beta_3 lnD_{ij} + \varepsilon_{ij}$ (Martinez-Zarzoso 2003)

The purpose of the gravity model of trade is to find a link between the level of trade between countries and their economic size (GDP or GNP), the geographical distance between both countries and a set of dummies. In 1966, Linnemann added the population of the importing country (N_i) and the population of the exporting county (N_j) to the equation. This is the so called augmented gravity model. The equation will now look like:

(4) $lnT_{ij} = \beta_0 + \beta_1 lnY_i + \beta_2 lnY_j - \beta_3 lnD_{ij} + \beta_4 lnN_i + \beta_5 lnN_j + \varepsilon_{ij}$

The gravity model of trade is still very popular. The main reason this model is very popular is its high explanatory power, but when Tinbergen started the model there was no theoretical foundation. The empirical results showed that the model was able to predict the level of trade, but there was no theoretical background for it. The first popular theoretical model was produced by Anderson in 1979. He used the Armington assumption where products are differentiated by the country of origin. After him various other papers were written to provide a theoretical foundation for the gravity model of trade. Examples are: Bergstrand in 1990, Deardoff in 1998, Eaton & Kortum in 2002, Anderson & van Wincoop in 2003. There were a few different approaches used in these papers to provide a theoretical foundation for the gravity model of trade. For instance Eaton & Kortum used the Ricardian framework, while Deardoff used the Heckscher-Ohlin model. Although those papers used different approaches for the theoretic model, they all came to the conclusion that the model has very strong empirical power. Deardoff provided a theoretical framework where he assumed identical Cobb-Douglas preferences. So consumers will spend a fraction β of their incomes on the products of the home country, country *i*. So the income of country *i* is

(5) $Y_i = \Sigma \beta Y = \beta Y^w$

We can easily transform equation (5) into equation (6).

(6) $\beta = Y_i / Y^w$

When we assume there are no trading costs like transport costs present and, then the trade between country *i* and country *j* will be:

(7) $T_{ij} = \beta Y_j = \frac{Y_i Y_j}{Y^w}$

Since there are homothetic preferences equation (5) can be easily transformed into $T_{ij} = \beta Y_j$. When we substitute equation (6) in the equation we get equation (7).

With Cobb-Douglas preferences this equation is the same as the gravity equation without transport costs therefore distance is not a relevant factor. We can see from equation (7), the larger the income of the host and/or home country, the larger the amount of trade. When there are transport costs involved the equation will be:

(8) $T_{ij} = \frac{Y_i Y_j}{t_{ij} Y^w}$

Since distance is the most contributing factor for transport costs, this looks exactly the same as the gravity equation by Newton in equation (1). (Deardoff 1998)

3 FDI

According to the OECD, a Foreign Direct Investment is an investment done by an enterprise in one country, investing in an enterprise in another country. This investment should be for a longer period and the investing enterprise must have some influence in the management of the enterprise where they invested in. When the enterprise has 10% or more of the voting power then the enterprise is considered to have some control over the enterprise in the host country. The 10% threshold does not always tell everything about the actual role of the home enterprise. An enterprise with less than 10% of the voting power can have more influence than an enterprise with more than 10% of the voting power. Although the threshold is arbitrary and it does not always tell everything about the real influence of the enterprise it is widely accepted to achieve statistical consistency across countries (OECD, 2008).

There are two ways to become a multinational. This is either through a Greenfield investment or through mergers and acquisitions (M&A). A Greenfield investment is a type of FDI where corporations start a new corporation in the host country. This is not the case in mergers and acquisitions where the corporation in the home country takes over a corporation in the host country. That is why a Greenfield investment is better appreciated in the host country than mergers and acquisitions, because there is more added value. Most of the FDI flows are through mergers and acquisitions which was almost 80% of total FDI flows in 2005. (OECD, 2007).
There are three types of mergers and acquisitions, namely:

- Horizontal: This type is the most common type of M&A. These M&A are when the firms involved are in the same industry. Most of the time, the goal of a horizontal M&A is to increase market power. A famous example of this, is the horizontal M&A between KLM and Air France.
- Vertical: the relationship between the merging firms is the one of supplier and distributor. The main motive for a vertical M&A is to reduce transaction costs and the dependency on other firms. An example of this is when Ford

merged with other companies to ensure that the raw materials needed to build a car were on time and of high quality. (The economist 2007).

- Conglomerate: These are M&A when the companies don't operate in the same industry. The reason why these companies decide to merge is to diversify their risk. One example is the merger of The Walt Disney Company and the American Broadcasting Company.

We can see from figure 1 that the size of FDI is growing rapidly. FDI grew in the period 1970 - 2003 from $55 billion to $514 billion. FDI is not only growing, it is also becoming more important in the world economy. In figure 2, we see that FDI accounted for 0,45% of world GDP in 1970. In 2003, it already accounted for 1,53%.

Figure 1. Development of world GDP, FDI and Trade

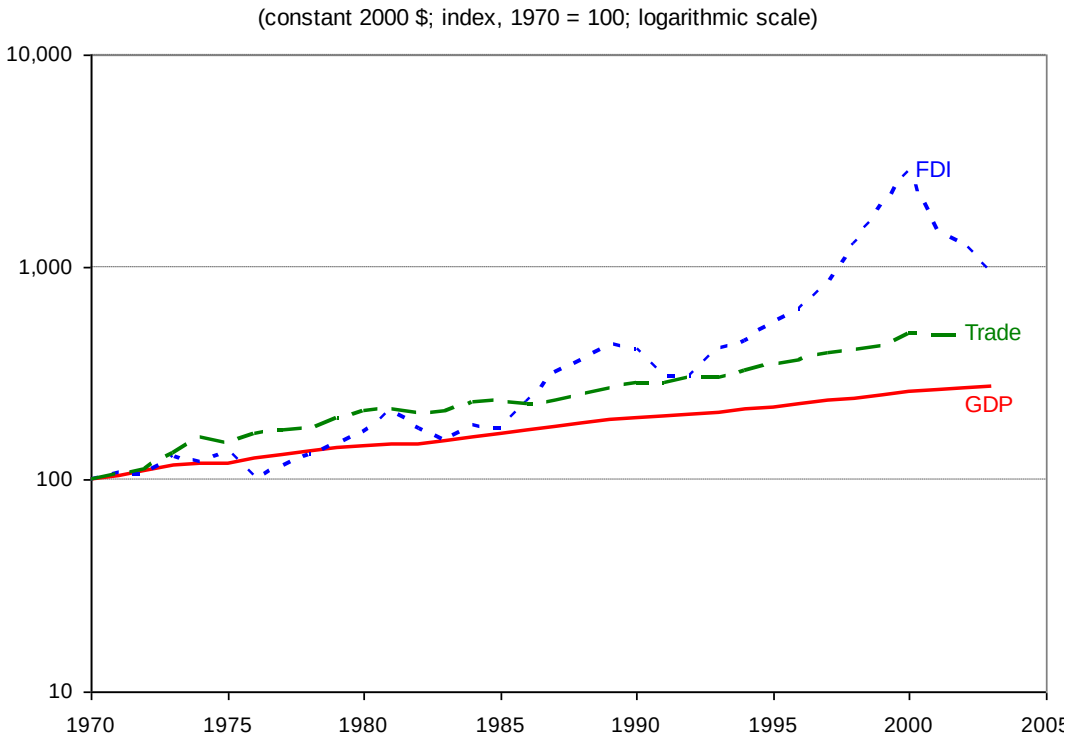
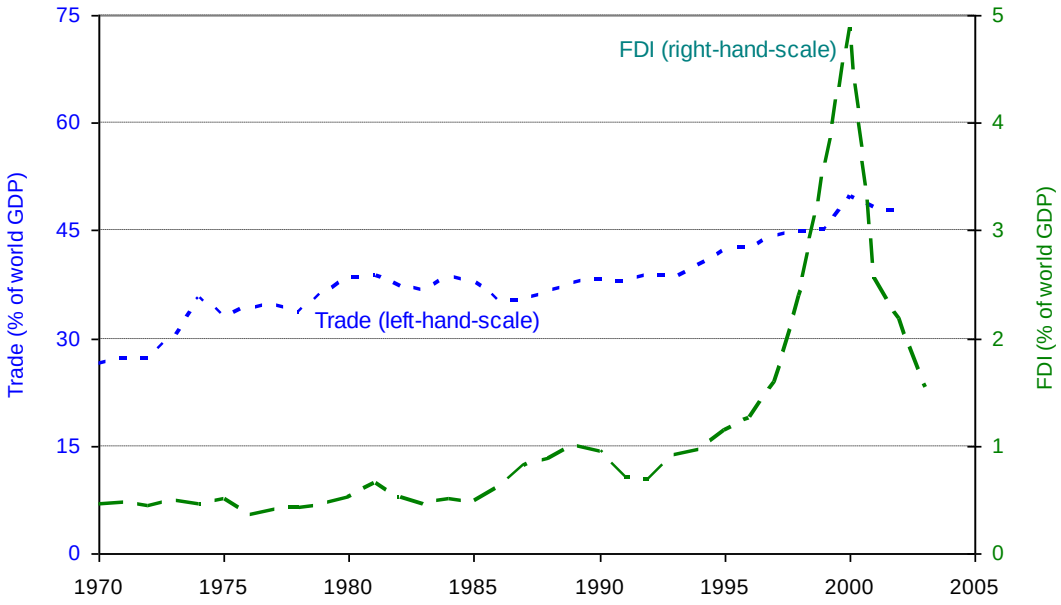


