MAGISTERARBEIT

Titel der Magisterarbeit

THE IMPACT OF LINGUISTIC DISTANCE ON BILATERAL TRADE VOLUMES BETWEEN AUSTIA AND ITS TRADE PARTNERS

verfasst von

Jochen Löb

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. oec.)

Wien, 2015

Studienkennzahl lt. Studienblatt	066 913
Studienrichtung lt. Studienblatt	Magisterstudium Volkswirtschaftslehre
Betreut von	Univ.-Prof. Dipl.-Ing. Dr. Robert M. Kunst

Contents

Abstract

Language and trade are two aspects of human interaction that can be put into direct relation. Cultural overlaps of languages and the ability of a society to communicate in different tongues increase bilateral trade between countries and still are of importance. By using two different datasets I show that there is still a negative impact of differences of languages on trade as well as a positive impact of multilingualism in the European context. I am using variables of linguistic distance to show in a gravity of trade flows analysis how similarities between languages and the effort of reaching a multilingual society increase bilateral trade.

1 Introduction

Linguistic distance plays an important role in the context of trade relations. Whenever an economic transaction is conducted or an offer is negotiated the tool of language is present. Though there might be platforms of trade and a huge supply of service providers that facilitate an individual's economic needs, the ignorance of language always puts one into a disadvantageous position in the play of asymmetric information. And it is expansive to buy a proper replacement for language skills.

In the European Union all official languages share the common privilege to have an equal status. On the first view the immense translation apparatus of the European Parliament and other European institutions might irritate as it seems inefficient and too elaborated. One might think "Why don't they just all speak English?". In a second view it is a dignified tool of respect and a cultural preservation programme that enlivens the power of diversity on this fusing continent.

Regarding on what permanent role language plays in every aspect of our life it is an impossible task to put the understanding between people and the interaction of languages into a numeric scale. Nevertheless language is an opportune gate to include culture, social interaction and political connection into a very general analysis on a macroeconomic level. But as we will see later on, language can be discerned as a static institution that dictates natural cultural borders or as a cheap opportunity to overcome those borders and increase bilateral trade. Together with physical and economic variables linguistic difference and thus linguistic distance variables have become elementary in studies about trade flows. Here I am using variables of linguistic distance to show in a gravity of trade flows analysis how language contains this opportunity.

Austria's geographical neighbourhood, which naturally in a huge part represents its trade partners, is built in a unique way inside the European Union. It is underestimated that cultural and with that linguistic and historic conditions predestine Austria to be a linking platform between all of the European Union's important linguistic streams. Language builds bridges not only in a human and a political way but especially in the economic sense. Understanding another country's culture is impossible without knowing of its language. That is why the role of direct communication in native and official languages must not be forgotten behind the enormous role of English as a lingua franca.

This paper is structured into two parts in which one focusses the pure interaction of countries as trading partners with their official languages and the other deals with the distribution of specific language skills within the population. The first case absolutely ignores that there are linguae francae in which economic transaction can be done and shows that there is a lot more behind the always present English interaction. In the second case the existence of English, French or German as a lingua franca is emphasized and the importance of a multilingual population shown.

These variables do not necessarily represent the economic reality as the quantity of usage of English or other languages in economic transactions is not addressed here. Further these variables serve as an instrument to quantify cultural and political ties and their impacts on trade between countries.

The first variable is based on pure calculation of symbols and shows cultural bonds as the linguistic development goes along with cultural and intellectual developments. The second variable which is based on the distribution of official languages and the language learning behaviour of a country's citizens has the potential to additionally show educational policies and political orientation in the observed countries. A third variable is constructed independently from pairwise connection, addressing the overall multilingual skills of a nation's society. Next to these three variables a couple of further variables used in other researches can be found in the literature review.

In the environment of a gravity research a linguistic distance variable is a good way to express linguistic difference and additionally cultural and historic ties. This is why I focus on this part of the empirical trade analysis and will compare an existing measure in newer and older data, using Austria as a centre of the study and a reference for questions about linguistic policies.

The questions that I want to adress here are:

Is there still an effect of languages on trade in the Austrian trade environment?

and

Is there an opportunity to improve linguistic skills and thus trade conditions in Austria?

The central and eastern European markets can be prospering for the Austrian economy if there is a willingness for openness and respectful cultural and thus linguistic handling especially from the Austrian side.

The gravity equation of trade is a suitable instrument to model all kinds of flows. It has been used massively in a huge amount of studies and the theoretical background has been established in the last decades.

2 Languages and trade

2.1 A linguistic trade history

Austrian trade is dominated by the economic partnership with Germany. Apart from many historic bonds between different parts of Germany and Austria, the two countries share their common German language which is spoken by close to all citizens natively and officially. Nearly all Austrians and Germans are able to speak in the standard High German language although this mostly written style is hard to find on the map in natively spoken form. Post world war II the division of the Third Reich had been successfully implemented politically but naturally the economic ties redeveloped and broadened between Austria and Western Germany. Due to the iron curtain the amount of trade with the East and especially with the Eastern European neighbours had been reduced and invoked an additional increase in trade with the Western hemisphere. The existence of a common language may have been one reason why trade between Austria and Germany developed such a close economic connection over the decades and centuries.

Countries that emerged out of the Austrian-Hungarian Empire and thus have a strong economic and infrastructural heritage to its former Capital, Vienna, have a very differentiated horizon in respect to their languages. Many of these countries like Slovenia, Slovakia, Czech Republic, Bosnia-Herzegovina and parts of Poland and Ukraine are linked by their tie to the Slavic linguistic family which is not understandable by Germanic speaking people and hard to learn. The very close Hungary and its Hungarian language are even more difficult to learn for German natives and the Finno-Ugric language, being a branch of the Uralic language family, has no obvious connection to any Slavic, Germanic or Roman languages but can only be found in Hungary, Finland, Estonia and parts of Russia. Most of the former Empire had been absorbed by the Eastern side of the bipolar cold war Europe and economic relations had been negatively influenced by governmental regulations mostly from the East.

With its neighbourhood to Italy and Switzerland and some historic connections to Romania and Venice which had in a minor grade been part of the Austrian-Hungarian Empire, Romance languages are of importance to the Austrian partnerships.

English as a lingua franca in the European Union and established medium for trading purposes has the potential of improving the intercultural exchange in the region although none of the countries mentioned have any linguistic or historical connection of importance to the English language or Great Britain. As the German language is by its roots connected to the English, Austrians, Germans and Swiss people have an advantage in learning English to their Slavic, Romance and Hungarian neighbours.

By studying lingua franca one may not forget that during the Napoleonic century, before and after, French has been the lingua franca in the European aristocratic society. Therefore it had made its influence in all the languages on the European continent. Longer ago but for a wider time span and much more intense the Latin language shaped all languages in the Roman hemisphere. This underlines a basic connection between European languages and Romance languages.

All these interactions and limitations that happen beyond the background of linguistic barriers are easy to talk about but very hard to measure in numbers. For the languages that are important for this thesis a theoretical background about roots, descent and historical development is available. Most of the European languages are Indo-European languages and similar in their grammatical and semiotic structures. There are three other language families in our dataset: In Turkey most of the population speak Turkish which is a Turkic language, mostly spread from Turkey over central Asia to western China. Further Hungarian and Finnish are Uralic languages that can only be found in Estonia and else east of the Ural mountains and thus make these two countries quite isolated in Europe in the linguistic context.

Japanese is a Japonic language and the Chinese languages, especially the mostly spoken and official Mandarin Chinese, are Sino-Tibetean languages. Japanese and Chinese are assumed to be related by a descend of many millenia but all relational theories are still controversial. A linguistic distance variable will underline that the two languages are quite far away from each other.

2.2 Literature Review

While there exists an intuition of how close one language is related to another, it is very difficult to put this relation into a quantitative frame. This is a summary of studies that tried to address the measurement of linguistic distance within the gravity model.

In the context of the gravity model a linguistic variable has been established for long. The simplest way to quantify linguistic distance is to create a binary variable showing 1 if two countries have the same official language or 0 if they do not. But is that representative if Slovenia and Croatia by that measure do not share one language while Switzerland and France do? Slovenian and Croatian are two different official languages which differ only slightly while French is official language in both France and Switzerland while only a minority in Switzerland speaks French as a mother tongue. Anderson and Van Wincoop (2004) account a 7% tax equivalent language barrier on the base of a binary variable in their analysis of trade costs.[2]

To find a measure which is more accurate, broader data is needed. An important factor here is the existence of a lingua franca in a trading area. English as such is the most important lingua franca today and bad English assumable to be the most spread language in the world.

Apart from the mostly used binary common language variable there are several other ways to quantify linguistic distance and put more weight to the difficulty of translation and other language related transaction costs.

Chiswick and Miller developed a measure for linguistic distance to English which is calculated from results of American students' foreign language tests. If for the American students it is more difficult to learn Japanese than French, Japanese is seen as more distant from English than French. Based on language scores for 43 languages and a fixed time period of 16 and 24 weeks, their variable has a range of 2. It ranges between 1 as a maximum distance, like between English and Japanese, and 3 as a minimum distance, like between English and Norwegian. They use this measure to show that linguistic closeness has a positive impact on English skills of immigrants in the United States.[6] Similar studies show how immigrants in Canada with a Romance language as mother tongue tend to settle in Quebec.[5] Furthermore, immigrants in Israel with Arabic linguistic background have the best Hebrew skills.[4]

Hutchinson used this measure to show that "...the further a country's primary language is from English, the lower trade will be between the United states and that country."[9] His sample consisted of 36 countries that do not have English as an official language. Due to their post world war II history and linkage to the United States Japan and South Korea were treated as special cases in the study as Chiswick and Miller's linguistic distance variable is very low but trade relations are very high.

Lohmann introduced a language barrier index (LBI) which uses 139 linguistic features in 10 categories from the World Atlas of Language Structures (WALS)[24], a phonological, grammatical and lexical property database constructed by 40 authors. All properties of the two languages are compared to each other. Whenever a property differs in the two languages 1 is added to a sum. The average over all these dissimilarities results in a measure between 0 and 1, relating 1 to no similarity and 0 to extreme similarity. Using this index in a gravity equation based on Rose (2004)[15] data Lohmann finds that a 10% increase in the Language Barrier Index might well cause a 7-10% decrease in bilateral trade flows.[12]

Isphording and Otten use a linguistic distance variable that is based on the Automated Similarity Judgement Program (ASJP)[18] which has been created by the Max Planck Anthropological Institute. By counting transforming actions between the symbols of two words with the same meaning in an international phonetic alphabet a variable is created which is fully independent of language learning skills and linguistic judgement. A number between 0 and around 110 is created with 0 meaning "no difference" and the higher the number the higher the difference between two languages. Nevertheless this variable resembles quite well an intuitive distant judgement. Moving from the lower quartile (As an example the distance between English and Russian) to the upper quartile (English and Japanese) trade between countries decreases by around 4,1%.[10]

Ku and Zussman conduct a study about the role of English as a lingua franca in international trade. They use a database that collects mean national scores in the Test of English as a Foreign Language (TOEFL) for more than 100 countries and 30 years. Integrating these average scores in their gravity equation they find significant results that higher English proficiency increases trade relations of countries where one of them does not necessarily need to have English as an official language. In a two-stage regression an English proficiency variable is estimated which depends on both countries' average TOEFL scores.[11]

Melitz compares two approaches: The effect of direct conversational possibility and the effect induced by translational possibilities. Further English as a lingua franca is analysed not to be any more effectively promoting trade as a conversational vehicle than any other European lingua franca. Non-European languages though are to his findings not so effective. He also finds that literacy and diversity of spoken languages in a country boost foreign trade. Melitz points out that the assumption of a language effect implies either no translation possibility or a free translation system. Hence in the case of direct communication research he assumes that translation is cost free. As first linguistic distance measure between countries he uses "open-circuit communication", a variable which is 1 if there exists a language which is either officially or widely spoken in the two mentioned countries. In the case of "widely spoken" he means that the language is spoken by more than 20% of the population in each of the two countries. Otherwise this binary variable is 0. For the other measure Melitz uses a variable called "direct communication". Here the percentage that can speak a specific language in both countries is used for a scale. 4% is a minimum and thus limits the number of languages for the 157 countries to 29. Percentages of speakers in the countries are multiplied and summed up for all languages that are spoken in both countries by more than 4% each. Melitz used the CIA factbook and Grimes (2000) as his data reference. [14]

The European Commission ordered a Special Eurobarometer to survey language skills of the European member states' citizens. Jan and Jarko Fidrmuc used this dataset from 2005 which includes native languages up to three foreign languages that people believe to be able to speak. They estimate "communicative probabilities", probabilities that two random people from two different countries can speak a common language, by multiplying average proficiency rates. They find that higher linguistic skills result in higher trade intensity. In the EU-15 the probability that two randomly chosen people from two different countries can communicate in English is 22 % and raises intra EU-15 trade by around 30%.[7]

To calculate a distance between all languages pairwise it is necessary to look at the languages themselves as there is a lack of data about skills of so many languages. One possibility is to reference language trees constructed by linguists. Apart from the fact that they sometimes differ, the problem of quantification is still not solved.

Here we want to focus on three measures, the first being the LDND based on the ASJP of the Max Planck Anthropological Institute and the other two based on the Eurobarometer for languages in the European Union.

3 The Gravity Model - Theoretical Backgrounds

This chapter is supposed to be a short review of the theoretical backgrounds of the gravity model in economics. The gravity model in general is a concept formulated by Isaac Newton who used it to describe the force that drags things downward on earth. The gravitational force (F) between two masses equals the product of these two masses, multiplied by a gravitational constant and divided by the square of the distance between these two masses:

$$F = g\frac{M_1 M_2}{dist_{12}^2}$$

In 1962 Jan Tinbergen used this model to analyse world trade flows and found that the Gross National Products (GNP_i, GNP_j) and distance between country i and j have a hugely significant impact on trade flows between these countries. His goal was to use this model of a normal international trade flow and find reasons for eventual discrepancies in real trade flows hinting at discrimination or preferential treatment in trade. He already implemented dummy variables to quantify deviations caused by trade agreements such as the Commonwealth or the Benelux countries and looked at the preference of trade with neighbouring countries. Further a Gini coefficient for export diversification was integrated to show that diversification has a positive influence on exports.[16]

Convinced by the immense empirical results, today it is common belief that this concept is well suited to describe any form of flow between subjects. It has become an often used technique to model flows in economics and has gained a huge amount of theoretical foundation in the recent 25 years. So, the gravity model is used to explain trade, immigrant and capital flows between countries or other entities and in each of these models it animates a huge amount of specified analyses. In this thesis the impact of linguistic distance variables on trade is reviewed.

Now if we are including elasticities to the gravity model the equation can be written this way:

$$F_{ij} = g\frac{M_1^{\beta_1} M_2^{\beta_2}}{d_{ij}^{\beta_3}}$$

.

The transformation from this basic gravity model to a linear regression can be shown by taking the logarithm of both sides of the equation.

We can express that in the following:

$$ln(F_{ij}) = ln(g) + \beta_1 ln(M_i) + \beta_2 ln(M_j) + \beta_3 ln(dist_{ij}) + u_{ij}.$$

The sign of the coefficients should express the positive or negative relationship between the dependant and the explaining variable and for that reason the signs in the general not estimated form are neutral which means positive here. We expect β_3 to be negative.