Figure 2. Trade and FDI (% of World GDP)



(Van Marrewijk, 2007)

4 Gravity model and FDI

In several papers the gravity model is used to explain FDI flows, but why is this model used? We saw that the model has high explanatory power when used to predict international trade. Helpman, Melitz and Yeaple (2003) recognize three ways how a company can sell their products abroad: the company can export their products, they can serve their customers through FDI or the company can license a foreign company to sell the product. If you look at these three channels you can see that trade and FDI are closely related. The choice between exporting the products, licensing another company or serve the customers through FDI can be made through the OLI framework. There are some additional costs in serving the market abroad through FDI. Such costs are: costs due to language differences and the cost of building and maintaining the plant in the foreign country. When the advantages of the OLI factors outweigh the additional cost described above then the company will have an incentive to use FDI. When talking about Ownership advantages you can think of some intangible assets as a good reputation or patents or a tangible asset like an advanced technique. The location advantages arise from the location of the plant in the host country. You can think of advantages like being more connected to the customers of the host country or lower transport costs. The last part of the OLI framework is the internalization advantage; this comes in play when the company licenses a foreign company to produce their product. When there is a risk that the foreign country will eventually have the knowledge to produce the product themselves then FDI becomes more interesting (Francesca di Mauro, 2000)

In a paper by Brainard (1993) he recognizes three equilibria: in the first equilibrium every company is a multinational, in the second equilibrium every company exports their products to the foreign country and the third equilibrium when there are multinationals as well as exporters. The three equilibria are derived from the following formula:

(9) $$\frac{F(w)}{C'(w,r)} - \frac{1-e^{(T+D)(1-\sigma)}}{2e^{(T+D)(1-\sigma)}}$$

Here F(w) is the fixed cost associated with the opening of a new plant. D represents the distance between the exporting country and the importing country, T represents trade barriers and C(w,r) are costs associated with the production of the input which depend on local wages w and firm specific input r. When the outcome of equation 9 is negative, then the additional costs of opening another plant are lower than the costs of exporting the products. In this case, the company will prefer FDI. When the outcome of equation 9 is positive, then the company will export the products. For one company the outcome can be positive, while the outcome for another company is negative for the same host country and the same industry. When this situation occurs, FDI and exports co-exist. With this theory, the proximity-concentration trade-off, we see that trade and FDI are substitutes for each company.

Other studies show that trade and FDI can also be complements. Helpman (1984) stated that when factor endowments are not substantial, then a country with the greater availability of capital will produce the capital-intensive goods. The country with the greater availability of labor will produce labor-intensive goods. But when factor endowments differ a lot the capital-intensive country will invest in the country with the greater labor availability. So R&D will flow to the labor-intensive country. In return, finished products will go to the country with the greater availability in capital. In this case trade and FDI are complements.

As we saw in the existing literature, it is not clear whether trade and FDI are substitutes or complements. What is clear, is that trade and FDI are closely related.

As we saw there are some similarities and some differences between trade and FDI. Since the equation for the gravity model of trade can explain most of the trade flows, it is interesting to see whether the gravity model of FDI is also a good fit to explain the FDI flows.

5 Research done so far

As said before, a lot of empirical research has been done to see if the gravity model holds for trade, but fewer papers do research to find if the gravity models holds for FDI. One of those papers was written by Markus Leibrecht and Aleksandra Riedl (2012). They tried to explain FDI flows for central and eastern European countries. The variables they included were of course the traditional variables: GDP of the host country, GDP of the home country and the distance between those countries. Other variables they included are: GDP per capita of the home country, wages, productivity, privatization revenues, political risk, inflation, import taxes, liberization of trade, a lag in the FDI level of the host country, a lag in the FDI level in the home country, index for the infrastructure, bilateral effective average tax rate and the market potential of the surrounding countries. But not all of the variables are significantly different from zero. According to Leibrecht and Riedl only the traditional three variables, GDP per capita, wages, privatization revenues, taxes on imports, bilateral effective average tax rate and the infrastructure are significant.

Another empirical paper was written by Gao (2009) he produced the empirical research done for 16 OECD countries and 5 transitioning countries. Gao added the common border dummy, the common language dummy and the GDP per capita to the standard gravity equation. Just as in the paper by Liebrecht and Riedl, the GDP per capita is significant just as the language dummy. The common border dummy is only significant when there are only European countries in the sample. The fact that the common language is highly significant in this paper caught my attention. In most papers the common language variable was significant in the gravity model of trade, but I could find some empirical papers where the common language was added to the gravity model of trade and was not significant. For example Baier & Bergstrand (2003) and Martinez-Zarzoso & Lehmann (2002) produced a paper where the dummy for the common language was not significant. Other papers agree with Gao so it will be interesting to see whether the empirical study which I will perform in the next section will support the findings of Gao or Baier & Bergstrand and Martinez-Zarzozo & Lehmann.

6 Explanation of the variables

FDI

To check whether the gravity equation for FDI holds, the data for FDI is essential. This data I have obtained from the OECD (2012). This is the Organization for Economic Co-operation and Development. In their database, they provide data for country to country FDI flows. Inward and outward FDI flows are available for all 34 OECD countries (Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan,. Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States) in the time period from 1985 to 2011. The database has data for a maximum of 311 partner countries. Unfortunately a lot of "blanks" appear in the data. This means that no data was provided by the national statistical offices in that country. As said before the data is provided by the national statistical offices in each country, and not by one worldwide office. This means that there can be a difference in the methods used to collect and the quality of the observations when the FDI data was collected. As a consequence, the outward FDI flow from country A to country B, will not always match the inward FDI flow in country B from country A. When we look at the data we see that differences exist in almost every observation and in some cases the difference is considerable. For my paper I will focus on the information provided by the home country since this are all OECD countries and I expect that the data from OECD countries will be more reliable than the data from non-OECD countries.

GDP

In my model I will also include the GDP levels of the home and the host country. The data for the GDP levels I obtained from the CEPII Gravity database. CEPII, a French research center in international economics, supplied the data that was used in the paper by Head, Mayer & Ries (2010). The data used for GDP comes from the World Bank's World Development Indicators. This data is almost perfect to use, but unfortunately the World Bank doesn't keep track of countries that merged or

separated. For example, data for Russia only exist from 1989 and further. That's why CEPII also used GDP estimates provided by Katherine Barbieri. When we combine both data sources, we get the GDP data per country we need.

Distance

The distance variable is obtained from the GeoDist database. This database created by the center of CEPII, is a collection of data provided by Mayer & Zignago (2005). They calculated the bilateral distances for 225 countries across the world. They provided multiple ways of calculating the distance between countries. The first one is widely used but is not the most accurate way of measuring the distances between countries in my opinion. In this method the distance between the capitals of both countries is measured. The distance between Amsterdam (capital of the Netherlands) and Washington D.C. (capital of the United States of America) is in this method 6.197 kilometers.

The second method is to calculate the distance between the economic centers of both countries. In some cases the capital of a country is not the economic center of a country. Again the example of the United States of America, CEPII considers the economic center of this country to be New York. The distance between Amsterdam and New York is 5.866 kilometers.

The third method which in my opinion is the most accurate and thus also the method that I will use in my empirical research is inspired by a model created by Head and Mayer (2002). In their paper they came up with a model to calculate the effective distances between states. They assume that states are the smallest unit for which trade flows are measured. But there exist smaller units: districts. So trade flows from district *k* to district *l* are given by x_{kl} . The function to measure the trade between state i and state j is:

(10) $$x_{\text{ij}} = \sum_{k\ \varepsilon\ i} \sum_{l\varepsilon\ j} x_{kl}$$

Head and Mayer also use the gravity equation to calculate the trade between the two districts. The equation then becomes

(11) $x_{ij} = \sum_{k\,\varepsilon\,i} \sum_{l\varepsilon\,j} GY_k Y_l d_{kl}^{\theta}$

Where Y_k and Y_l stands respectively for the total income for district k and the total income of district l, d is the distance between district k and l and θ is a parameter which we expect to be negative and G is a constant. The same gravity equation applies then also for the trade between two states.

(12) $GY_i Y_j d_{ij}^{\theta} = \sum_{k\,\varepsilon\,i} \sum_{l\varepsilon\,j} GY_k Y_l d_{kl}^{\theta}$

Which can be simplified to the effective distance between the states i and j:

(13) $d_{ij} = (\sum_{k\,\varepsilon\,i} \left(\frac{Y_k}{Y_i}\right) \sum_{l\,\varepsilon\,j} \left(\frac{Y_l}{Y_j}\right) d_{kl}^{\theta})^{\frac{1}{\theta}}$

Papers by Head and Mayer (2000), Helliwell and Verdier (2001) and Anderson and van Wincoop (2001) uses θ = 1. In that case the formula reduces to the average distance between economic centers.

As said before the method to calculate the distances is not exactly the same equation that is created by Head and Mayer (2002). The method that we use replaces income by population. So the formula becomes:

(14) $d_{ij} = \sum_{k\,\varepsilon\,i} \left(\frac{Pop_k}{Pop_i}\right) \sum_{l\,\varepsilon\,j} \left(\frac{Pop_l}{Pop_j}\right) d_{kl}$

We take the top 25 populated cities per country to measure the weighted distance between countries.

When we take a look at what the distance is between the Netherlands and the United States according to this method we see that the distance is much greater. The distance is now 7.282 kilometers. When we compare that with the 6.197 kilometers for the distance between the capitals of both countries and the 5.866 kilometers between the economic centers of both countries we see that the difference between these three methods is huge. Choosing one method over the other can have a huge impact on your results.

As said before, I will use the last method; this is in my opinion the most accurate method. Although most of the investments will be between the economic centers of countries, I think it is inaccurate to assume that there is only one city where investments occur. Also, in the method where we measure the distance between the capitals, only one city per country is considered. In the final method we look at the

25 most populated cities, in that method the most important regions of the country is captured and this will give a reliable measure for the distance between countries.

Population

The population coefficient is calculated using data from the CEPII Gravity database. They have collected data from the World Development Indicators.

Language

The common language variable I also obtained from the CEPII GeoDist database. The CEPII database collected their data from the CIA World Factbook and the website of Ethnologue, an organization who intend to provide a list of all the known living languages in the world. With all these information CEPII managed to create a database where the official languages, the languages spoken by at least 20% of the population in a country and the languages spoken by at least 9% of the population per country are registered. An official language is assigned when the language is used within the government (Mc Arthur, 1998). This does not necessarily mean that the official language is spoken by many inhabitants. For example according to the CIA World Factbook (2011) the Maori language which is an official language in New Zealand was in 2006 only spoken by 3,9% of the inhabitants of New Zealand. According to the paper by Melitz (2007) the languages that are spoken by at least 20% of the population and are also spoken in another country in the world are also recognized as an official language for that country. Melitz calls this open-circuit communication, he finds 15 of those open circuit communication languages[1].

The CEPII database relies on numbers created by the CIA World Factbook and the website of Ethnologue to recognize which languages are used by more than 9% of the population in each country. The database has a maximum of four languages spoken by more than 9% of the population.

With all this data available it is easy to see if a common language exists among country pairs. When a common language exists the dummy variable will represent

[1] Arabic, Chinese, Danish, Dutch, English, French, German, Greek, Hindi, Malay, Persian (Farsi), Portuguese, Spanish, Swedish and Turkish

the value of 1 and when there is no common language then the dummy variable will be 0.

Colony

Another variable which I will include in my model is colony. This data is also available via the CEPII Gravity database and again their information comes from the CIA World Factbook. This dummy variable is a 1 when a colony relationship ever existed and a 0 if such a relationship not existed. The same method applies for a colony after 1945. The dummy will be 1 when a colonial relationship between the two countries existed and 0 when this is not the case.

Contiguity

The last variable I will include in my model is whether two countries share a border. This data is also from the CEPII Gravity database. When they do the dummy variable will be 1 and when the two countries don't share a border then the dummy will be 0.

The data in the CEPII database contains data only to the year 2006. The data regarding FDI is from 1985 and newer. Therefore my model contains data from 1985 to 2006.

7 Empirical section

7.1 The model

In the empirical section of my thesis I want to see if a common language and/or a colonial past is related to the FDI flows between countries. I will include the control variables: GDP of the home country, GDP of the host country, distance, the population of the home country, the population of the host country and contiguity. Further I will add the variables that I want to investigate: common official language, common spoken language, ever had a colonial relationship and a colonial relationship after 1945. My empirical model will then be as follows:

(15) $$\ln(FDI_{ijt}) = \beta_0 + \beta_1 \ln(\mathrm{GDP_{it}}) + \beta_2 \ln(\mathrm{GDP_{jt}}) + \beta_3 \ln(\mathrm{DIS_{ij}}) + \beta_4 \ln(\mathrm{POP_{it}}) + \beta_5 \ln(\mathrm{POP_{jt}}) + \beta_6 \mathrm{LANGOFF_{ij}} + \beta_7 \mathrm{LANGETH_{ij}} + \beta_8 \mathrm{COL_{ij}} + \beta_9 \mathrm{COL45_{ij}} + \beta_{10} \mathrm{CONTIG_{ij}} + \varepsilon_{ij}$$

7.2 Expectations of the variables

GDP of the home country

When the home country has a high level of GDP this indicates that there is a lot of economic activity in the home country. When there is a lot of economic activity in the home country this means that there are a lot of firms who can export/invest abroad. So I expect that β_1 will have a positive sign.