As a first example of an economic equation we can look at a model by McCallum. He showed the influence of the border between the United States and Canada on trade between these countries with a gravity model using export and import observations of 10 Canadian provinces and 30 American states in 1988 accounting for 90% of bilateral trade. The study was released at the same time when the North Atlantic Free Trade Agreement (NAFTA) came in place and predicted that east-west trade in Canada will be drawn towards a north-south pattern due to the diminishing influence of the border. He first shows that Canadian provinces trade more with each other than with states in the USA and thus implies that national borders still interfere trade. Then, by pretending there was no border, he predicts higher province-to-state trade. Overall McCallum was predicting a shift in trade patterns but no extraordinary increase in total trade. He stated this equation:

$$x_{ij} = a + by_i + cy_j + ddist_{ij} + eDummy_{ij} + u_{ij}.$$

Here x_{ij} is the logarithm of trade flows between region i and region j. y_i and y_j are the logarithms of each region's GDP and $dist_{ij}$ resembles the logarithm of the distance between the capitals of the regions. u_{ij} is the error term and the dummy variable is 1 for interprovincial trade and 0 for province-to-state trade. a is a constant and the coefficients $b-e$ should be interpreted as elasticities between each variable and the dependent variable. [13]

Anderson and van Wincoop have found some missing theoretical background on this study and put the whole gravity equation into a firmer model. They established a gravity theory which implies multilateral resistance terms. These multilateral resistance terms smaller McCallums border effect due to the former omitted variable bias. Additionally McCallum only took intra-provincial trade in Canada into account whereas Anderson and van Wincoop took inter-state trade in the USA as there comparative measure. This showed the overestimated coefficient due to the fact that Canada is a much smaller economy than the US. They redid the study for the year 1993 and resulted in a border effect of 20-50% instead of 2200% as in McCallums estimations. As McCallum they are conducting a comparative static exercise, this time being able to solve the general-equilibrium model before and after removing the border.[1]

The multilateral resistance terms explained a lot of the gravity equation's theoretical background what lay in the dark before . Every country is assumed to have an individual characteristic which is determined by many country specific properties. That by itself influences the country's trade with any potential partner in international trade.

They can be included into the gravity model as follows. The multilateral resistance terms are Π_i and P_j. Π_i measures the multilateral resistance that the exporting country, marked with i, has to all other countries. The higher that resistance term the more difficult it is for the exporting country to export into other markets. P_j measures the multilateral resistance the importing country, marked with j, has to all other markets. The higher P_j the more difficult it is for country j to import from other countries.

$$(\Pi_i)^{(1-\sigma)} = \sum_j \left(\frac{t_{ij}}{P_j}\right)^{(1-\sigma)} \frac{Y_j}{Y_w}$$

$$(P_j)^{(1-\sigma)} = \sum_j \left(\frac{t_{ij}}{\Pi_i}\right)^{(1-\sigma)} \frac{Y_i}{Y_w}$$

Here σ is the elasticity of substitution between all goods and t_{ij} describes the bilateral trade distance which can consist of many factors that describe trade costs. Y_j, Y_i and Y_w describe country j's and i's GDP and the world GDP respectively. The following gravity model can be established:

$$X_{ij} = \frac{Y_i Y_j}{Y_w} \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{(1-\sigma)}$$

So the bilateral trade volume does not only depend on the bilateral trade resistance which is reflected by t_{ij} but also by the outward multilateral resistance Π_i and the inward multilateral resistance P_j. Some effects can be recorded. If trade costs or barriers between the importing country j and the rest of the world rise, expressed in a rise of P_i, the relative price of exporting country i is decreasing which makes trade between these two countries more attractive. If on the other hand the outward multilateral resistance Π_i of exporting country i is rising, meaning that it has become more difficult for country i to export to the rest of the world, and the bilateral resistance to country j remains constant then bilateral trade will increase.[2]

Furthermore, Anderson and van Wincoop established a categorization of trade costs and justified why the gravity approach is a reasonable one to address the influences of trade costs. They calculate a 170% total trade barrier as an ad-valorem tax equivalent for a representative rich country. These barriers consist of transportation, the border effect and commodity distribution. Anderson and van Wincoop trace many barriers back to economic policy instruments such as tariffs, quotas and exchange rates as well as the more essential infrastructure investments, law enforcement, property rights, information, regulation and last but not least the main focus of this thesis: language. Policy related trade costs are assumed to make 10% of national income and many questions about international macroeconomics are tightly linked to the implications of trade costs. They argue that the gravity model constitutes the main connection between trade costs or trade barriers and the trade flows.[2]

In their study Anderson and van Wincoop are able to split the 170% trade ad-valorem tax into 2 main origins: 55% local distribution costs and 74% international trade costs. Further divided the 170 % for a representative industrialized country are 21% transportation costs, 44% border trade barriers and 55% retail and wholesale distribution costs. 44% border trade barriers can be broken down to 8% policy barrier, 7% language barrier, 14% currency barrier, 6% information cost and 3 % security costs.[2]

Haonan Wu focuses on the role of the physical distance variable and confirms that the distance coefficient in the gravity model measures "the relative unit transportation cost between short distance and long distance rather than the absolute level of average transportation cost."[8] He confirms that by repeating earlier findings for 1991-2006 data. Further he shows that the distance coefficient is not affected by technological advance. On the other hand he finds that with an increase in technology increases the sensitivity of the distance variable to the oil price changes. Differencing between agriculture, manufacturing and services the nonetheless decreasing behaviour of the three distance coefficients was rather strictly decreasing in the agricultural sector and shock behaviour in the other two sectors.

Based on Anderson and van Wincoop's model Baldwin and Taglioni [3] show that the main assumption of Anderson and van Wincoop, namely that the inner and the outward multilateral resistance can be assumed to be equal, doesn't hold for panel data and introduce a solution to this problem as well as pointing out three very common mistakes made in gravity studies which they call the gold, silver and bronze medal mistakes.

Although the problems in a gravity model estimation can be addressed with different dummy types such as time-varying country-pair dummies, time dummies and country dummies, the huge amount of data and the hence resulting amount of dummies are not always easily applied.

4 Data

To work with the data country names had to be replaced by the International Monetary Fund's International Financial Statistic's country codes. With these a panel data regression in Stata was conducted.

4.1 Data sources

The countries I included into this study are these 24 countries which have been under the 20 most important Austrian export partners in the last 30 years. Hereby importance is measured by export value from Austria to the destination country and additionally Austria is included as well. Countries that are included in this study are the following: Austria, Belgium, Canada, China, Croatia, Czech Republic, Denmark, Finland, France, Germany, Hungary, Italy, Japan, Spain, Netherlands, Poland, Romania, Russian Federation, Slovakia, Slovenia, Sweden, Switzerland, Turkey, United Kingdom, USA.

There are several sources which provide trade data. From Statistics Austria the data for Figures 4.1 and 4.2 of foreign trade volumes has been extracted. The countries are sorted in their size of trade volume with Austria.

In the Appendix (Section 7.2) we can see that over the last two decades the export and import shares of Eastern European countries such as Russia, Czech Republic or Slovakia have increased while shares of Western European countries like Netherlands, Belgium and Sweden have decreased. Further while trade with China and Turkey increases monotonous, that with Japan and the United Kingdom rather decreases. As other countries also take a monotonous increasing share, we can see that the number of important trade partners is increasing as well.

For our regressions we take UN Comtrade data [23]. It is import data, always reported by the importer. It is total trade value in current international Dollar . The dataset should contain $25 * 24 * 16 = 9600$ observations. For some reasons there were

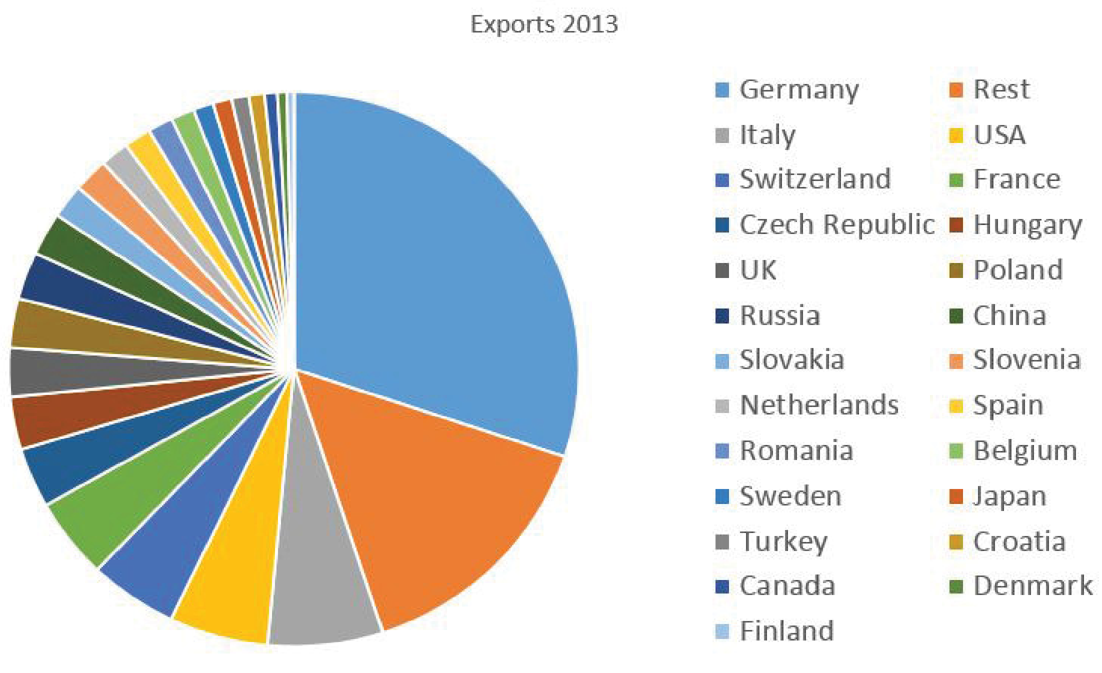


Figure 4.1: Export shares Austria 2013

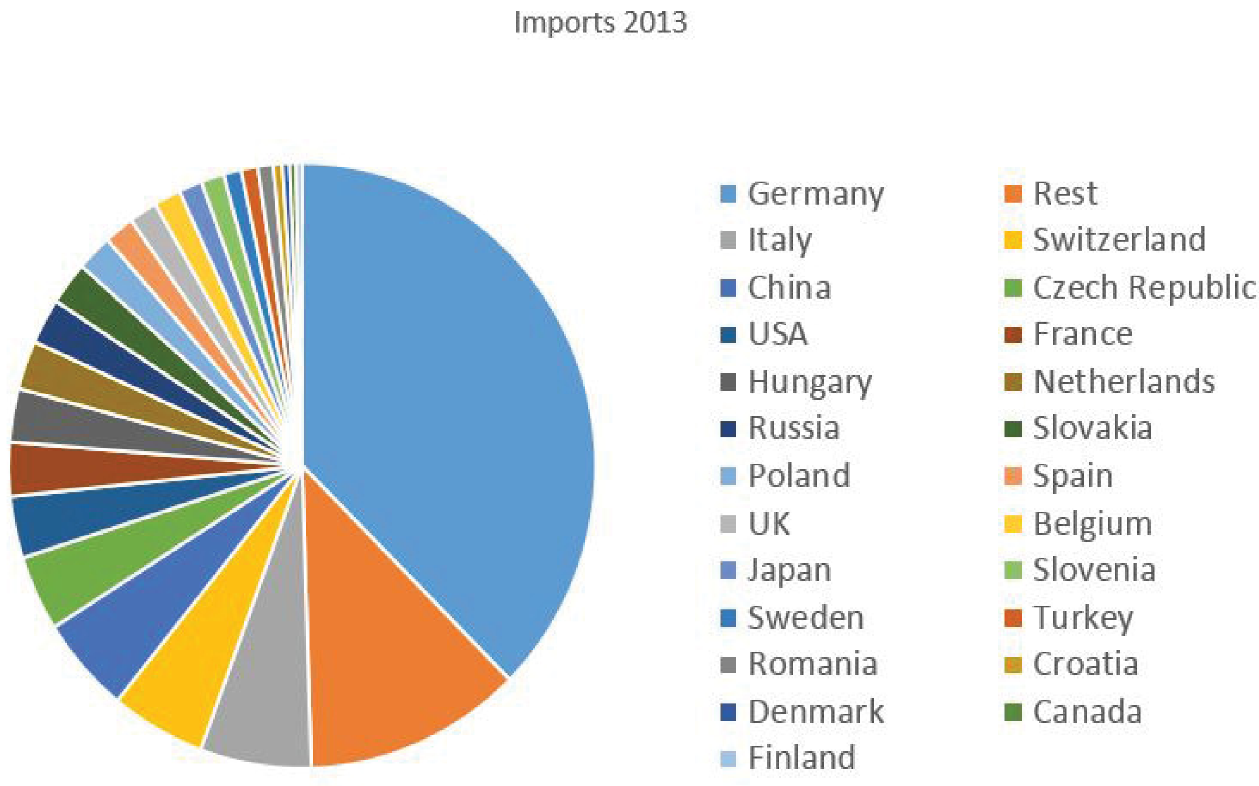


Figure 4.2: Import shares Austria 2013

70 self trade observations (for example Slovenia Slovenia) and two missing observations between Sweden and Turkey in 2001 and 2002. As the trade volume between Sweden and Turkey is increasing I estimate these two missing values by a linear approximation. For 1997 and 1998 trade data for Belgium and Netherlands has been published together. As Luxembourg is not that big and Belgium is still one of Austria's most important trade partners without Luxembourg I ignored this fact and treated the 1997 and 1998 data as Belgian trade data.

The Centre d'Ètudes et d'Informations Internationales (CEPII [19]) provides a package of physical distance variables which are well suited for gravity studies. The GeoDist variables are taking economic centres of countries into account by weighting the distances of the biggest cities with these cities' population share on total population. The weighted distance variables are calculated as follows, proposed by Head and Mayer (2002):

$$d_{ij} = \left(\sum_{k \in i} \frac{pop_k}{pop_i} \sum_{l \in j} \frac{pop_l}{pop_j} d_{kl}^{\theta} \right)^{\frac{1}{\theta}} .$$

Population and distance data is used from a Canadian world population counting initiative called "population clock". [21] Here the variable "pop_k" is used for the population of an agglomeration k in country i. θ is a measure for sensitivity of trade flows to the bilateral distance d_{kl}. It is set to 1 in *distw* and to -1 in *distwces* .

GDP is measured in current Dollars, converted at current exchange rates.

To compare results, all regressions are repeated with an often used dataset from Andrew Rose. [22] I reduced the existing huge dataset and deleted all the countries that are not part of the selection based on Austria's foreign trade relations. The data is an unbalanced panel that reaches from 1948 to 2000.

4.2 How can we measure linguistic distance?

Isphording and Otten [10] suggest three measures which they implement in a dataset published by Rose (2004) covering 178 countries over 52 years and compare regression results of the impact of these three measures on international trade.

The first and most common measure is the introduction of a bilateral dummy variable which equals 1 if two countries share the same language and otherwise 0. This is commonly used but very tricky to implement. Here linguistic difference is a binary variable. So do Austria an Switzerland share the same language? If yes, do Austria and the Netherlands share the same language? I would say no, but I am sure that communication between these two languages is much more intuitive than communication between German and Hungarian speaking individuals. That is reason to find a broader scale for evaluation of linguistic distance.

There are measures based on test scores as these of Chiswick and Miller (1999) who use data of American language student's test results. The problems for these measures is the assumption that people everywhere have the same problems on learning languages as Americans do. Similarly Ku and Zussman (2010) interpret the international TOEFL (Test of English as a Foreign Language) results to get a comparable measure. This unfortunately only represents distances languages have in comparison to English.

In the following three measures are presented that treat each language equally.