GDP of the host country

When the host country has a high level of GDP this indicates that there is a high level of economic activity in the host country. When there is some economic activity it is easier for a company to invest in. Most of the time when there is economic activity this indicates that at least basic infrastructure is available; this of course, is a positive value factor for companies searching for a place to invest. Also when there is more economic activity, the potential market is in most cases larger. So I expect that β_2 will have a positive sign.

Distance

In several papers, for example the paper written by Martinez-Zarzoso & Nowak-Lehmann (2003) we see that physical distance ($\ln(DIS_{ij})$) has a significant negative effect in the gravity model of trade. This can be explained by the fact that, in most cases the greater the distance, the higher the transportation costs will be. So it will be unattractive to serve markets where the distance is greater. As we saw in the OLI framework it may be interesting for a company, when the transportation costs are high, to serve the potential market via FDI. This is also true according to the proximity-concentration trade off by Brainard (1993). So according to these theories the coefficient for distance should be more positive in the gravity model of FDI than the coefficient for distance if we look at the gravity model of trade. This means that if a potential market is far away, then the most likely method of serving this market is through FDI, so that the distance coefficient for FDI is higher than for trade, but the coefficient is still negative as transportation costs rise when distance increases. This is also shown in the paper by Gopinath and Echeverria (2004) where distance has a significant (5% level) negative coefficient for the trade/FDI ratio. My expectation is thus that β_3 will have a negative sign.

Population in the home country

The sign of this variable can be positive or negative. When the population is big in the home country, it is possible that firms will produce more in their own country and will not export or invest in another country. There is significant potential of workers and the home country can fulfill the needs of the local firms (absorption effect). In this case the population coefficient will be negative. But when the population is big the consumption of the home country will also be high. To supply those needs, the production must go up. Therefore it is possible that FDI will go up (economies of scale). Whether the coefficient is positive or negative will depend on if the absorption effect or the economies of scale dominates (Martinez-Zarzoso & Nowak-Lehmann 2003). So it is difficult to predict the sign of β_4.

Population in the host country

The expectation for this variable is unsure, just as the expectation for the population in the home country. When the population is high there is a high availability of labor for a company that is considering investing in the foreign country. When this is the situation the coefficient β_5 will be positive. On the other hand, when the population is big, consumption is also high. To provide enough products for their own population, more labor is required to meet the increasing demands. Under normal price elasticity, when the demand for a product or service goes up, so does the price. In this case the demand for labor goes up as well as labor wages. This can result in a firm choosing another country to invest in. Another explanation why the coefficient could be negative is the case when the population of the host county grows; there will be more initiatives to start a company by the local population. The government may prefer local business over foreign investments and may implement barriers of entry for foreign companies. When these effects dominate, the sign of the coefficient β_5 will be negative. So it is not yet clear whether β_5 will be positive or negative.

Common Official language

I expect that countries that share an official language will have higher FDI flows than countries that don't share an official language. Although there are currently programs that can translate one language to the other, it is easier if both countries can communicate in a common language. Melitz (2007) mentions the importance of direct communication where inhabitants of two different countries can communicate without the help of any translation service. Although there are no direct costs for use of these translation services (internet or a dictionary) there is a higher chance of some kind of faux pas. So I expect that FDI flows will be higher if both countries have a common language, so I expect that the sign of β_6 will be positive.

Common Spoken Language

My prediction for the coefficient for a common spoken language is the same as for the common official language. A common spoken language exists when two countries share a language that is spoken by at least 9% of the population in each

country. So that when a common spoken language exists it will be easier to communicate and thus β_7 will be positive.

Colonial relationship

When countries have had a colonial relationship, in most cases, the countries will share some cultural values. So it will be easier for the companies to adapt to the norms and values that are present in the host county and it will also be easier for the population of the host country to meet the standards of the multinational. So I expect that the sign for β_8 will be positive.

Colonial relationship after 1945

When the colonial relationship still exists after 1945 then my expectation is that this will give a positive coefficient. I expect to find cultural similarities between countries that are connected after 1945. These similarities will make communication easier between those countries. I expect that the cultural similarities are stronger when the colonial relationship ended recently. So I expect that β_9 is positive.

Contiguity

When countries are close enough to each other that they share at least one common border, then most of the time the countries are much alike. I expect that this stimulates the investments from one country into the other country. So my expectation is that β_{10} is positive.

7.3 Results

7.3.1 Dataset 1985-2006

Table 1. Results for dataset 1985-2006

				Cluster		
Independent variables	Estimate	OLS	Robust	Home	Host	Home & Host
Home GDP	2,01	72,16**	64,20**	12.03**	27.59**	33,74**
Host GDP	0,86	75,32**	71,76**	13.42**	13.81**	30,61**
Distance	-0,55	-31,51**	-30,96**	-9.60**	-7.89**	-12,53**
Home Population	-1,24	-44,08**	-39,20**	-7.39**	-18.22**	-20,80**
Host Population	-0,23	-17,14**	-16,37**	-4.31**	-3.49**	-7,07**
Common Official Language	0,69	8,34**	9,18**	3.11**	4.14**	4,23**
Common Spoken Language	0,53	7,06**	8,21**	3.15**	3.27**	3,79**
Colony	0,98	12,48**	14,15**	5.13**	6.47**	6,04**
Colony after 1945	-0,50	-4,57**	-4,94**	-1.76	-1.73	-1,80
Contiguity	0,54	7,44**	8,09**	2.80**	2.16*	3,10**
Adjusted R^2	0,52					

Notes: All variables except the dummies are expressed in natural logarithms. T-statistics are in parentheses. ** denotes significance at the 1% level, * denotes significance at the 5% level. The number of observations is 16380.

Expectations versus Results

Table 2. Expectations and results per variable

Explanatory variable	Expected Result	True Result
GDP of the home country	Positive	Positive
GDP of the host country	Positive	Positive
Distance	Negative	Negative
Population in the home country	Positive/Negative	Negative
Population in the host country	Positive/Negative	Negative
Common Official language	Positive	Positive
Common Spoken language	Positive	Positive
Colonial relationship	Positive	Positive
Colonial relationship after 1945	Positive	Negative
Contiguity	Positive	Positive

The GDP's for the home and host country have, as expected, a positive influence on FDI. This means that when wealth increases in the home and host country it is highly likely that more investments will take place. On average the FDI flows between the home country and the host country will go down when distance increases as can be seen by the negative coefficient for distance. The size of the population in the home country has a negative effect on FDI. An explanation for this negative coefficient can be that the absorption effect dominates the economies of scale effect. The working potential is large enough to fulfill the needs of the companies in the home country. Also the population in the host country has a negative effect on FDI; this also could be explained as the host country is self-sufficient. The host country don't require firms coming to their country, therefore the inflow of FDI is smaller. We can derive from the positive coefficients for the language variables that an easy communication is important for FDI. The coefficient for a colonial relationship is positive which could indicate that the countries share some common values which stimulate FDI. Unexpected is the negative coefficient for countries that are involved in a colonial relationship after 1945. One possible

explanation could be that the inhabitants of the colonized country were exploited and norms and values were not accepted by the colonizing country. As an example we can take the relationship between India (dependant colony) and Britain (imperial power). Because the British didn't respect the holy rituals of the Indian inhabitants it resulted in the Sepoy mutiny in 1857 (Sebastian Sanne, 2003). Because the British did not show any respect and exploited the country and the inhabitants, the hate against Britain is still present in parts of India. In the colonized period, India was a very important country for Britain to invest in. But after the decolonization the economic interaction between India and Britain diminished rapidly (Tomlinson 1978). This hate could explain why the existence of a colonized relationship after 1945 has a negative effect on FDI. When countries share a common border this means that the distance between countries cannot be extremely far, when countries share a border this has an additive positive influence on FDI as can be seen from the positive coefficient for contiguity. This is also expected.

OLS

We see that all the variables are highly significant, at the 1% confidence level. We see that GDP of the home country and GDP of the host country are extremely significant.

Regression with robust standard errors

It is possible that the data is not normal divided. It can be that there are outliers and that some heteroskedasticity exists. This will influence the standard errors of the variables. When the standard errors of a variable are like the one in the left panel of figure 3, where the variance is higher when the distance between x and $\bar{x}$ increases then the standard errors in the OLS regression are too small. The variance in OLS is calculated as

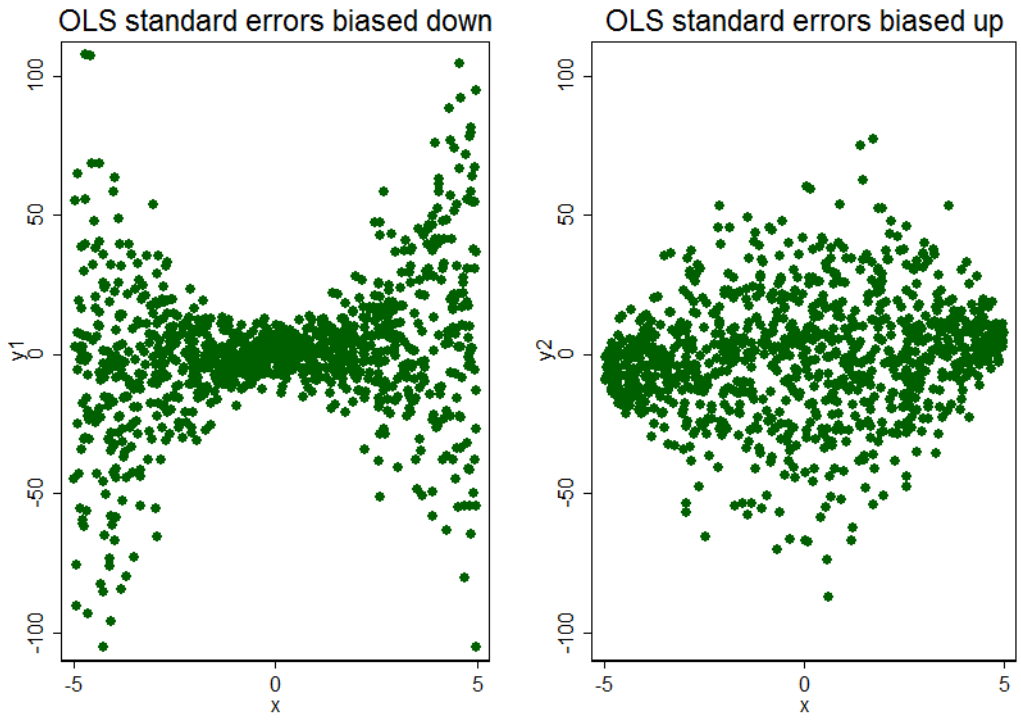
(16) $s^2 = \frac{1}{n-1} \Sigma (x_{i-\bar{x}})^2$ (Moore et al., 2003).

In the left panel of figure 3 we see that when x is far from its mean then the variation is larger than in a normal distribution, but in the calculation for the variance there is no correction for this. When we correct for this to obtain robust standard errors we

will get higher standard errors than in an OLS regression for the variables like the one in the left panel of figure 3.

When the standard errors are distributed as in the right panel of figure 3, then the standard errors are too high. We see that when x is far from its mean the variance of the error term is small. So the standard error is actually smaller than calculated in the OLS regression. (Chris Auld, 2012)

Figure 3. Two types of heteroskedasticity



First I will check if heteroskedasticity exists in my model using the Breusch-Pagan/Cook-Weisberg test. The outcome can be found below in table 3.

Table 3. Test for heteroskedasticity 1985-2006

Chi-square	481.71
Prob>chi-square	0.0000

This table indicates that there is heteroskedasticity present in this model.

STATA offers the possibility to correct for these flaws. (Chen, Ender, Mitchell & Wells, 2003). In OLS, we assume that all the standard errors are independent and

identically distributed. If we want to obtain robust standard errors we relax the assumption that the standard errors are identically distributed. Since we now have robust standard errors, we have quite accurate p-values. These results can be found in the column "Robust" in table 1. The coefficients don't change, but the t-statistics do change. All the variables are still significant at the 1% confidence level and the effects on the t-statistics are small. So if there is heteroskedasticity in my model, then it doesn't affect my model so that variables become insignificant.

Regression with clusters

It is possible that a relationship exists between the residuals within a cluster. If we use the cluster command in STATA we relax the assumption that standard errors are independent within a cluster. Although there are a lot of possible clusters I choose to cluster on the home country, the host country and the home-host country combination. We can see in table 1, that clustering has a huge impact on the t-values of all variables. Although the t-values dropped the effect on the significance of the variables is small. A colony after 1945 is no longer significant in all three of the clusters I created. This was the variable in OLS with the least significance in this model. So after clustering, the effect of being a colony after 1945 has no significant impact on the amount of FDI between countries. Contiguity still has in the home country cluster and the home-host country cluster a significant influence on the flows on the 1% significance level, but it is only significant on the 5% confidence level if we use the host country cluster.

Fixed Effects

We saw that all the variables were significant at the 1% confidence level. According to some papers these results are biased because there might be some country-specific or time-specific factors that influence the result, but these factors were not included in the model. An example of a country-specific factor that might influence the outcome is the size of a country. Since this factor (in most cases) does not change over time it will suffice to create a dummy for each country in our model. The same I will do for each year in my model. This can correct for events like technological progress.

The formula now becomes:

(17) $\ln(FDI_{ijt}) = \beta_0 + \beta_t + \beta_i + \beta_j + \beta_1 \ln(\mathrm{GDP_{it}}) + \beta_2 \ln(\mathrm{GDP_{jt}}) + \beta_3 \ln(\mathrm{DIS_{ij}}) + \beta_4 \ln(\mathrm{POP_{it}}) + \beta_5 \ln(\mathrm{POP_{jt}}) + \beta_6 \mathrm{LANGOFF_{ij}} + \beta_7 \mathrm{LANGETH_{ij}} + \beta_8 \mathrm{COL_{ij}} + \beta_9 \mathrm{COL45_{ij}} + \beta_{10} \mathrm{CONTIG_{ij}} + \varepsilon_{ij}$

β_0 is the intercept which is the same for each year, home country and host country. B_t is an intercept which changes for each year *t* and is the same for each home country and host country. B_i depends on the home country *i* and is the same for each year and host country. B_j depends on the host country *j* and is the same for each year and home country.