4.2.1 Linguistic Distance Normalized and Divided

One such measure is the Automatic Similarity Judgement Program (ASJP) developed by the Max Planck Institute of Evolutionary Anthropology. It implements a Levenshtein distance on the phonetic transcription of similar words in different languages. If we have two chains of symbols which we will call words, the Levenshtein distance calculates the distance between these words by counting the number of operations that are needed to transform one word into the other. Allowed operations here are substitutions, insertions or deletions.

As an example this shall be applied to the phonetic transcription of the German word "Sonne" (zɔnə) and the English word "sun" (sʌn) : to change z into s is one substitution, as well as changing ɔ into ʌ. By deleting ə, the transformation is completed and the Levenshtein distance is calculated as 3. It is possible to show that this transformation is a bijective function on the set of phonetic words, such that for the calculation it doesn't matter from which word we begin the transformation.

This transformation is repeated for each country on a so called Swadesh list which consists of 40 words that by Swadesh (1952) are suggested to be included in all existing languages. The only problem that occurs now is that different words have very different length and thus the measure is biased. To cope for that these distances are divided through all other Levenshtein distances that result from possible word comparisons:

$$LDND_{AB} = \frac{\sum_i (d_{ii})/n}{\sum_{i \neq j} (d_{ij})/(n(n-1))}$$

LDND stands for Linguistic Distance Normalized and Divided. Here d_{ii} describes the Levenshtein distance of word i between the languages of country A and country B. These distances are summed up and divided by the number of countries n. In the denominator all possible combinations of words with different meanings and their averaged distance ensures that any random similarity between phonetic inventories doesn't bias the measure.

Here we have the LDND values from the Max Planck Anthropological Institute's Swadesh-list language database ASJP (Automated Similarity Judgement Program). The values have been computed with a program by Eric Holman (2011).

As Belgium's population consists of 60 % Flemish and 40% French speaking people we weight the values of these languages by these percentages and use the result as their LDND value. For Switzerland the Bernese dialect has been chosen to represent the majority of Swiss dialects of the German language. The French minority of around 20 % has not been considered here.

A table of the 21 languages and their LDND values as calculated with Holman's program can be found in the Appendix as well as a scatter plot of the 600 values.

4.2.2 Communication Condition Indicator

Jan and Jarko Fidrmuc used the Eurobarometer 243, "Europeans and their languages", a study ordered by the European Commission to estimate probabilities about the ability of two randomly chosen individuals from two countries to communicate with each other in predefined languages.[7] While they used the report from 2005 a new report has been published by the European Commission in 2012.[17]

Based on the probabilities given in the Eurobarometer from 2012 I follow Jan and Jarko Fidrmuc and calculate communicative probabilities that show with what probability two individuals from two different countries can talk to each other in a specific language. Then I form an indicator by adding up these probabilities. The non-EU countries are dropped in this constellation. In the Eurobarometer we can find percentages of how many people in a EU country speak the official languages as a native language.

Further there is a survey included which gives information about percentages of how many people are able to speak other languages which are not their native language. In the Figure 4.4 this information is listed up to a third language that people can speak additionally to their native language. This table is from the Eurobarometer and includes 24 EU countries. By looking at the first rows in nearly all countries we can see that in close to all countries English is the most often mastered non native language. This table underlines the role of English as an in fact lingua franca in the European Union. But under the impression of so many countries in which the majority is not able to speak English (we have to ignore countries with native English tongue like UK here) we may not overlook that other languages still play an important role in communication between countries, their citizens and thus their economies.

The following picture 4.4 shows us the percentages of how many people in a country feel able to hold a conversation in at least two non native languages.

Here I'm going to show how I calculate the communication condition index for which I use the Eurobarometer's language ability frequencies. It is as if we calculate the communicative probability that two random individuals from two different countries i and j can speak the same language. The only assumption that contradicts the concept of a probability is that I assume that individuals with matching languages in two countries are disjunctive in respect to languages. Here I mean that due to lack of further information on which individuals speak which language, two individuals from country i and country j that can both speak language l_1 and l_2 are counted double in the indicator. Thus there is theoretically no upper limit for the CCI.

This can be justified by the assumption that if two people can speak more than one equal language, their communication will be even better. Let us imagine for example an Austrian and a Dutch person that both speak French and work on a project together. The Dutch is very good in German and so they might be most comfortable in negotiating in German. If they now recruit a third person from France that can't speak German the team will have to change the language. It is well possible that all three team members speak English and the team will further communicate that way. If only one of the three

D48a Thinking about the languages that you speak, which language is your mother tongue?

	State language(s), official languages that have an official status in the EU
BE	Dutch 55%, French 38%, German 0.4%
BG	95%
CZ	98%
DK	96%
DE	87%
EE	80%
IE	English 93% , Irish 3%
EL	99%
ES	Spanish 82%, Catalan 8%, Galician 5%, Basque 1%
FR	93%
IT	97%
CY	95%
LV	71%
LT	92%
LU	Luxembourgish 52%, French 16%, German 2%
HU	99%
MT	Maltese 97%, English 4%
NL	94%
AT	93%
PL	95%
PT	95%
RO	93%
SI	93%
SK	88%
FI	Finnish 94%, Swedish 5%
SE	93%
UK	88%

Figure 4.3: mother tongues and official languages from the Eurobarometer 386 report

BE		
	English	38%
	French	45%
	German	22%
BG		
	English	25%
	Russian	23%
	German	8%
CZ		
	English	27%
	Slovakian	16%
	German	15%
DK		
	English	86%
	German	47%
	Swedish	13%
DE		
	English	56%
	French	14%
	German	10%
EE		
	Russian	56%
	English	50%
	Finnish	21%
EL		
	English	51%
	French	9%
	German	5%
ES		
	English	22%
	Spanish	16%
	Catalan	11%
FR		
	English	39%
	Spanish	13%
	German	6%
IT		
	English	34%
	French	16%
	Spanish	11%
CY		
	English	73%
	French	7%
	Greek	5%
LV		
	Russian	67%
	English	46%
	Latvian	24%
LU		
	French	80%
	German	69%
	English	56%
HU		
	English	20%
	German	18%
	French	3%
MT		
	English	89%
	Italian	56%
	French	11%
NL		
	English	90%
	German	71%
	French	29%
AT		
	English	73%
	French	11%
	Italian	9%
PL		
	English	33%
	German	19%
	Russian	18%
RO		
	English	31%
	French	17%
	Italian	7%
SI		
	Croatian	61%
	English	59%
	German	42%
SK		
	Czech	47%
	English	26%
	German	22%
FI		
	English	70%
	Swedish	44%
	German	18%
SE		
	English	86%
	German	26%
	French	9%
UK		
	French	19%
	English	10%
	German	6%

Figure 4.4: Second languages in the EU, Eurobarometer 386 report

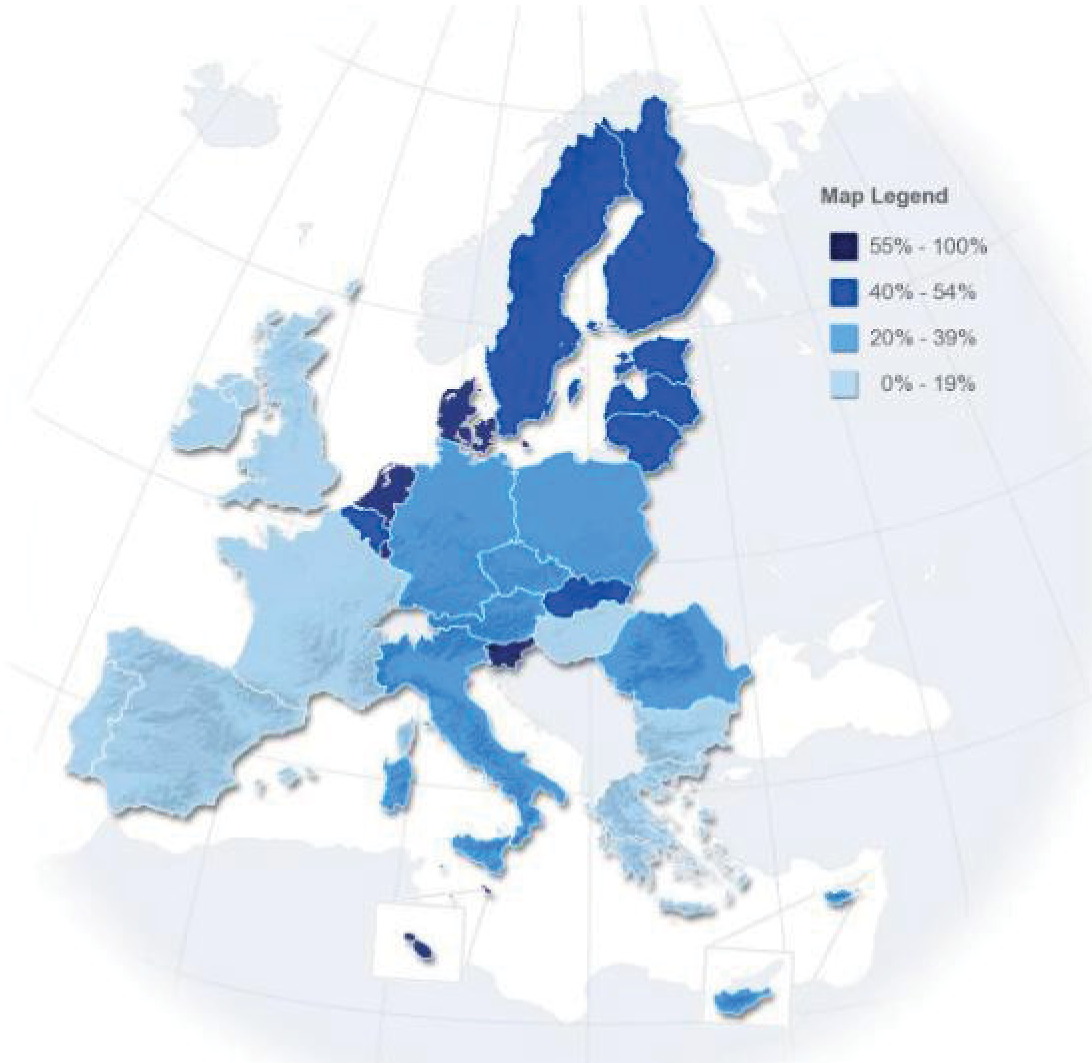


Figure 4.5: Map of people that are able to speak at least two more languages fluent next to their mother tongue, Eurobarometer 386 report

individuals now doesn't speak English they wouldn't be able to communicate if not the fact that the other two can speak French. This means that multilingualism has the advantage of increasing the probability of being able to communicate. For some reason there are people that are essentially important in special markets or technical fields who are uncomfortable with the English language. Here multilingualism makes sense and English as the one and only lingua franca in the future must fail as every other language as long as there are different countries with different official and native languages. Here multilingualism reduces translation and explanation costs immensely.

In the following way the index is constructed:

First we add the percentages of people that speak language l_k natively to the percentages of people that speak l_k non natively for all languages in a country. Here we only take these people that can speak the language good enough so they can conduct a conversation in that language.

$$k \in \{Austrian, Belgian, ...\} = L$$

$$i, j \in \{AUT, BEL, CAN, CRO, CHI, CZE, ...\}$$

We use the relative frequency of people in country i that speak the official language k as their mother tongue $N_{k,i}$ and add the relative frequency of individuals who can speak language k but cannot refer to it as their mother tongue $S_{k,i}$ to receive the total relative frequency of individuals in country i who can speak language k, $T_{k,i}$:

$$T_{k,i} = N_{k,i} + S_{k,i}$$

To calculate the index for the country pair i and j we first calculate the probability that two individuals from country i and country j can both speak language k, $P_{k,i,j}$:

$$P_{k,i,j} = T_{k,i} \cdot T_{k,j}$$

To finish the index we sum all these probabilities up over all languages.

$$CCI(i,j) = \sum_{k \in L} P_{k,i,j}$$

4.2.3 Multilingualism Index

It would be interesting to know if a multilingual society already fosters a country's trade. Sometimes the question comes up if learning languages has any sense in times of good English skills and constantly improving automated translation programs. Although the internet seems to connect people all over the world, an extreme diversified and country specific development takes place. Different cultural spheres keep being separated in social networks, news channels and trading platforms. This might be related to the linguistic backgrounds of people and the assumed freedom in the net has not erased the advantage of asymmetric information through language skills.

By adding all percentages of spoken languages in a country I construct an index that thus only measures how "much" language is spoken in a country. For this I use information from a website[20] that extracted the relevant data from the "Europeans and their languages" - Eurobarometer in 2012.

Here is an example of how this is done: If in country i the languages $\{x, y, z\} \in \Lambda$ are spoken by the percentages $a_i(x), b_i(y)$ and $c_i(z)$ and in country j the languages $\{z, v\} \in \Lambda$ are spoken by $d_j(z)$ and $e_j(v)$, be Λ the set of all relevant languages, then the MI index is calculated by adding all percentages:

$$MI_{sum}(i,j) = a_i(x) + b_i(y) + c_i(z) + d_j(z) + e_j(v)$$

The index MI_{single} is only constructed for countries and not for country pairs such as MI_{sum}. For country i the index would be

$$MI_{single}(i) = a_i(x) + b_i(y) + c_i(z).$$

4.2.4 Weighted LDND

At last I try to weight the LDND with the CCI to integrate the language learning results of the CCI into the more static linguistic conditions of the LDND. The LDND values are divided through the CCI values, so that a negative effect is to be expected.

5 Regression Results

The model is based on the typical gravity structure where the bilateral trade volume is explained by the two countries' GDPs, their bilateral physical distance and a vector of other possibly relevant variables, including the here looked at linguistic variables, subsumed under the gravitational "un-constant" g:

$$ln(X_{ijt}) = ln(g) + ln(Y_{it}) + ln(Y_{jt}) + ln(d_{ij}) + \epsilon_{ijt}$$

Here g consists of different variables, depending on the regression. In most regressions bilateral variables such as "have a common border", "were the same country once", "common colonizer post 1945" and "have ever had a colonial relationship" were included in g.

In the following tables different regression results are shown in the columns. An R means that a pooled regression has been conducted by taking the reporting country as group variable. Here a general trend can be observed if the variables show any significant effect. Whenever an a is in the column, the dataset from Andrew Rose's homepage going from 1948-2000 was used. Where is put the letter b the later data from 1997-2012 is used. In $R1$ and $R2$ a fixed effects and a random effects regression has been conducted. With a Hausman test we can see that the fixed effects model should be preferred but due to the huge datasets the coefficients change only minimally between the two models. In $R3$, a set of variables has been inserted, such as "border", meaning the two countries have a common border, "comctry", meaning the two countries have once been the same country in history, and "colony", meaning one of the countries has once been colonized by the other. "lndl" means that at least one of the two countries is landlocked and "comcol" that the two countries had a common colonizer post 1945. In $R3$ a set of time dummy variables has been included.

If there is a P, then a random effects regression has been conducted with country pairs as panel group variable. In $P1$ this has been done without and in $P2$ with time dummy variables. The typical variable set has been included in both. Due to the fact that the variables we are interested in do not vary over time a fixed effects model cannot be applied to a country pair regression. Thus a random effects regression has been applied in that case.

All numbers that are put into the table are significant on a 2 % level. All coefficients of the linguistic distance variables, physical distance variables and GDP variables are significant on a 1% level. An "x" means that the variable has not been included in a regression. "om" says that the variable has been omitted due to collinearity. If there is a "/" inserted then the coefficient has not been significant on a 2% level. In the following tables the coefficients of the different variables in the different regressions are listed together with each regression's R^2. Further information about standard errors or p-values can be observed in the Appendix where all the Stata outputs can be found.