Table 4. Fixed effects, dataset 1985-2006

				Cluster		
Independent variables	Estimate	OLS	Robust	Home	Host	Home & Host
Home GDP	1,10	8,85**	8,02**	3.61**	5.90**	6,70**
Host GDP	0,66	10,45**	10,04**	6.82**	6.95**	7,94**
Distance	-1,07	-43,36**	-41,56**	-12.10**	-12.52**	-16,00**
Home Population	-2,85	-3,82**	-3,76**	-1.00	-2.89**	-2,82**
Host Population	-2,34	-10,07**	-9,69**	-5.45**	-4.23**	-6,63**
Common Official Language	0,53	6,66**	6,85**	3.08**	3.16**	6,82**
Common Spoken Language	-0,15	-2,04*	-2,06*	-1.42	-0.96	-1,00
Colony	1,04	14,86**	15,46**	5.03**	6.99**	6,82**
Colony after 1945	-0,18	-1,71	-1,78	-0.52	-0.64	-0,71
Contiguity	0,51	7,84**	7,47**	2.38*	2.69**	3,08**
Adjusted R^2	0,67					

Notes: All variables except the dummies are expressed in natural logarithms. T-statistics are in parentheses. ** denotes significance at the 1% level, * denotes significance at the 5% level. The number of observations is 16380.

When we look at the results after including those fixed effect dummies in table 4 we see that the coefficients for each variable have changed significantly. The coefficient for GDP has decreased for the home and for the host country. The importance of the

GDP of the home country almost halved from 2,02 to 1.10. The negative coefficients for distance and for the population for the home and the host country became even more negative. The effect on the coefficient for countries sharing an official language are small. The coefficient for countries sharing a language spoken by at least 9% of the population has changed. Even the sign of the coefficient changed from positive to negative. I have no explanation why having a common spoken language would have a negative effect on FDI. Although significantly different from zero for OLS and robust standard errors the effect is very small. The coefficients for colony and contiguity were almost unaffected and are still positive. The effect of countries that had a colonial relationship on FDI is still negative, but the effect is very small.

OLS

The variables that look at the GDP for the home and host country are still significant, but the t-values have dropped. It is even clearer that distance has an effect on FDI in the fixed effects model. The population for the origin and destination country now has a much stronger negative effect than in the case of the original model, but again an enormous drop in the t-values. My assumption is that there are variables which are not captured in the model, contributed to the coefficients of those variables in the previous model. A common official language is still a variable which has a significant positive effect on FDI. This cannot be said of countries which share a language spoken by at least 9% of the population. This variable is no longer significant at 1%, but it is still significant at the 5% confidence level. This is not the case for countries that have a colonial relationship after 1945. The coefficient is still negative, but it is not significant anymore. The t-statistics of countries that ever had a colonial relationship is still significant and did not change much. The same is the case for contiguity.

Regression with robust standard errors

The effect of correcting for heteroskedasticity is just as in the model without the fixed effects very small. The effect on the t-values for each variable is very small and all the variables which were significant at the 1%/5% confidence level are still

significant at the 1%/5% confidence level. So as in the model without the fixed effects the influence of heteroskedasticity is very small.

Regression with clusters

As in the model without the fixed effects; I will add a cluster on home country, host country and home-host country to the model. The t-values of home GDP and host GDP drop even further after adding the clusters, but the variables still remain significant at the 1% confidence level. The t-values for distance and host population also drop but the values are still very high and thus remain significant. This is not the case for home population, the t-value is very low in the fixed effects model compared to the normal model, but dropped even more after adding clustering to the model. It dropped so much that the home population has no significant effect after clustering for the home country. A common spoken language is also no longer significant. All other variables remain roughly as significant as before.

7.3.2 Dataset 2000-2006

To see whether the coefficients displayed in the previous paragraph are still relevant I took a sample of the data where I removed all the data prior to the year 2000. The results can be found in table 5.

Table 5. Results for dataset 2000-2006

				Cluster		
Independent variables	Estimate	OLS	Robust	Home	Host	Home & Host
Home GDP	2,37	61,02**	54,97**	13.61**	27.08**	34,15**
Host GDP	0,86	51,98**	51,34**	13.49**	15.27**	30,81**
Distance	-0,70	-26,57**	-26,30**	-9.99**	-10.01**	-15,37**
Home Population	-1,61	-41,50**	-37,24**	-9.72**	-20.41**	-23,25**
Host Population	-0,24	-12,37**	-12,02**	-5.03**	-3.68**	-7,12**
Common Official Language	0,57	4,71**	5,14**	2.63*	3.12**	3,24**
Common Spoken Language	0,39	3,41**	4,00**	2.09*	2.09*	2,57**
Colony	1,17	9,18**	10,04**	4.47**	5.63**	5,53**
Colony after 1945	-0,31	-1,78	-2,08*	-0.80	-1.09	-1,12
Contiguity	0,78	6,93**	7,06**	3.81**	2.68**	3,83**
Adjusted R^2	0,55					

Notes: All variables except the dummies are expressed in natural logarithms. T-statistics are in parentheses. ** denotes significance at the 1% level, * denotes significance at the 5% level. The number of observations is 7979.

What we can see from the results in table 5 is that the coefficients for each variable are roughly the same as in the regression with the dataset 1985-2006. None of the variables changed sign, so all coefficients that were positive remained positive and all coefficients that were negative remained negative.

OLS

The t-statistics in the dataset 2000-2006 are lower for all variables. An explanation for this can be that if there are outliers then the effect on the t-statistics in the dataset 1985-2006 is much smaller than in the dataset 2000-2006. In the dataset 1985-2006 all variables were significant at the 1% confidence level. We see that this is not the case for the dataset 2000-2006. The variable colony after 1945 is no longer significant. All other variables are also significant at the 1% confidence level in this dataset.

Regression with robust standard errors

I will start by checking if heteroskedasticity is present in the model 2000-2006.

Table 6. Test for heteroskedasticity 2000-2006

Chi-square	225.12
Prob>chi-square	0.0000

So also in this model there is heteroskedasticity present.

Just as in the model with data for all years we see that the t-values with robust standard errors don't differ a lot from the OLS model. All variables that were significant at the 1% level are still significant at the 1% level. Colony after 1945 is now significant at the 5% level. What we see that after getting robust standard errors the t-statistics for home GDP, host GDP, distance, home population and host population went down and for common official language, common spoken language, colony, colony after 1945 and contiguity the t-statistics went up. This also happened in the dataset for 1985-2006.

Regression with clusters

Just as we saw in the dataset 1985-2006, the t-statistics for each variable after clustering dropped compared to the OLS model.

For the home country cluster, we see that the significance for the two variables that compare the home and host country sharing a language dropped to 5%. There is no evidence that a colonial relationship after 1945 influences the FDI flows between countries.

For the host country cluster, we see that the common spoken language is only significant at the 5% confidence level; all other variables are at the same significance level as the OLS model.

The variables in the home-host country cluster are significant at the same confidence levels as the OLS model.

Fixed after 2000

Table 7. Fixed effects, dataset 2000-2006

				Cluster		
Independent variables	Estimate	OLS	Robust	Home	Host	Home & Host
Home GDP	1,35	4,79**	4,40**	4.03**	4.39**	4,57**
Host GDP	0,78	5,10**	5,05**	4.88**	5.42**	5,43**
Distance	-1,23	-30,61**	-30,19**	-9.92**	-13.81**	-17,70**
Home Population	-4,07	-1,59	-1,49	-1.06	-1.65	-1,48
Host Population	-3,53	-3,27**	-3,05**	-2.14*	-2.49*	-3,02**
Common Official Language	0,29	2,47*	2,56*	1.72	1.48	1,69
Common Spoken Language	0,01	0,13	0,13	0.09	0.08	0,08
Colony	0,92	8,21**	8,21**	3.83**	5.05**	4,94**
Colony after 1945	0,22	1,35	1,50	0.59	0.78	0,84
Contiguity	0,62	6,10**	5,62**	2.68*	2.71**	3,13**
Adjusted R^2	0,69					

Notes: All variables except the dummies are expressed in natural logarithms. T-statistics are in parentheses. * denotes significance at the 1% level, ** denotes significance at the 5% level. The number of observations is 7979.

We see that the fixed effects model is very different from the normal model. The estimates for some variables are much higher and for other variables much lower. For example, the estimate for host population was -0,24 in the normal model, in the fixed model the estimate is -3.53. So the size of the host population has a much greater impact on the FDI flows in the fixed effects model. The opposite is true for home GDP, in the normal model the coefficient was more than double the coefficient in the fixed effects model. This variable has become less important in the fixed effects model. Although very close to zero it is remarkable that the coefficient for countries that had a colonial relationship after 1945 is positive. In all previous models the coefficient had been negative.

OLS

Home population which was significant at the 1% confidence level in the normal model, in the fixed effects model the variable is not significant different from zero. The standard error for this variable is now very high. Also a common spoken language has no significant effect on FDI. Common official language still has a positive effect on FDI, but only at the 5% level. Although we saw that having a colony after 1945 is positive it is not significant. All other variables are still significant at the 1% confidence level.

Regression with robust standard errors

Just as in the previous models the effect of correcting to robust standard errors is small. The t-statistics are almost the same as in the OLS model and none of the variables improved or decreased in the levels of confidence.

Regression with clusters

In most cases, the t-statistics drop after clustering. Clustering on the home country results in insignificance for common official language. Host population and contiguity are now only significant at the 5% confidence level. The other variables are the same as in the OLS model.

After clustering on the host country we see that the variables are almost the same as in the home country cluster, only the t-statistic for contiguity increased so much that it became significant at the 1% confidence level.

Compared to the OLS model each variable remains at the same confidence level in the home-host country cluster, except for common official language which lost its significance.

7.3.3 The final model

We see that there are a lot of different methods to choose from. For my model I will choose the fixed effects model. With this model I will correct for country-specific and time-specific variables that were not included in the model. There are some

variables that I not included in the model that may or may not influence the FDI flows. An example of this is the size of the home and the host country. Without using the fixed effects model I assume that all country and time-specific variables are in the model and this is not realistic.

Since it is highly likely that the results within a cluster are more similar to each other than the results in another cluster I will use clustering in my model. I collected the FDI flows based on the outward FDI, provided by the home country. Therefore I will cluster my results on the home country. It is most likely that some institutions will present higher or lower FDI outflows compared to the actual FDI outflows. By clustering on home country we will correct for these errors. We saw that there was a high possibility that there was some heteroskedasticity in the model. After clustering the standard errors are robust. So the model that I will use is the fixed effects model with clustering on the home country.

Table 8. Overview for both datasets for the final model

Independent variables	Estimate complete model	Complete model	Estimate after 2000	After 2000
Home GDP	1,10	3.61**	1,35	4.03**
Host GDP	0,66	6.82**	0,78	4.88**
Distance	-1,07	-12.10**	-1,23	-9.92**
Home Population	-2,85	-1.00	-4,07	-1.06
Host Population	-2,34	-5.45**	-3,53	-2.14*
Common Official Language	0,53	3.08**	0,29	1.72
Common Spoken Language	-0,15	-1.42	0,01	0.09
Colony	1,04	5.03**	0,92	3.83**
Colony after 1945	-0,18	-0.52	0,22	0.59
Contiguity	0,51	2.38*	0,62	2.68*
Adjusted R^2	0,67		0,69	

In this section I will discuss the results I obtained from the fixed effects model, clustered on the home country. When we compare the results from the dataset 1985-

2006 to the results in the dataset 2000-2006, we see that there are some variables whose t-values drop, but only a common official language loses its significance. An explanation for this can be found in the reduced importance of having the same official language. As children receive more education regarding foreign languages, there is an increased chance of a common language which is spoken in the home and host country. When there is no common language, communication has become easier with the help of the internet, translation programs and translators.
I will compare the result of the dataset 1985-2006 with fixed effects and clustering on the home country to three papers where the gravity equation on FDI was executed.

- Talamo (2007) did a regression where fixed effects for the home country and fixed effects for the host country were added. The standard errors were robust standard errors and the home country is an OECD country.
- Gao (2009) did an OLS regression and a regression where fixed effects for the host-home country were added, because of these fixed effects all information regarding the relation between home and host countries are captured by the fixed and thus removed from the regression. Therefore I will compare some variables with the OLS model and some variables with the fixed effects model. All standard errors are robust standard errors. The home country is in all cases an OECD country.
- A comparable model performed by Paniagua (2011) is an OLS model with robust standard errors and there is no restriction on the home country.

Although these models are not exactly the same as my model I will make a comparison with the common variables when present.

As expected the home country GDP is highly significant and positive. The estimate for home GDP is greater than one; this means that when GDP rises in the home country the amount of FDI will grow with a greater amount. This variable had in the papers by Gao and Paniagua a significant positive influence on FDI. Host country

GDP is in my model also a positive significant variable. An increase in the GDP of the host country results in an increase in the FDI flow. This variable is not significant in the paper by Gao and Paniagua, but even more surprising the estimate is negative in the paper by Gao. This is a result which was also surprising and alarming to the author himself and he was unable to provide a reasonable explanation. Distance has a very significant negative effect on FDI. This variable which was significant in all papers and in all papers it had a negative coefficient. We see that the size of the home countries population is not significant, so it is not significant whether the absorption effect or the economies of scales effect (as described in the expectations of the variable) dominates. The size of the population in the home country is also not significant in the paper by Talamo. Although the estimate for home population is positive (2,594) it is not significantly different from zero. The estimate for the host population is negative and highly significant. A rise in the host population thus results in a decline in the FDI flow. The high significance and the high negative estimator are also results obtained by Talamo. Countries sharing an official language will have a positive influence on the amount of FDI. This is also what we see in the paper by Gao, Talamo and Paniagua. Although an official language does not tell us anything about how many people speak the language in the country it has a positive effect on FDI. Countries that share a common spoken language have no significant influence on the FDI flow. The estimate is very close to zero and we can't say with 95% confidence that it is different from zero. It is surprising to see that the estimate for this variable is negative, but as said before the estimate is very close to zero and it cannot be ruled out that the coefficient is in fact positive. A colonial relationship will have a positive influence on the FDI flow. A positive relationship between a colonial relationship and FDI can also be found in the paper by Paniagua. I suggest that there may be some cultural similarities between countries that had a colonial relationship which will stimulate the FDI flow. There is no significant effect on FDI when the colonial relationship ended after 1945. Contiguity had a positive, insignificant effect in the model by Gao. In my model the effect is also positive, but in contrast to Gao the

variable is significant. This means that countries will on average invest more in a country with a common border than investing in another country.

8 Conclusions and Summary

Although FDI flows are small compared to trade, FDI is growing very fast. The value of FDI in 2003 is almost ten times higher than it was in 1970. A good model to predict the choice for a company where to invest is the gravity model. This model was mainly used in Newton's law of universal gravitation and the gravity model of trade, but it remains a very strong empirical power for the gravity model of FDI. It took a few years for a theoretical framework for the gravity model of FDI to be established, but well worth the wait.