5.1 Linguistic Distance Normalized and Divided

As the gravity equation is theoretically supposed to be an expenditure equation we would like to have coefficients near to 1 for the GDPs. In the dataset from Rose the coefficients are closer to 1 than in the newer one. The physical distance variable is, as expected, negative through all regressions.

Table 5.1: Linguistic Distance Normalized and Divided

Var	R1a	R1b	R2a	R2b	R3a	R3b	R4a	R4b	P1a	P1b	P2a	P2b
R-sq	0.70	0.57	0.70	0.59	0.70	0.60	0.81	0.50	0.78	0.66	0.81	0.50
GDPrep	1.72	0.21	1.69	0.22	1.72	0.21	0.9	0.1	1.36	0.27	0.87	0.16
GDPpar	1	0.46	1	0.45	0.99	0.45	0.97	0.44	1.5	0.3	1.06	0.14
dist	-1.12	-0.77	-1.13	-0.76	-1.12	-0.6	-1.1	-0.58	-1.59	-0.29	-1.22	/
border	x	x	x	x	/	0.65	/	-0.67	-0.67	1.16	/	1.64
comcol	x	x	x	x	1	om	0.98	om	1.4	om	/	om
comctry	x	x	x	x	om	0.88	om	0.84	om	/	om	-1
lndl	x	x	x	x	/	x	-0.06	x	/	x	x	x
LDND	-1.17	-0.9	-1.16	-0.9	-1.11	-0.57	-1.13	-0.58	-1.21	-0.75	-1.23	-0.83

The coefficients for the LDND are all significantly negative. This can be taken as an argument that there are static differences in languages and thus in culture and society that have a negative impact on the amount of trade. Comparing the older and the new dataset it can be assumed that the effect is decreasing over time. There seems to be no regularity in the coefficients for borders, colonies and landlocked countries. Depending on the regression they are changing signs or are not significant. By including time dummies the effect of the LDND has been decreased.

5.2 Communication Condition Indicator

Table 5.2: Communication Condition Indicator

Var	R1a	R1b	R2a	R2b	R3a	R3b	R4a	R4b	P1a	P1b	P2a	P2b
R-sq	0.74	0.67	0.75	0.68	0.74	0.69	0.87	0.63	0.8	0.72	0.83	0.59
GDPrep	1.89	0.18	1.83	0.19	1.89	0.18	0.66	0.1	1.8	0.22	0.73	0.1
GDPpar	0.81	0.36	1.82	0.36	0.81	0.35	0.80	0.34	1	0.24	0.39	0.16
dist	-1.08	-1.5	-1.08	-1.49	-0.9	-1.08	-0.87	-1.05	-1.22	-1	-0.75	-0.78
border	x	x	x	x	0.3	0.81	0.32	0.84	/	1.16	0.79	1.5
colony	x	x	x	x	0.85	-0.6	/	-0.6	/	/	/	/
comctry	x	x	x	x	om	0.28	om	0.23	om	-0.78	om	-1.49
lndl	x	x	x	x	/	x	/	x	0.58	x	-0.39	x
CCI	0.8	/	0.76	/	0.8	/	0.8	/	0.66	/	0.61	0.29

The CCI is an indicator that gives a measure of how well two random people of two random countries can communicate with each other. In the newer data only the last regression was able to give a significant result. It seems that a better communicative condition increases trade. Yet this indicator has less impact than the static LDND measure. Still the higher the overlap of languages is that two societies are able to use the more they are willing to trade. This might be related to educational policy. Nations with similar languages and and a close tie in politics might increase schooling of each others languages. In some cases of smaller countries that lie close to a bigger country such as the Benelux countries to Germany and France or the Scandinavian countries to Germany and the UK cultural entertainment as TV and internet sources encourage people to learn another language. In this context it is possible to observe in Figure 4.4 that smaller countries like Slovenia or the Netherlands have much better foreign language skills than bigger countries like the UK or France. The possibility to swerve from one language to another simplifies communication and increases trust in trading.

5.3 Multilingualism Index

The MI is a measure for how multilingual a nation's society is. The higher the number of languages per person the higher is the MI. For some reason the MI coefficients are negative in the newer dataset. In the first dataset the coefficients are very high but the country pair regression with time dummies is not significant. The two different trends can also be observed in the scatter plots in the Appendix. It seems that for the newer data a multilingual society has a negative impact on trade. On the other hand the effect for the other dataset is immensely positive except for the last regression where the coefficient is not significant.

Table 5.3: Sum of Multilingualism Indices

Var	R1a	R1b	R2a	R2b	R3a	R3b	R4a	R4b	P1a	P1b	P2a	P2b
R-sq	0.82	0.71	0.82	0.72	0.84	0.73	0.77	0.65	0.87	0.76	0.81	0.64
GDPrep	1.63	0.18	1.6	0.19	1.59	0.19	0.51	0.11	1.67	0.23	0.84	0.11
GDPpar	1.09	0.35	1.1	0.35	1.15	0.35	1.12	0.33	1.18	0.24	0.49	0.12
dist	-1.28	-1.4	-1.28	-1.4	-1.08	-1.17	-1.09	-1.16	-1.15	-1.05	-0.88	-0.97
border	x	x	x	x	0.25	0.47	0.26	0.49	/	0.85	0.8	1.1
colony	x	x	x	x	0.27	-0.32	0.28	-0.32	/	/	1.16	/
comctry	x	x	x	x	om	0.14	om	/	om	-0.66	om	-1.42
lndl	x	x	x	x	0.26	x	0.19	x	0.68	x	-0.62	x
MIsum	5.24	-2.25	5.3	-2.27	5.99	-2.08	5.64	-2.16	8.51	-3.09	/	-3.84

Table 5.4: Multilingualism Indices

Var	R1a	R1b	R2a	R2b	R3a	R3b	R4a	R4b
R-sq	0.76	0.69	0.82	0.73	0.77	0.70	0.79	0.62
GDPrep	1.63	0.18	1.60	0.19	1.59	0.19	0.5	0.11
GDPpar	1.09	0.35	1.1	0.35	1.15	0.35	1.12	0.33
dist	-1.26	-1.41	-1.26	-1.4	-1.06	-1.17	-1.06	-1.16
border	x	x	x	x	0.26	0.47	0.26	0.49
colony	x	x	x	x	0.25	-0.35	0.27	-0.31
comctry	x	x	x	x	om	0.13	om	/
lndl	x	x	x	x	0.27	x	x	x
MI2	2.66	-1.11	2.67	-1.11	3.05	-1.02	2.88	-1.06
MI1	om	om	4.24	-2	om	om	om	om

5.4 Weighted LDND

Table 5.5: LDND x CCI

Var	R1a	R1b	R2a	R2b	R3a	R3b	R4a	R4b	P1a	P1b	P2a	P2b
R-sq	0.74	0.67	0.74	0.67	0.73	0.68	0.86	0.62	0.81	0.72	0.82	0.58
GDPrep	1.81	0.18	1.74	0.19	1.82	0.18	0.65	0.1	1.63	0.22	0.73	0.1
GDPpar	0.84	0.36	0.85	0.35	0.84	0.35	0.83	0.34	1.1	0.24	0.45	0.11
dist	-0.93	-1.46	-0.93	-1.46	-0.87	-1.06	-0.83	-1.04	-1.06	-0.97	-0.83	-0.73
border	x	x	x	x	0.14	0.73	0.15	0.76	/	1.12	0.66	1.46
colony	x	x	x	x	/	-0.52	/	-0.51	/	/	/	/
comctry	x	x	x	x	om	0.34	om	0.29	om	-0.77	om	-1.52
lndl	x	x	x	x	/	x	0.15	x	/	0.72	-0.48	x
LDCCI	-0.75	/	-0.73	/	-0.73	/	-0.73	-0.06	-0.67	/	-0.52	-0.29

The combination of the two measures lessens the results of a language effect. Still a correction of the LDND by the communicative potential keeps the negative effect of the overall linguistic bilateral situation. This can be interpreted in a way that the condition to be able to communicate with each other decreases the effect of language distance. So by learning another countries languages the static distance between these two languages and the resulting negative impact on trade can be decreased.

One of the two variables "comcol" and "comctry" are omitted in most of the regressions as their values are mostly the same in this dataset. The "colony" variable has not been included in these tables as it was never significant.

6 Conclusion

In all the regressions we can find that the linguistic variables have a significant influence on the trade volume. It is shown that language still influences trading behaviour in our times although the effects have decreased comparing the 50 years before 2000 to these 15 years before 2012. Sophisticated translation software and English language as a lingua franca decrease these effects over time but are not able to remove this linguistic trade barrier. Linguistic distance can be coped with by education and technology.

In case of the LDND variable we were expecting a negative sign for the coefficients. The farer the languages are apart from each other in the sense of a Levenshtein distance the higher this linguistic distance variable is. By calculating operations needed to change the sound of a word from one language into the other a distance is calculated that is independent of learning behaviour or test results. This variable can be seen as static because the languages and their official pronunciation does not change as fast as learning behaviour or language skills. We can see a highly significant negative impact of this static linguistic distance on the bilateral trade volumes between the countries. This variable can also be interpreted as a cultural distance manifested in language. Countries with a higher linguistic and thus cultural distance trade less with each other than countries with higher linguistic similarities.

The CCI coefficient is positive as the higher the CCI the closer the two countries are in respect to their linguistic skills. An indicator is created that measures how well two random individuals from two random countries might communicate with each other. Here a positive effect of overlapping language skills between nations can be found and it can be assumed that learning other European languages increases trade with the European partners that use these languages.

The MI shows that a multilingual society increases trade and learning languages of whichever kind can have a positive impact on trade. Yet, the results from the newer dataset seem to say the opposite. Nonetheless the overall result seems to show that linguistic distance, language learning behaviour of nation's societies as well as multilingualism have a positive impact on trade and the importance of linguistic diversity may not be underestimated.

Under these circumstances an increasing importance of integrating Hungarian and Slavic languages into the educational policy would be an appropriate policy for the Austrian schooling system. Due to geographical and historical ties trade between Austria and its Slavic and Hungarian neighbours is high in comparison to other non-Slavic western countries. By increasing the efforts of learning these languages the advantage Austria has in comparison to other western European countries could be strengthened and a huge part of eastern integration could take place.

Overall it can be stated that linguistic conditions still have an important impact on trade and economic relations. The effect of languages on trade decreases over time and technological advances as well as depersonalization of trade transactions will diminish the importance of person-to-person communication. Nevertheless the process of political and cultural integration and with that standards and needs of the countries will still be dependent on linguistic properties and skills. This might be even more important within the European Union than in the international trade environment.

By learning different languages and trying to be able to communicate in foreign languages native languages are not destroyed. The process helps raising the own language to another level. Still the tie between language and culture should not be underestimated and language learning should be increased in a magnitude of languages and not only in English. Learning a foreign language successfully increases the success of learning a further language and thus multilingualism is a useful education.

As software development and the integration of English into everyday language is progressing world wide there might be plenty of changes in this area of research which should be addressed in the coming decades.

Bibliography

[1] James E. Anderson and Eric van Wincoop, 2003. ”gravity with gravitas: a solution to the border puzzle”, The American Economic review, Vol.93, No.1 pp. 170-192

[2] James E. Anderson, Eric van Wincoop: Trade Costs, 2004, Journal of Economic Literature, Vol.42, No.3 p. 691-751

[3] Richard Baldwin and Daria Taglioni, 2006. ”Gravity for dummies and dummies for gravity equations”, Discussion Paper Series No.5850 International Macroeconomics and international trade

[4] Michael Beenstock, Barry R. Chiswick, Gaston L. Repetto: The effect of linguistic distance and country of origin on immigrant language skills: Aplication to Israel, 2001, International Migration Volume 39, Issue 3 p. 33-60

[5] Barry R. Chiswick, Paul W. Miller: Language choice among immigrants in a multi-lingual destination, 1994, Journal of Population Economics 7 p. 119-131

[6] Barry R. Chiswick, Paul W. Miller: Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages, 1994, IZA Discussion paper series, No. 1246

[7] Jan and Jarko Fidrmuc, 2009. ”Foreign Languages and Trade”, Brunel University west London, Working paper no. 09-14

[8] Haonan Wu, 2015: Re-visiting the distance coefficient in Gravity Model, Cornell University, arXiv:1503.05283

[9] William K. Hutchinson: ”Linguistic Distance” as a Determinant of Bilateral Trade, 2003, Department of Economics, Vanderbilt University, Working Paper No. 01-W30R

[10] Ingo Eduard Isphording and Sebastian Otten, 2013. "The Costs of Babylon—Linguistic Distance in Applied Economics", Review of International Economics, Wiley Blackwell, vol. 21(2), pages 354-369, 05.

[11] Hyejin Ku, Asaf Zussman: Lingua Franca: The role of English in international trade, 2010, Journal of Economic Behaviour and Organization 75 (2010) 250-260

[12] Johannes Lohmann: Do language barriers affect trade?, 2010, Economic Letters 110 (2011) 159-162

[13] John McCallum, 1995. "National borders matter: Canada-U.S. regional trade patterns", The American Economic Review, Vol. 85 No.3 pp.615-623

[14] Jaques Melitz: Language and Foreign Trade, 2002, University of of Strathclyde, CEPR Working paper no. 3590, 2006 draft

[15] Andrew K. Rose: Do We Really Know that the WTO Increases Trade?, 2004, American Economic Review, 94(1) p.98-114

[16] Jan Tinbergen, 1962. "Shaping the world economy: Suggestions for an international Economic Policy", "Appendix VI: An analysis of world trade flows", The Twentieth Century Fund

[17] TNS Opinion and Social.2012. "Special Eurobarometer 386 - Europeans and their languages", European Commission

Internet sources:

[18] The Automated Similarity Judgment Program (ASJP), www.asjp.clld.org

[19] Centre d'Ètudes Prospectives d'Informations Internationales, http://www.cepii.fr/

[20] www.languageknowladge.eu

[21] www.populationmondiale.com

[22] world trade dataset from Andrew Rose's study "Do we really know that the WTO increases trade?", http://faculty.haas.berkeley.edu/arose/recres.htm

[23]

[24] The World Atlas of Languages Structures Online, www.wals.info

7 Appendix

7.1 Linguistic Distance Variables

LDND	Belg	BERNE	DANIS	DUTCH	ENGLI	French	GER	SWE	ITA	ROM	ESP	CRO	CZE	POL	RUS	SLOVA	SLOVE	FIN	HUNG	TURK
Belg	.00																			
BERN	81.47	.00																		
DANI	80.55	77.53	.00																	
DUTC	46.48	69.12	67.21	.00																
ENGL	76.14	71.29	66.83	60.73	.00															
Frenc	46.48	93.81	93.88	92.96	91.55	.00														
GER	72.28	50.12	68.07	48.83	68.62	95.73	.00													
SWE	81.48	72.59	52.92	68.10	64.15	94.85	69.79	.00												
ITA	82.90	91.90	90.68	87.56	90.25	78.24	86.36	92.89	.00											
ROM	83.07	90.98	89.73	87.24	86.64	78.89	88.17	93.11	55.34	.00										
ESP	87.93	95.06	91.72	91.82	93.34	84.03	92.98	92.63	56.74	73.71	.00									
CRO	90.59	88.76	91.01	90.57	86.23	90.60	91.98	89.42	89.24	89.48	91.74	.00								
CZE	92.78	89.60	94.38	93.29	89.83	92.27	92.92	93.25	90.23	90.89	90.16	43.72	.00							
POL	94.46	93.78	97.12	94.71	95.02	94.20	96.50	97.63	91.84	93.88	91.04	55.20	44.05	.00						
RUS	92.54	91.77	94.86	92.87	92.13	92.21	92.04	96.65	95.47	95.39	95.78	57.48	56.69	57.64	.00					
SLOV	92.20	90.89	95.70	92.45	90.60	91.94	93.54	92.07	92.03	92.50	91.06	40.91	32.82	53.57	55.26	.00				
SLOVI	90.68	87.09	92.23	90.84	88.98	90.51	89.75	91.61	88.76	89.32	90.68	28.36	35.68	46.38	51.63	44.25	.00			
FIN	100.13	98.05	101.83	101.44	104.24	98.81	99.04	97.53	102.69	100.05	97.38	100.49	97.34	97.59	100.11	98.25	99.91	.00		
HUNG	97.63	96.59	98.31	98.33	95.19	96.93	98.33	97.73	101.23	99.81	101.97	94.37	94.31	96.19	97.67	96.13	93.93	85.04	.00	
TURK	100.34	97.85	101.79	101.31	101.54	99.36	98.37	100.94	99.04	98.20	98.31	98.73	97.05	98.77	96.84	99.60	96.06	98.04	94.21	.00
JAP	101.87	101.76	102.40	100.87	99.98	102.86	99.64	100.38	98.70	102.01	97.93	103.79	101.47	98.79	101.76	101.98	101.34	96.70	101.49	100.41
CHIN	100.27	99.75	101.86	100.11	102.17	100.42	100.93	101.47	97.04	101.02	95.50	99.32	101.66	101.37	100.53	100.74	101.13	99.61	100.90	100.70

	JAP
CHINES	98.98

Figure 7.1: Linguistic Distance Normalized and Divided

Country	AUT	BEL	CZE	DEN	FIN	FR	GER	HUN	ITA	NED	POL	ROM	SVK	SVE	ESP	SWE	UK
AUT	x																
BEL	0.57702	x															
CZE	0.3366	0.1362	x														
DEN	1.0649'	0.43208	0.3027	x													
FIN	0.6784	0.30632	0.216	0.7503	x												
FR	0.4428	0.93354	0.1143	0.3636	0.2838	x											
GER	1.3262'	0.54628	0.2967	0.9375	0.5666	0.4068	x										
HUN	0.3167	0.14122	0.081	0.2566	0.1724	0.1167	0.2908	x									
ITA	0.3531	0.262	0.0918	0.2924	0.238	0.2957	0.2128	0.0728	x								
NED	1.3492'	1.25874'	0.3495	1.1077'	0.7578	0.6633	1.2333'	0.1887	0.3524	x							
POL	0.4176	0.16796	0.1176	0.3731	0.2652	0.1401	0.3691	0.066	0.1122	0.4319	x						
ROM	0.2513	0.2589	0.0837	0.2666	0.217	0.279	0.1974	0.0671	0.2005	0.3283	0.1023	x					
SVA	0.3944	0.14808	0.7046	0.327	0.2216	0.1146	0.359	0.052	0.0884	0.3902	0.1276	0.0806	x				
SVE	0.8213	0.31828	0.2223	0.7048	0.4886	0.2553	0.7378	0.118	0.2006	0.8292	0.2745	0.1829	0.2458	x			
ESP	0.1606	0.0836	0.0594	0.1892	0.154	0.2132	0.1232	0.044	0.1826	0.198	0.0726	0.0682	0.0572	0.1298	x		
SWE	0.8795	0.45974	0.2712	0.9827	1.1045'	0.4347	0.7464	0.2215	0.3068	0.9847	0.3332	0.2819	0.2808	0.6166	0.1892	x	
UK	0.7921	0.54354	0.2736	0.871	0.6968	0.5625	0.6336	0.2125	0.3636	0.9797	0.3348	0.3361	0.268	0.6034	0.2156	0.8755	x

Figure 7.2: Communication Condition Indicator

country	MI
AUT	159.9
BLG	178.12
CZE	136.00
DEN	198.25
FIN	199.4
FR	143.52
GER	153.63
HUN	126.83
ITA	129.27
NED	183.03
POL	146.25
ROM	142.4
SVA	165.79
SVE	209.71
ESP	141.52
SWE	189.42
UK	133.02

Figure 7.3: Multilingualism Index

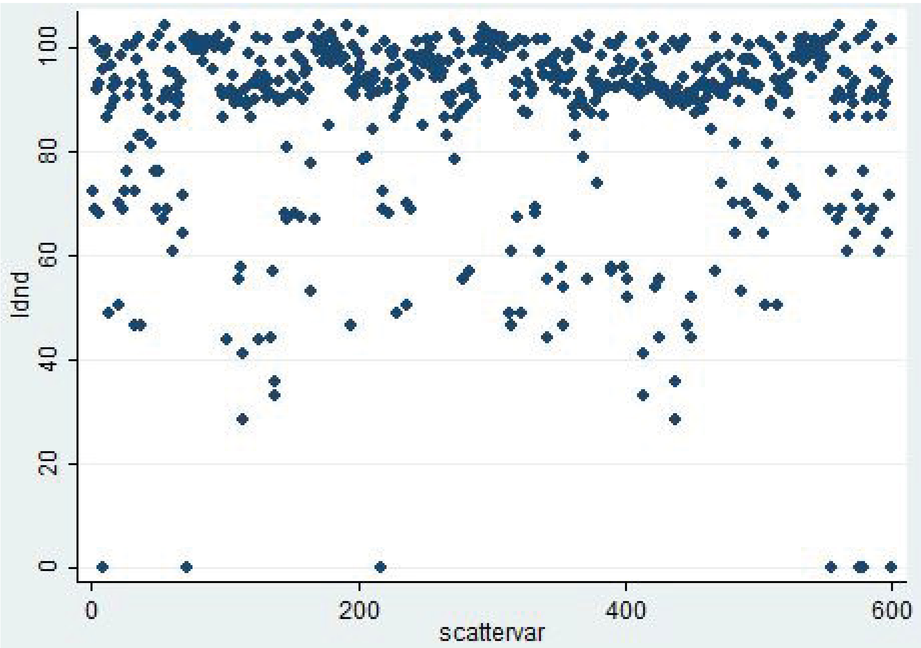
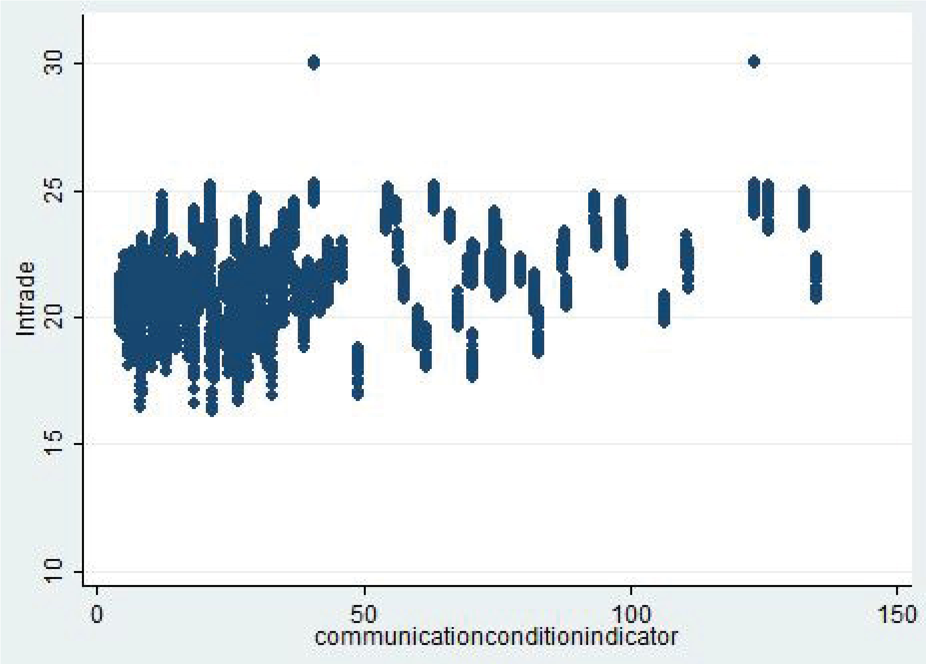
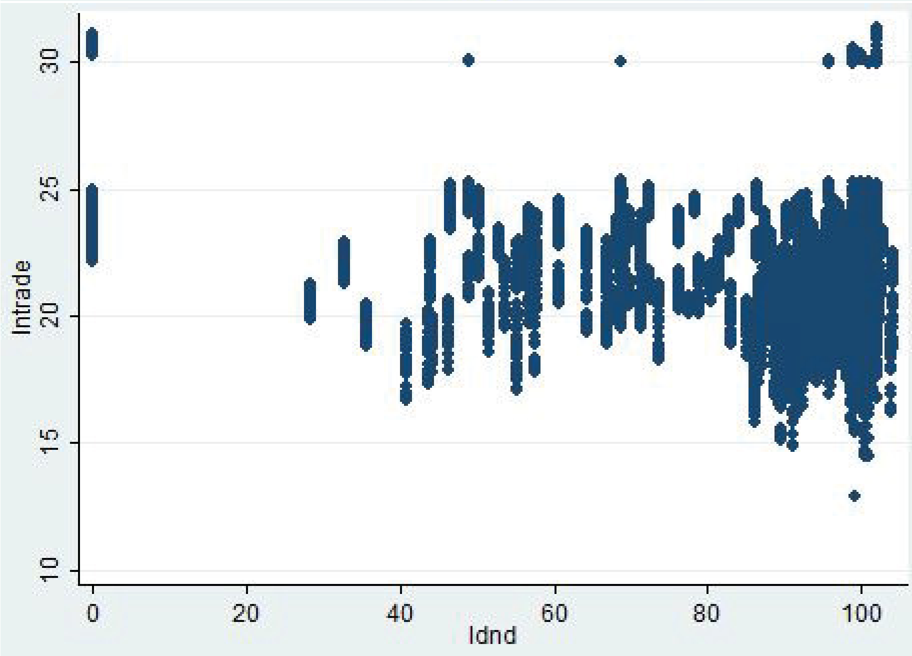
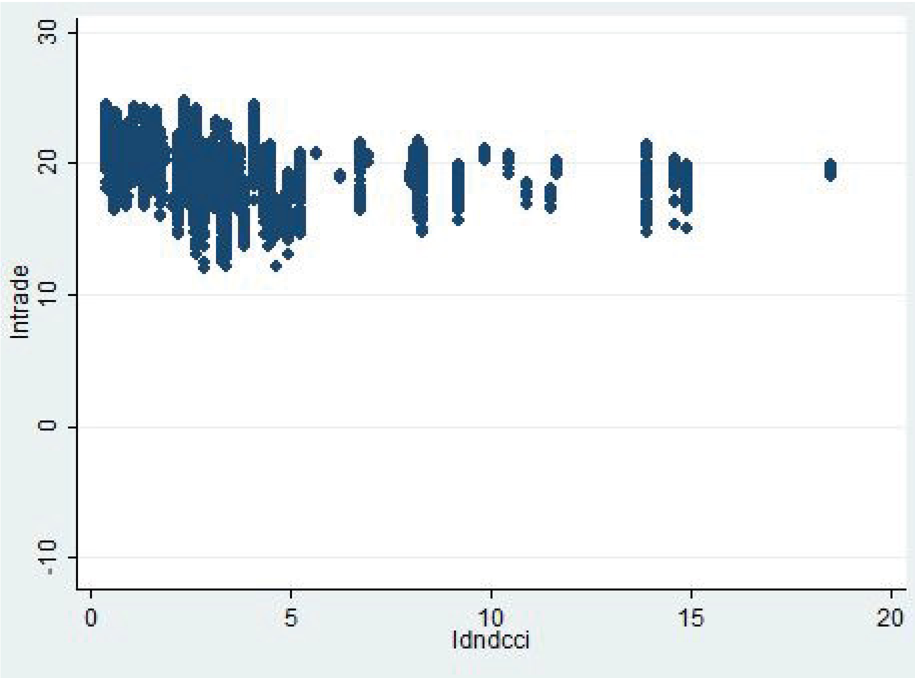


Figure 7.4: Distribution of the LDND. The measure becomes more sensitive in the range between 85 and 105.

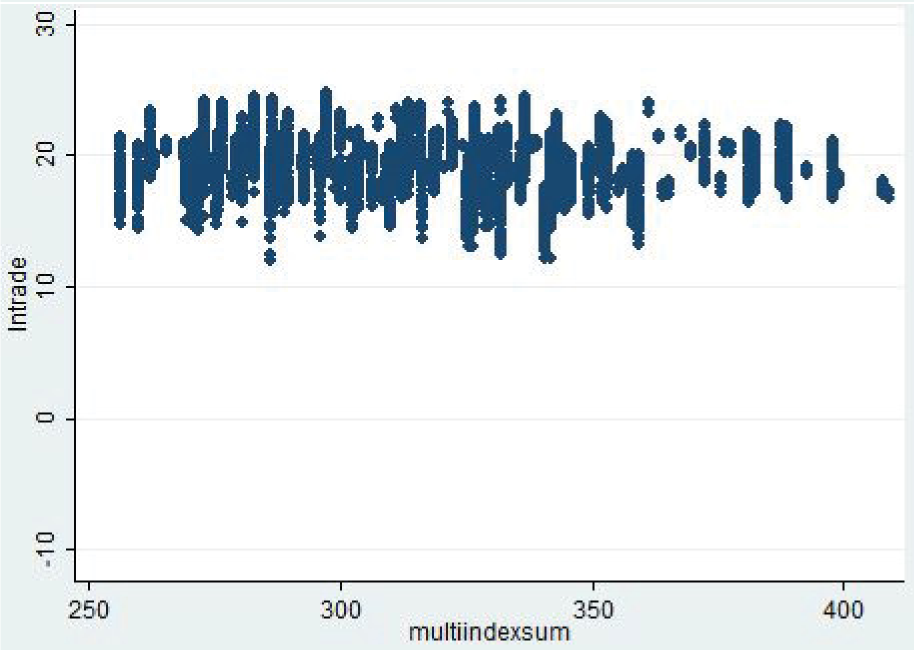
Plots of LDND, CCI, LDND/CCI and MI to ln trade Rose dataset:





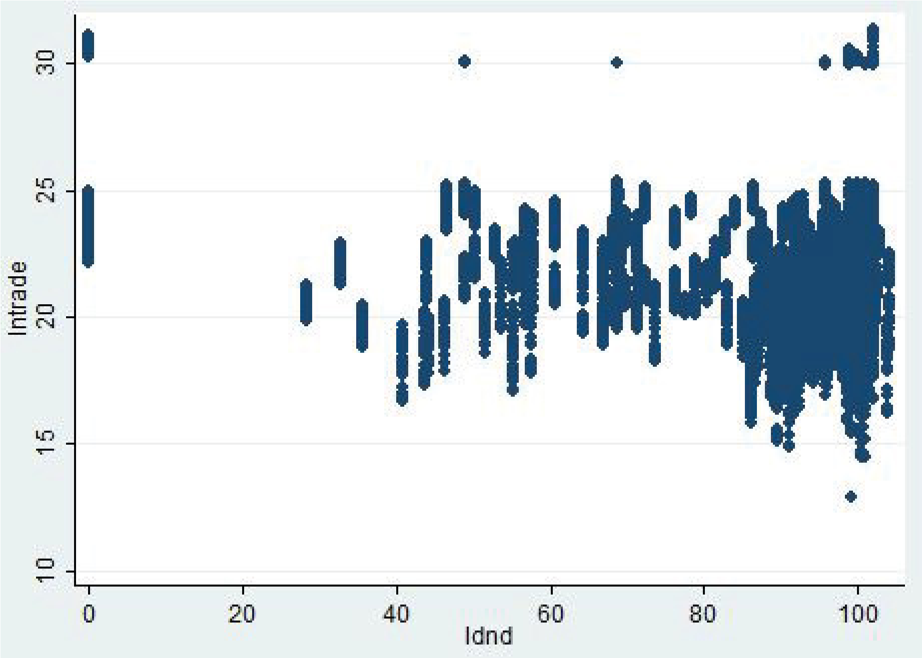


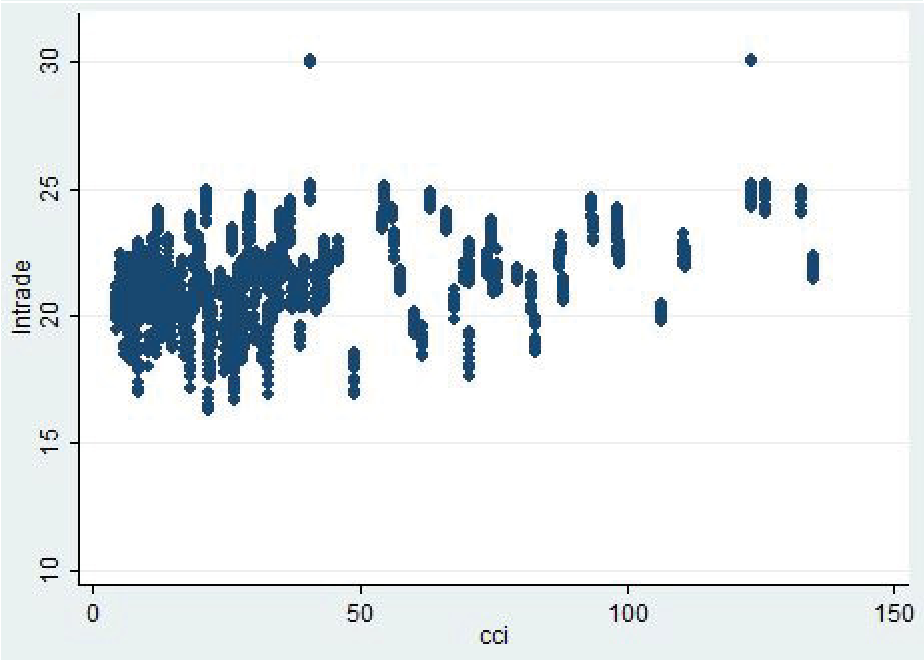
30
20
10
0
-10
lntrade
0
5
10
15
20
ldndcci

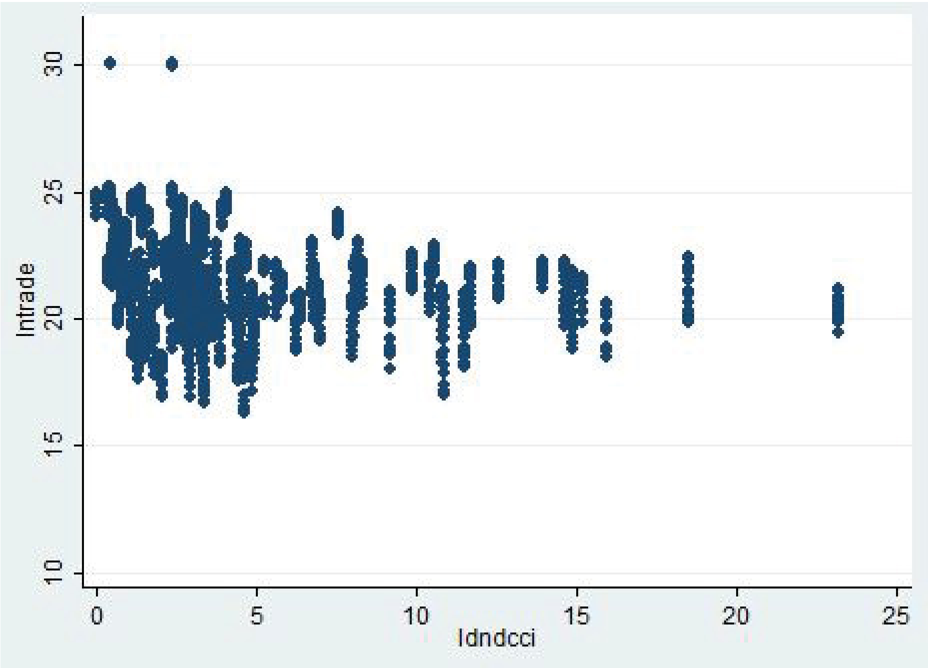


30
20
10
0
-10
lntrade
250
300
350
400
multiindexsum

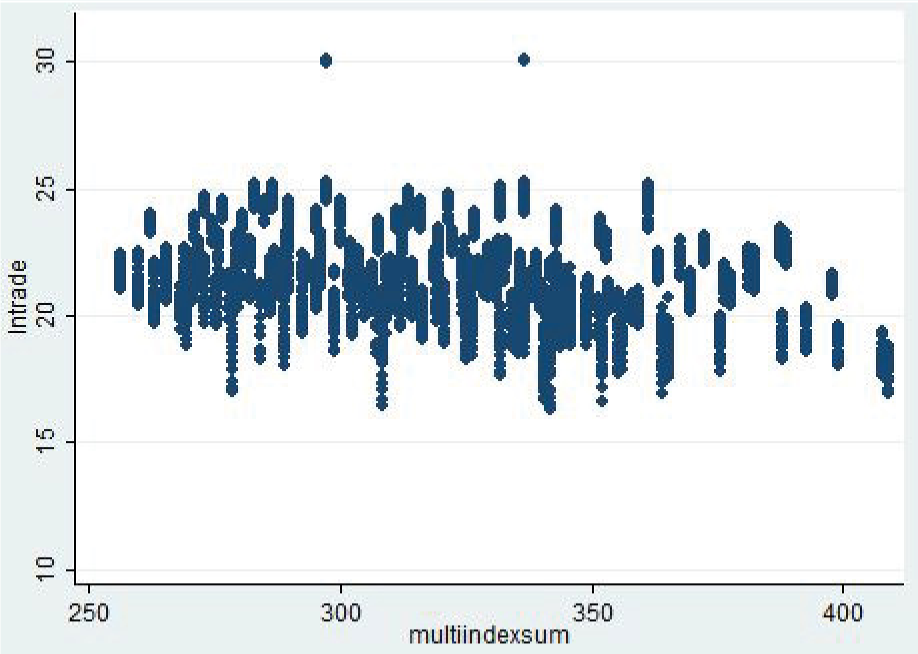
Plots of LDND, CCI, LDND/CCI and MI to ln trade in new dataset:







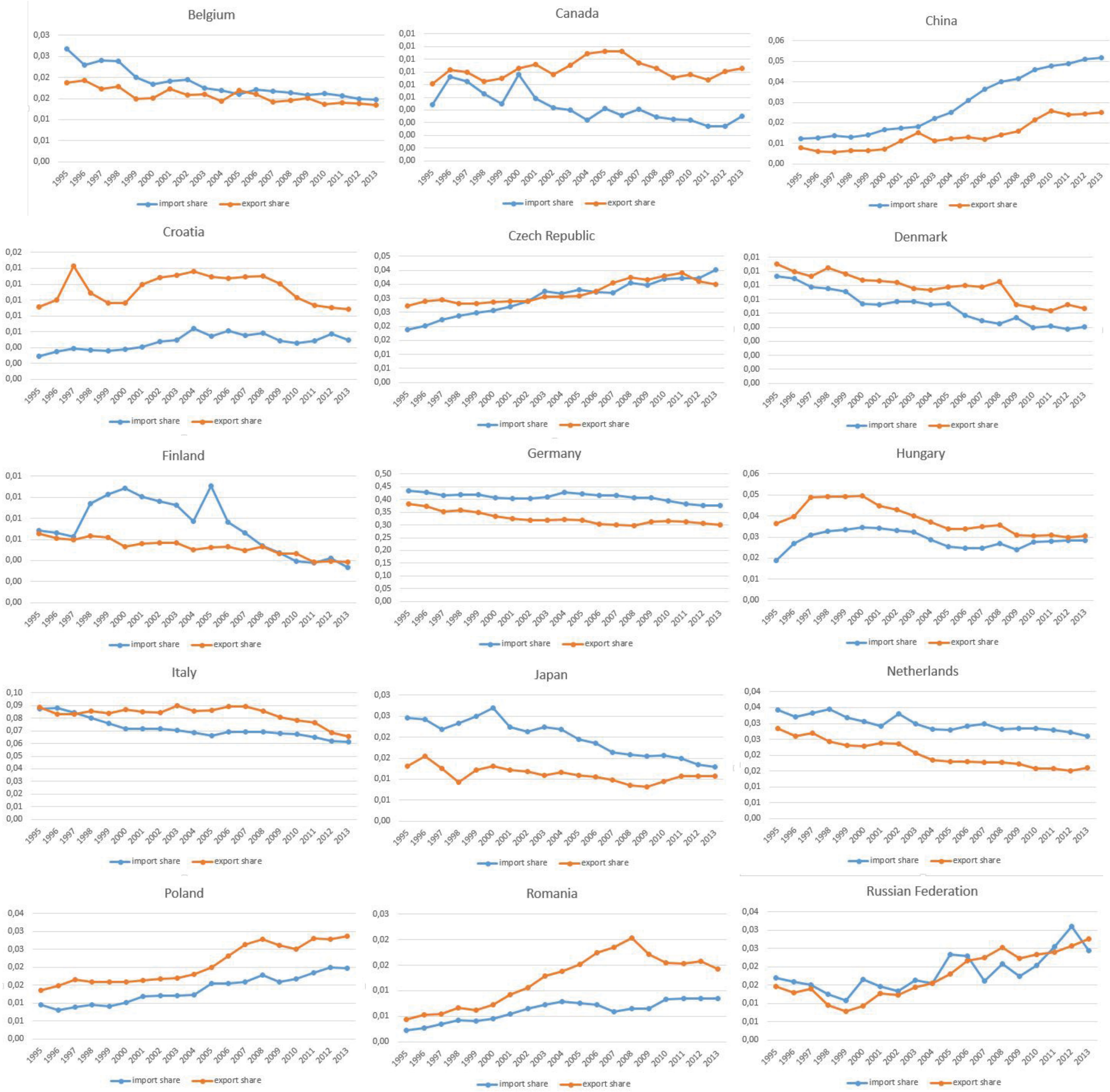
lntrade
ldndcci

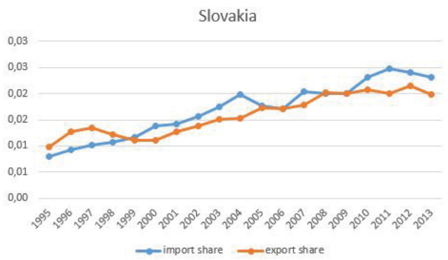


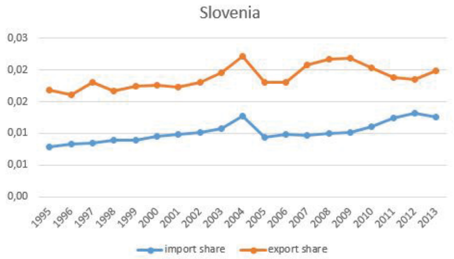
lntrade
multiindexsum

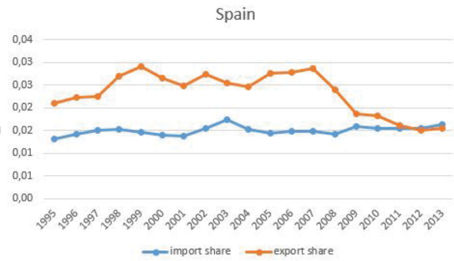
7.2 Import and Export Shares

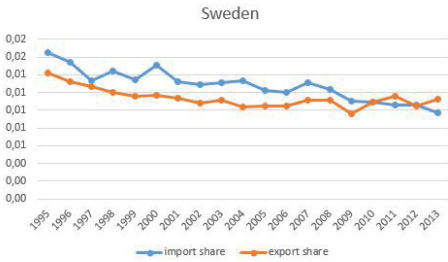
Import and export shares from 1995-2013 of different countries on Austria's trade