The variables that I tested in this paper are:

- GDP of the home country
- GDP of the host country
- Distance between the home and the host country
- The size of the population of the home country
- The size of the population of the host country
- Home and host country sharing an official language
- Home and host country sharing a spoken language by at least 9% of the population
- Home and host country had a colonial relationship
- Home and host country had a colonial relationship after 1945
- Home and host country share a border

The method that is used in this paper to test the variables is the fixed effects model where I clustered the data on the home country. I choose the fixed effects model, because there are most likely some country- or time specific factors not included in the model. When the fixed effects model is used a correction takes place for these factors. Clustering was required, because it was highly likely that the results within one cluster were more similar than the results in another cluster. The FDI data was collected based on the data provided by the home country, therefore I have clustered my data on the home country. The four variables that were the primary variables of

interest in this paper were two variables concerning common languages and two variables concerning colonies. A common official language has shown to be a significant positive effect on FDI in the dataset 1985-2006 and the dataset 2000-2006. A common spoken language has also shown this effect in the dataset 1985-2006. In the dataset 2000-2006, this variable was no longer significant. Reasons for this can be; because of improved education and improved technology, communication is easier even when no common spoken language exists. A colonial relationship means that on average there will be more potential for FDI between those countries. A colonial relationship after 1945 is not significant differently from zero in both datasets. It is remarkable that when the variables would be significant, that in the dataset 1985-2006 the variable is negative and in the dataset 2000-2006 the variable is positive. Although it is difficult to draw conclusions from this, becausethere is no significance, an explanation for this can be that in the dataset 1970-2006 there is still some resentment present, while in the dataset 2000-2006 the resentment faded and cultural similarities dominate.

GDP of the home country and GDP of the host country are as expected positive and significant, distance has as expected a negative and significant coefficient. The size of the population of the home country has no significant effect on FDI. When the size of the population of the host country goes up the amount of FDI goes down on average. We have seen that distance has a negative effect and on top of that, when countries share a common border the amount of FDI is on average even higher.

Some improvements can be done in a next study. When collecting the data for FDI, I only looked at the data provided by the home country. This data, provided by the host country, does not always match and in some cases differs greatly from the data provided by the home country. In a next study it may be interesting to see what the outcome will be if we rely on the data provided by the host country or average the data that is supplied by the home and host country.

References

Anderson, James E. (1979). A *Theoretical Foundation for the Gravity Model*. American Economic Review, 69, 1, 106-116

Anderson, James E. & Wincoop, Eric van (2003). *Gravity with Gravitas: A Solution to the Border Puzzle*. American Economic Review, Vol. 93 (1), 170-192

Ault, Chris (2012). "The Intuition of Robust Standard Errors", *University of Victoria*, http://chrisauld.com/2012/10/31/the-intuition-of-robust-standard-errors/ (downloaded 10th May 2013)

Baier, Scott L. & Bergstrand, Jeffrey H. (2003). *Endogenous Free Trade Agreements and the Gravity Equation*. Atlanta: Georgia Institute of Technology

Bergstrand, Jeffrey H. (1990). *The Heckscher-Ohlin-Samuelson Model, The Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade*. The Economic Journal, 100, 1216-1229

Brainard, S. Lael (1993). *An Emperical Assessment of the Proximity-Concentration Tradeoff Between Multinational Sales and Trade*. National Bureau of Economic Research, No. 4580

Chen, X., Ender, P., Mitchell, M. & Wells, C. (2003). *Regression with STATA*. UCLA Academic Technology Services

CIA World Factbook (2011). "New Zealand". *Central Intelligence Agency*, https://www.cia.gov/library/publications/the-world-factbook/geos/nz.html (downloaded 11th April 2013)

Deardoff, Alan (1998). *Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?*. The Regionalization of the World Economy, Chicago, University of Chicago Press

Eaton, Jonathan & Kortum, Samuel (2002), *Technology, Geography and Trade*. Econometrica, Vol 70 (5), 1741-1779

Gao, Shen (2009). *The Predictive Capacity of the Gravity Model of Trade on Foreign Direct Investment*. Nationalekonomiska Institutionen

Gopinath, Munisamy & Echeverria, Rodrigo (2004). *Does Economic Development Impact the Foreign Direct Investment-Trade Relationship? A Gravity-Model Approach*. American Journal of Agricultural Economics, vol. 86(3), 782-787

Head, Keith (2003). *Gravity for Beginners*. Vancouver, University of British Columbia

Head, Keith & Mayer, Thierry (2002). *Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade*. Working Papers 2002-01, CEPII Research Center

Head, Keith Mayer, Thierry & Ries, John (2010). *The Erosion of Colonial Trade Linkages after Independence*. Journal of International Economics, 81(1): 1-14

Helpman, E. (1984), *A simple theory of international trade with multinational corporations*. Journal of Political Economy 92, pp. 451-472.

Helpman, Elhanan, Melitz, Marc J. & Yeaple, Stephen R. (2003). *Export Versus FDI*. NBER Working Paper No. 9439

Leibrecht, Markus & Riedl, Aleksandra (2012). *Modelling FDI based on a spatially augmented gravity model: Evidence for Central and Eastern European Countries*. University of Lüneburg, Working Paper Series in Economics No. 239

Linnemann, Hans (1966). *An Econometric Study of International Trade Flows*. Amsterdam: North Holland Publishing Company

Marrewijk, Charles van (2007). *International Economics Theory, Application and Policy*. New York: Oxford University Press

Martinez-Zarzoso, Inmaculda (2003). *Gravity Model: An Application to Trade Between Regional Blocs*. Atlantic Economic Journal, Vol 3 (2), 174-187

Martinez-Zarzoso, Inmaculda & Lehman, Felicitas Nowak. (2003). *Augmented Gravity Model: An Empirical Application To Mercusar-European Union Trade Flows*. Journal of Applied Economics, Vol VI (002), 201-316

Mayer, Thierry & Zignago, Soledad (2011). *Notes on CEPII's distances measures: The GeoDist database*. CEPII Working Paper 2011-25

Mauro, Francesca di (2000). *The Impact of Economic Integration on FDI and Exports: a Gravity Approach*. Centre for European Policy Studies, No. 156

McArthur, Tom (1998). *Official Language*. Concise Oxford Companion to the English Language, Oxford University Press

Melitz, Jacques (2007). *Language and Foreign Trade*. CEPR Discussion Papers 3590

Moore, David S., McCabe, George P., Duckworth, William M., Sclove, Stanley L. (2003). *The Practice of Business Statistics*. New York: W.H. Freeman and Company

OECD (2007). *International Investment Perspectives: Freedom of Investment in a Changing World*. Paris: OECD

OECD (2008). *OECD Benchmark Definition of Foreign Direct Investment*. Fourth Edition. Paris: OECD

OECD (2012). *FDI flows by partner country*. Paris: OECD

Paniagua, Jordi (2011). *FDI Gravity Equation: Models, Estimations and Zero*. Catholic University of Valencia

Sanne, Sebastian (2003). *British Colonialism in India and its Influence on Indian Society*

Talamo, Giuseppina MariaChiara (2007). *Institution, FDI and the Gravity Model*, Journal of International Economics, VOL. 65(2), 269-296

The Economist (2009). *Moving on up*. London, The Economist Newspaper Limited

Tinbergen, Jan (1962). *Shaping the World Economy*. New York: The Twentieth Century Fund

Tomlinson, B.R. (1978). *Foreign Private Investment in India 1920-1950*. Modern Asian Studies Vol. 12(4), 655-677