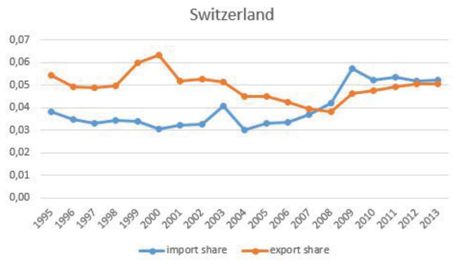


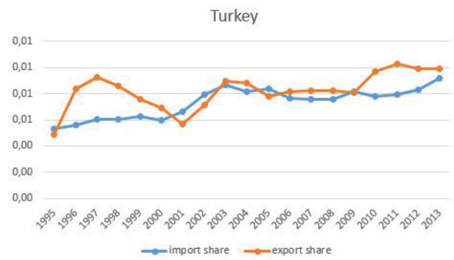

Slovakia
import share
export share

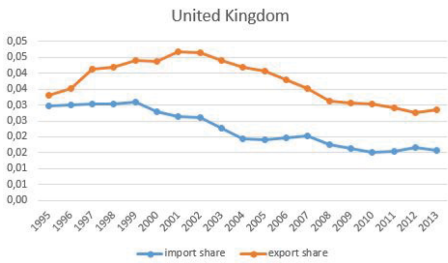

Slovenia
import share
export share

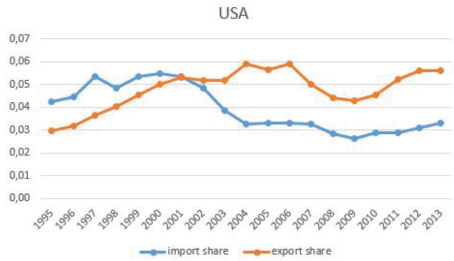

Spain
import share
export share

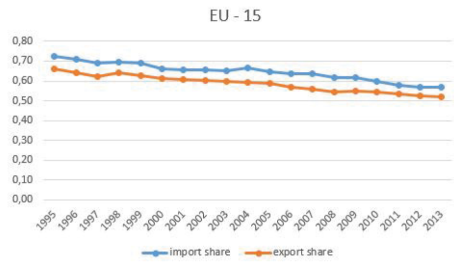

Sweden
import share
export share

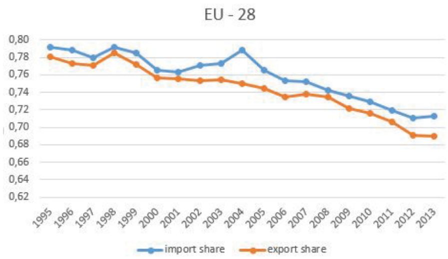

Switzerland
import share
export share

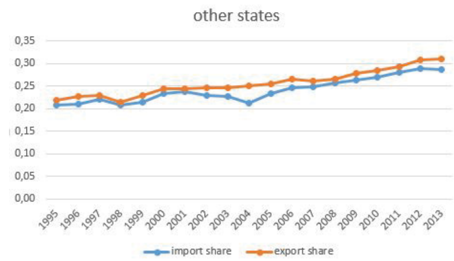

Turkey
import share
export share


United Kingdom
import share
export share


USA
import share
export share


EU - 15
import share
export share


EU - 28
import share
export share


other states
import share
export share

7.3 Regression results

```
Fixed-effects (within) regression               Number of obs      =     15602
Group variable: cty1                            Number of groups   =        24

R-sq:  within  = 0.8070                         Obs per group: min =        92
       between = 0.6578                                        avg =     650.1
       overall = 0.7014                                        max =       872

                                                F(4,15574)         =  16277.32
corr(u_i, Xb)  = -0.5818                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.715292   .0134514   127.52   0.000     1.688926    1.741658
      lngdp2 |   .9951518   .0062936   158.12   0.000     .9828156    1.007488
      lndist |  -1.119916   .0107327  -104.35   0.000    -1.140954   -1.098879
  lnlingdist |   -1.16592   .0442016   -26.38   0.000     -1.25256    -1.07928
       _cons |  -20.05643    .307335   -65.26   0.000    -20.65884   -19.45402
-------------+----------------------------------------------------------------
     sigma_u |  1.0770966
     sigma_e |  .86974989
         rho |  .60530937   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(23, 15574) =   264.96           Prob > F = 0.0000

. estimates store fixed

. xtreg lntrade lngdp1 lngdp2 lndist lnlingdist, re i(cty1)

Random-effects GLS regression                   Number of obs      =     15602
Group variable: cty1                            Number of groups   =        24

R-sq:  within  = 0.8069                         Obs per group: min =        92
       between = 0.6634                                        avg =     650.1
       overall = 0.7066                                        max =       872

                                                Wald chi2(4)       =  64720.48
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |    1.69108    .013254   127.59   0.000     1.665103    1.717058
      lngdp2 |   1.000405   .0062865   159.13   0.000     .9880832    1.012726
      lndist |  -1.128066   .0107186  -105.24   0.000    -1.149074   -1.107058
  lnlingdist |  -1.159519   .0442994   -26.17   0.000    -1.246344   -1.072693
       _cons |  -19.47059   .3183107   -61.17   0.000    -20.09447   -18.84671
-------------+----------------------------------------------------------------
     sigma_u |  .47124414
     sigma_e |  .86974989
         rho |  .22694214   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.5: LDND in Rose dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      9472
Group variable: cty1                            Number of groups   =        25

R-sq:  within  = 0.6271                         Obs per group: min =       352
       between = 0.4490                                        avg =     378.9
       overall = 0.5764                                        max =       384

                                                F(4,9443)          =   3969.32
corr(u_i, Xb)  = 0.0740                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |    .209841   .0101412    20.69   0.000      .189962      .22972
      lngdp2 |   .4551903   .0038172   119.25   0.000     .4477077    .4626729
      lndist |  -.7687439   .0147645   -52.07   0.000    -.7976855   -.7398022
      lnldnd |  -.8963226   .0548641   -16.34   0.000    -1.003868   -.7887771
       _cons |   10.69953   .3871187    27.64   0.000     9.940692    11.45836
-------------+----------------------------------------------------------------
     sigma_u |  .77377657
     sigma_e |   1.018028
         rho |  .36617095   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(24, 9443) =   108.49            Prob > F = 0.0000

. estimates store fixed

. xtreg lntrade lngdp1 lngdp2 lndist lnldnd, re i(cty1)

Random-effects GLS regression                   Number of obs      =      9472
Group variable: cty1                            Number of groups   =        25

R-sq:  within  = 0.6270                         Obs per group: min =       352
       between = 0.4963                                        avg =     378.9
       overall = 0.5876                                        max =       384

                                                Wald chi2(4)       =  15878.68
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .2224326   .0098125    22.67   0.000     .2032004    .2416648
      lngdp2 |   .4540214   .0038136   119.05   0.000     .4465469    .4614959
      lndist |  -.7634958   .0147211   -51.86   0.000    -.7923487    -.734643
      lnldnd |  -.8976552   .0548935   -16.35   0.000    -1.005244    -.790066
       _cons |   10.32574    .395182    26.13   0.000     9.551197    11.10028
-------------+----------------------------------------------------------------
     sigma_u |   .5523031
     sigma_e |   1.018028
         rho |  .22739991   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.6: LDND in new dataset pooled without other country specific variables

Fixed-effects (within) regression   Number of obs = 15602
Group variable: cty1   Number of groups = 24

R-sq: within = 0.8075   Obs per group: min = 92
between = 0.6568   avg = 650.1
overall = 0.7010   max = 872

F(8,15570) = 8161.71
corr(u_i, Xb) = -0.5817   Prob > F = 0.0000

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	1.716099	.0136718	125.52	0.000	1.689301	1.742897
lngdp2	.9948357	.0067854	146.61	0.000	.9815355	1.008136
lndist	-1.120426	.012236	-91.57	0.000	-1.14441	-1.096442
lnlingdist	-1.110319	.0454554	-24.43	0.000	-1.199417	-1.021222
border	-.0050172	.0311309	-0.16	0.872	-.0660374	.0560031
colony	.1120546	.0626878	1.79	0.074	-.0108208	.2349299
comcol	1.008992	.1743969	5.79	0.000	.6671533	1.35083
comctry	0	(omitted)				
landl	-.0308826	.0228364	-1.35	0.176	-.0756445	.0138793
_cons	-20.30657	.3138062	-64.71	0.000	-20.92167	-19.69147
sigma_u	1.0761165					
sigma_e	.86877918					
rho	.60540795	(fraction of variance due to u_i)				

F test that all u_i=0: F(23, 15570) = 262.16   Prob > F = 0.0000

Figure 7.7: LDND in Rose dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      9472
Group variable: cty1                            Number of groups   =        25

R-sq:  within  = 0.6515                         Obs per group: min =       352
       between = 0.5243                                        avg =     378.9
       overall = 0.6042                                        max =       384

                                                F(7,9440)          =   2521.43
corr(u_i, Xb)  = 0.1162                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .2100842   .0098111    21.41   0.000     .1908523    .2293161
      lngdp2 |   .4544846   .0038382   118.41   0.000     .4469609    .4620082
      lndist |  -.5983578   .0158852   -37.67   0.000    -.6294963   -.5672193
      lnldnd |  -.5728366   .0562873   -10.18   0.000    -.6831719   -.4625013
      border |    .651675   .0414722    15.71   0.000     .5703806    .7329694
      colony |  -.0181806   .0630145    -0.29   0.773    -.1417027    .1053414
      comcol |          0  (omitted)
     comctry |   .8754235   .0617539    14.18   0.000     .7543727    .9964744
       _cons |   7.865808    .395427    19.89   0.000     7.090686     8.64093
-------------+----------------------------------------------------------------
     sigma_u |  .75150375
     sigma_e |   .9842136
         rho |  .36829611   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(24, 9440) =   104.98             Prob > F = 0.0000
```

Figure 7.8: LDND in new dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =     15602
Group variable: cty1                            Number of groups   =        24

R-sq:  within  = 0.8259                         Obs per group: min =        92
       between = 0.7474                                        avg =     650.1
       overall = 0.8160                                        max =       872

                                                F(60,15518)        =   1227.15
corr(u_i, Xb)  = -0.0006                        Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.8978792	.0382933	23.45	0.000	.8228198	.9729385
lngdp2	.9747726	.0065549	148.71	0.000	.9619242	.987621
lndist	-1.098129	.0117123	-93.76	0.000	-1.121087	-1.075172
lnlingdist	-1.126423	.0433596	-25.98	0.000	-1.211413	-1.041433
border	.0257858	.0297101	0.87	0.385	-.0324494	.0840209
colony	.1127715	.0597094	1.89	0.059	-.004266	.2298089
comcol	.9778448	.1661348	5.89	0.000	.6522011	1.303488
comctry	0	(omitted)				
landl	-.0633266	.0218256	-2.90	0.004	-.1061074	-.0205458

Figure 7.9: LDND in Rose dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      9472
Group variable: cty1                            Number of groups   =        25

R-sq:  within  = 0.6618                         Obs per group: min =       352
       between = 0.0738                                        avg =     378.9
       overall = 0.5040                                        max =       384

                                                F(22,9425)         =    838.40
corr(u_i, Xb)  = -0.0350                        Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.106038	.0121045	8.76	0.000	.0823105	.1297655
lngdp2	.4434737	.0038616	114.84	0.000	.435904	.4510434
lndist	-.5804646	.0157114	-36.95	0.000	-.6112623	-.549667
lnldnd	-.5773511	.0554952	-10.40	0.000	-.6861336	-.4685685
border	.674462	.040919	16.48	0.000	.594252	.754672
colony	-.0042238	.0621344	-0.07	0.946	-.1260206	.1175731
comcol	0	(omitted)				
comctry	.8355118	.0609478	13.71	0.000	.716041	.9549826

Figure 7.10: LDND in new dataset pooled with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      7752
Group variable: pair                            Number of groups   =       272

R-sq:  within  = 0.8873                         Obs per group: min =         4
       between = 0.7259                                        avg =      28.5
       overall = 0.7805                                        max =        53

                                                Wald chi2(8)       =  58257.89
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	1.35913	.0227654	59.70	0.000	1.31451	1.403749
lngdp2	1.49666	.0190326	78.64	0.000	1.459357	1.533964
lndist	-1.590425	.0609101	-26.11	0.000	-1.709807	-1.471043
lnlingdist	-1.207622	.2797653	-4.32	0.000	-1.755952	-.659292
border	-.671267	.2032196	-3.30	0.001	-1.06957	-.272964
colony	.5981883	.4398163	1.36	0.174	-.2638358	1.460212
comcol	1.40724	.6830116	2.06	0.039	.0685621	2.745919
comctry	0	(omitted)				
landl	.1970421	.106266	1.85	0.064	-.0112355	.4053197
_cons	-19.08167	1.231139	-15.50	0.000	-21.49466	-16.66868
sigma_u	.82935193					
sigma_e	.53164336					
rho	.70875466	(fraction of variance due to u_i)				

Figure 7.11: LDND in Rose dataset country pair random effects regression with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      9472
Group variable: pairid                          Number of groups   =       592

R-sq:  within  = 0.4757                         Obs per group: min =        16
       between = 0.6965                                        avg =      16.0
       overall = 0.6606                                        max =        16

                                                Wald chi2(7)       =   9239.26
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.2731196	.0050427	54.16	0.000	.263236	.2830032
lngdp2	.2987349	.0050411	59.26	0.000	.2888546	.3086153
lndist	-.2947083	.0458318	-6.43	0.000	-.3845369	-.2048797
lnldnd	-.7492156	.2080947	-3.60	0.000	-1.157074	-.3413575
border	1.164091	.1551407	7.50	0.000	.8600204	1.468161
colony	.2246903	.2431471	0.92	0.355	-.2518693	.7012498
comcol	0	(omitted)				
comctry	.0493117	.220172	0.22	0.823	-.3822174	.4808408
_cons	9.169832	.9387766	9.77	0.000	7.329863	11.0098
sigma_u	.95577592					
sigma_e	.51169842					
rho	.77722648	(fraction of variance due to u_i)				

Figure 7.12: LDND in new dataset country pair random effects regression with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      7752
Group variable: pair                            Number of groups   =       272

R-sq:  within  = 0.9110                         Obs per group: min =         4
       between = 0.7352                                        avg =      28.5
       overall = 0.8117                                        max =        53

                                                Wald chi2(60)      =  75370.49
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.8733144	.0271032	32.22	0.000	.8201931	.9264357
lngdp2	1.058241	.0229686	46.07	0.000	1.013223	1.103259
lndist	-1.026776	.0524717	-19.57	0.000	-1.129618	-.923933
lnlingdist	-1.226312	.2197841	-5.58	0.000	-1.657081	-.7955431
border	.1622151	.1626714	1.00	0.319	-.156615	.4810453
colony	.302358	.3456554	0.87	0.382	-.3751142	.9798302
comcol	.3338117	.5393848	0.62	0.536	-.7233632	1.390987
comctry	0	(omitted)				

Figure 7.13: LDND in Rose dataset country pair random effects regression with other country specific variables and time dummies included

a

```
Random-effects GLS regression                   Number of obs      =      9472
Group variable: pairid                          Number of groups   =       592

R-sq:  within  = 0.6490                         Obs per group: min =        16
       between = 0.5090                                        avg =      16.0
       overall = 0.4969                                        max =        16

                                                Wald chi2(22)      =  16179.33
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.1159797	.0051152	22.67	0.000	.1059541	.1260053
lngdp2	.1387074	.0051134	27.13	0.000	.1286853	.1487295
lndist	-.019029	.0472699	-0.40	0.687	-.1116763	.0736183
lnldnd	-.826992	.213996	-3.86	0.000	-1.246417	-.4075676
border	1.642036	.1596619	10.28	0.000	1.329104	1.954967
colony	.5235977	.2500707	2.09	0.036	.0334681	1.013727
comcol	0	(omitted)				
comctry	-1.005871	.2268371	-4.43	0.000	-1.450464	-.5612786
t97	-1.042847	.0274738	-37.96	0.000	-1.096695	-.9889998
t98	-1.003174	.0273657	-36.66	0.000	-1.05681	-.949538
t99	-.9921392	.0274043	-36.20	0.000	-1.045851	-.9384276
t00	-.86198	.0274598	-31.39	0.000	-.9158002	-.8081598
t01	-.8211785	.0274028	-29.97	0.000	-.874887	-.76747
t02	-.7336014	.0271223	-27.05	0.000	-.7867602	-.6804426
t03	-.5506178	.0266121	-20.69	0.000	-.6027765	-.498459
t04	-.4647828	.0255872	-18.16	0.000	-.5149327	-.4146329
t05	-.3477619	.0254754	-13.65	0.000	-.3976927	-.297831
t06	-.255806	.0252299	-10.14	0.000	-.3052557	-.2063563
t07	-.0946795	.0251627	-3.76	0.000	-.1439975	-.0453614
t08	.0171114	.025151	0.68	0.496	-.0321836	.0664064
t09	-.2615582	.0251602	-10.40	0.000	-.3108714	-.2122451
t10	-.1070037	.0251539	-4.25	0.000	-.1563046	-.0577029
t11	.040029	.0251524	1.59	0.112	-.0092689	.0893269
t12	0	(omitted)				
_cons	17.60228	.972252	18.10	0.000	15.6967	19.50786
sigma_u	.95864287					
sigma_e	.41715539					
rho	.84079024	(fraction of variance due to u_i)				

Figure 7.14: LDND in new dataset country pair random effects regression with other country specific variables and time dummies included

```
Fixed-effects (within) regression               Number of obs      =      3330
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.8514                         Obs per group: min =        16
       between = 0.7751                                        avg =     222.0
       overall = 0.7393                                        max =       551

                                                F(4,3311)          =   4742.89
corr(u_i, Xb)  = -0.7363                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.892051   .0251467    75.24   0.000     1.842746    1.941355
      lngdp2 |   .8137204   .0110006    73.97   0.000     .7921517    .8352891
      lndist |  -1.080239   .0269865   -40.03   0.000    -1.133151   -1.027327
       lncci |   .7997728   .0225202    35.51   0.000     .7556179    .8439276
       _cons |  -27.73415   .5329374   -52.04   0.000    -28.77907   -26.68923
-------------+----------------------------------------------------------------
     sigma_u |  1.1375238
     sigma_e |  .68301743
         rho |  .73500713   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(14, 3311) =   145.09           Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lncci, re i(cty1)

Random-effects GLS regression                   Number of obs      =      3330
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.8512                         Obs per group: min =        16
       between = 0.7805                                        avg =     222.0
       overall = 0.7468                                        max =       551

                                                Wald chi2(4)       =  18407.81
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.828339   .0247212    73.96   0.000     1.779886    1.876791
      lngdp2 |   .8229152   .0111316    73.93   0.000     .8010977    .8447327
      lndist |  -1.081625   .0273208   -39.59   0.000    -1.135173   -1.028077
       lncci |   .7856042    .022681    34.64   0.000     .7411503    .8300581
       _cons |  -26.40292   .5357758   -49.28   0.000    -27.45302   -25.35282
-------------+----------------------------------------------------------------
     sigma_u |  .37378417
     sigma_e |  .68301743
         rho |  .23046566   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.15: CCI in Rose dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      2144
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.7544                         Obs per group: min =        16
       between = 0.7145                                        avg =     142.9
       overall = 0.6696                                        max =       256

                                                F(4,2125)          =   1632.12
corr(u_i, Xb)  = 0.1222                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1795894   .0120506    14.90   0.000     .1559571    .2032217
      lngdp2 |   .3618465   .0052188    69.34   0.000     .3516121     .372081
      lndist |  -1.497157   .0361524   -41.41   0.000    -1.568055   -1.426259
       lncci |   .0158535   .0274948     0.58   0.564    -.038066     .069773
       _cons |    15.3546   .4539926    33.82   0.000     14.46428    16.24492
-------------+----------------------------------------------------------------
     sigma_u |  .76462087
     sigma_e |  .71698461
         rho |  .53211858   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(14, 2125) =    98.24           Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lncci, re i(cty1)

Random-effects GLS regression                   Number of obs      =      2144
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.7543                         Obs per group: min =        16
       between = 0.7335                                        avg =     142.9
       overall = 0.6779                                        max =       256

                                                Wald chi2(4)       =   6530.83
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1913963   .0117323    16.31   0.000     .1684015    .2143911
      lngdp2 |   .3609036   .0052282    69.03   0.000     .3506566    .3711506
      lndist |  -1.494107   .0361629   -41.32   0.000    -1.564985   -1.423229
       lncci |    .019286   .0274008     0.70   0.482    -.0344185    .0729906
       _cons |   14.99746   .4616051    32.49   0.000     14.09274    15.90219
-------------+----------------------------------------------------------------
     sigma_u |  .50955233
     sigma_e |  .71698461
         rho |  .33558233   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.16: CCI in new dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      3330
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.8543                         Obs per group: min =        16
       between = 0.7753                                        avg =     222.0
       overall = 0.7430                                        max =       551

                                                F(8,3307)          =   2423.82
corr(u_i, Xb)  = -0.7334                        Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	1.894213	.0252621	74.98	0.000	1.844682	1.943744
lngdp2	.8119143	.0117112	69.33	0.000	.7889523	.8348762
lndist	-.9024315	.036177	-24.94	0.000	-.973363	-.8315
lncci	.7980811	.0246662	32.36	0.000	.7497185	.8464438
border	.3003742	.0452368	6.64	0.000	.2116793	.3890691
colony	.0773396	.1014034	0.76	0.446	-.1214801	.2761593
comcol	.8365981	.2580156	3.24	0.001	.3307117	1.342485
comctry	0	(omitted)				
landl	.0511179	.0485143	1.05	0.292	-.0440033	.146239
_cons	-28.95878	.5592096	-51.79	0.000	-30.05521	-27.86235
sigma_u	1.1278115					
sigma_e	.67674435					
rho	.73526106	(fraction of variance due to u_i)				

```
F test that all u_i=0:     F(14, 3307) =   143.85           Prob > F = 0.0000
```

Figure 7.17: CCI in Rose dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      2144
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.7796                         Obs per group: min =        16
       between = 0.7251                                        avg =     142.9
       overall = 0.6903                                        max =       256

                                                F(7,2122)          =   1072.36
corr(u_i, Xb)  = 0.1286                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1823679   .0114354    15.95   0.000     .1599421    .2047938
      lngdp2 |   .3522673   .0052475    67.13   0.000     .3419766     .362558
      lndist |   -1.07839   .0483047   -22.32   0.000    -1.173119   -.9836603
       lncci |   .0311458   .0262583     1.19   0.236    -.020349     .0826406
      border |   .8147061   .0557578    14.61   0.000     .7053605    .9240516
      colony |  -.6018616   .0914672    -6.58   0.000    -.7812364   -.4224868
      comcol |          0  (omitted)
     comctry |   .2796529    .080806     3.46   0.001     .1211855    .4381202
       _cons |   12.48995   .4993503    25.01   0.000     11.51068    13.46922
-------------+----------------------------------------------------------------
     sigma_u |  .75749655
     sigma_e |   .6797126
         rho |  .55396346   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(14, 2122) =   106.08            Prob > F = 0.0000
```

Figure 7.18: CCI in new dataset pooled with other country specific variables included

Fixed-effects (within) regression		Number of obs =	3330
Group variable: cty1		Number of groups =	15
R-sq: within = 0.8735		Obs per group: min =	16
between = 0.8115		avg =	222.0
overall = 0.8652		max =	551
		F(60,3255) =	374.52
corr(u_i, Xb) = 0.1097		Prob > F =	0.0000

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.6559964	.0980318	6.69	0.000	.4637861	.8482067
lngdp2	.8027823	.0111389	72.07	0.000	.7809423	.8246223
lndist	-.8713716	.0340816	-25.57	0.000	-.9381951	-.804548
lncci	.795235	.0232217	34.25	0.000	.7497044	.8407656
border	.3209564	.0425166	7.55	0.000	.2375944	.4043185
colony	.0973775	.0952777	1.02	0.307	-.0894329	.2841878
comcol	.8754752	.2424859	3.61	0.000	.4000348	1.350916
comctry	0	(omitted)				
landl	.0085669	.0457255	0.19	0.851	-.0810867	.0982206

Figure 7.19: CCI in Rose dataset pooled with other country specific variables included

Fixed-effects (within) regression		Number of obs =	2144
Group variable: cty1		Number of groups =	15
R-sq: within = 0.7939		Obs per group: min =	16
between = 0.5195		avg =	142.9
overall = 0.6319		max =	256
		F(22,2107) =	368.96
corr(u_i, Xb) = 0.0645		Prob > F =	0.0000

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.0990836	.0139888	7.08	0.000	.0716503	.126517
lngdp2	.3415533	.0052029	65.65	0.000	.3313499	.3517567
lndist	-1.051917	.0469507	-22.40	0.000	-1.143992	-.9598426
lncci	.0469404	.0255306	1.84	0.066	-.0031273	.0970081
border	.8386343	.0541618	15.48	0.000	.7324181	.9448505
colony	-.5984535	.0887637	-6.74	0.000	-.7725273	-.4243798
comcol	0	(omitted)				
comctry	.2342313	.0785475	2.98	0.003	.0801925	.38827

Figure 7.20: CCI in new dataset pooled with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      2217
Group variable: pair                            Number of groups   =        81

R-sq:  within  = 0.8933                         Obs per group: min =         4
       between = 0.7496                                        avg =      27.4
       overall = 0.7986                                        max =        51

                                                Wald chi2(8)       =  17356.75
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	1.801678	.04246	42.43	0.000	1.718458	1.884898
lngdp2	1.003621	.0395828	25.35	0.000	.9260399	1.081202
lndist	-1.217844	.2177742	-5.59	0.000	-1.644674	-.7910149
lncci	.6562349	.1193009	5.50	0.000	.4224095	.8900603
border	-.2426805	.2753249	-0.88	0.378	-.7823075	.2969465
colony	.3729973	.5778829	0.65	0.519	-.7596323	1.505627
comcol	1.177769	.795888	1.48	0.139	-.3821423	2.737681
comctry	0	(omitted)				
landl	.5789346	.1543011	3.75	0.000	.2765099	.8813593
_cons	-28.45541	1.595369	-17.84	0.000	-31.58227	-25.32854
sigma_u	.69934843					
sigma_e	.45095206					
rho	.70631967	(fraction of variance due to u_i)				

Figure 7.21: CCI in Rose dataset country pair random effects regression with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      2144
Group variable: pairid                          Number of groups   =       134

R-sq:  within  = 0.5093                         Obs per group: min =        16
       between = 0.7621                                        avg =      16.0
       overall = 0.7262                                        max =        16

                                                Wald chi2(7)       =   2518.14
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.2242182	.0077125	29.07	0.000	.2091018	.2393345
lngdp2	.2426876	.0084751	28.64	0.000	.2260766	.2592985
lndist	-.9956068	.1765336	-5.64	0.000	-1.341606	-.6496073
lncci	.0772116	.0825496	0.94	0.350	-.0845827	.2390059
border	1.158945	.2292311	5.06	0.000	.7096607	1.60823
colony	-.5696209	.383742	-1.48	0.138	-1.321741	.1824996
comcol	0	(omitted)				
comctry	-.7831184	.3196105	-2.45	0.014	-1.409543	-.1566934
_cons	13.79056	1.342314	10.27	0.000	11.15967	16.42144
sigma_u	.72017608					
sigma_e	.45070745					
rho	.71856498	(fraction of variance due to u_i)				

Figure 7.22: CCI in new dataset country pair random effects regression with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      2217
Group variable: pair                            Number of groups   =        81

R-sq:  within  = 0.9309                         Obs per group: min =         4
       between = 0.7505                                        avg =      27.4
       overall = 0.8334                                        max =        51

                                                Wald chi2(58)      =         .
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =         .
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.7299055	.0527862	13.83	0.000	.6264465	.8333644
lngdp2	.3854441	.0411747	9.36	0.000	.304743	.4661451
lndist	-.7456114	.166891	-4.47	0.000	-1.072712	-.418511
lncci	.612091	.0910152	6.73	0.000	.4337044	.7904775
border	.7871508	.2136963	3.68	0.000	.3683137	1.205988
colony	.5281466	.4413976	1.20	0.231	-.3369768	1.39327
comcol	.7012346	.6079102	1.15	0.249	-.4902475	1.892717
comctry	0	(omitted)				
landl	-.3864791	.1233572	-3.13	0.002	-.6282548	-.1447034

Figure 7.23: CCI in Rose dataset country pair random effects regression with other country specific variables and time dummies included

```
Random-effects GLS regression                   Number of obs      =      2144
Group variable: pairid                          Number of groups   =       134

R-sq:  within  = 0.6889                         Obs per group: min =        16
       between = 0.5803                                        avg =      16.0
       overall = 0.5867                                        max =        16

                                                Wald chi2(22)      =   4407.58
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.1024271	.0077488	13.22	0.000	.0872397	.1176144
lngdp2	.1150002	.0083954	13.70	0.000	.0985454	.131455
lndist	-.7821491	.1820709	-4.30	0.000	-1.139001	-.4252968
lncci	.28574	.0853335	3.35	0.001	.1184893	.4529907
border	1.495878	.2365882	6.32	0.000	1.032173	1.959582
colony	-.5374989	.3959214	-1.36	0.175	-1.313491	.2384928
comcol	0	(omitted)				
comctry	-1.493891	.3301973	-4.52	0.000	-2.141066	-.8467166
t97	-.9860632	.0498536	-19.78	0.000	-1.083774	-.888352
t98	-.8717663	.0496193	-17.57	0.000	-.9690182	-.7745143
t99	-.867551	.0496531	-17.47	0.000	-.9648694	-.7702327
t00	-.7836348	.0499215	-15.70	0.000	-.8814791	-.6857906
t01	-.7381006	.0497742	-14.83	0.000	-.8356562	-.640545
t02	-.661076	.0493041	-13.41	0.000	-.7577102	-.5644418
t03	-.4925466	.0484591	-10.16	0.000	-.5875247	-.3975686
t04	-.4157236	.0464867	-8.94	0.000	-.5068358	-.3246114
t05	-.3350892	.0463543	-7.23	0.000	-.4259419	-.2442365
t06	-.2582181	.0459387	-5.62	0.000	-.3482563	-.1681799
t07	-.0530098	.0458892	-1.16	0.248	-.142951	.0369313
t08	.0770477	.0459048	1.68	0.093	-.012924	.1670195
t09	-.2658359	.0458897	-5.79	0.000	-.3557781	-.1758936
t10	-.1187259	.0458893	-2.59	0.010	-.2086673	-.0287846
t11	.0594361	.0458954	1.30	0.195	-.0305172	.1493894
t12	0	(omitted)				
_cons	19.56589	1.392969	14.05	0.000	16.83572	22.29606
sigma_u	.72339539					
sigma_e	.35885977					
rho	.80250876	(fraction of variance due to u_i)				

Figure 7.24: CCI in new dataset country pair random effects regression with other country specific variables and time dummies included

```
Fixed-effects (within) regression               Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8539                         Obs per group: min =        60
       between = 0.8688                                        avg =     416.3
       overall = 0.8159                                        max =       551

                                                F(4,6640)          =   9702.76
corr(u_i, Xb)  = -0.6125                        Prob > F           =    0.0000

------------------------------------------------------------------------------
        lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
         lngdp1 |   1.630098   .0181664    89.73   0.000     1.594486     1.66571
         lngdp2 |   1.092509   .0086492   126.31   0.000     1.075554    1.109464
         lndist |   -1.28256   .0182646   -70.22   0.000    -1.318364   -1.246755
lnmultiindexsum |   5.238773   .1170399    44.76   0.000     5.009337    5.468209
          _cons |  -54.14014   .7946082   -68.13   0.000    -55.69783   -52.58245
----------------+-------------------------------------------------------------
        sigma_u |  .66564596
        sigma_e |  .68616493
            rho |  .48482462   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 6640) =   110.65           Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lnmultiindexsum, re i(cty1)

Random-effects GLS regression                   Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8538                         Obs per group: min =        60
       between = 0.8729                                        avg =     416.3
       overall = 0.8204                                        max =       551

                                                Wald chi2(4)       =  38606.26
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
         lngdp1 |   1.599273   .0176955    90.38   0.000      1.56459    1.633955
         lngdp2 |   1.098877   .0086233   127.43   0.000     1.081976    1.115779
         lndist |  -1.284446   .0182978   -70.20   0.000    -1.320309   -1.248583
lnmultiindexsum |   5.307416   .1164971    45.56   0.000     5.079086    5.535746
          _cons |  -53.93469   .7982565   -67.57   0.000    -55.49924   -52.37013
----------------+-------------------------------------------------------------
        sigma_u |  .31733651
        sigma_e |  .68616493
            rho |  .17619965   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.25: MI sum in Rose dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7477                         Obs per group: min =       256
       between = 0.7555                                        avg =     256.0
       overall = 0.7170                                        max =       256

                                                F(4,4076)          =   3020.15
corr(u_i, Xb)  = 0.2298                         Prob > F           =    0.0000

----------------------------------------------------------------------------------
         lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+----------------------------------------------------------------
          lngdp1 |   .1849305   .0090131    20.52   0.000     .1672598    .2026012
          lngdp2 |   .3483599   .0039318    88.60   0.000     .3406514    .3560684
          lndist |  -1.402925   .0228944   -61.28   0.000    -1.447811    -1.35804
lnmultiindexsum  |  -2.250432   .1404176   -16.03   0.000    -2.525727   -1.975137
           _cons |   27.93658   .8832638    31.63   0.000      26.2049    29.66826
-----------------+----------------------------------------------------------------
         sigma_u |  .63371014
         sigma_e |  .71014869
             rho |  .44330354   (fraction of variance due to u_i)
----------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 4076) =    96.72           Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lnmultiindexsum, re i(cty1)

Random-effects GLS regression                   Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7477                         Obs per group: min =       256
       between = 0.7725                                        avg =     256.0
       overall = 0.7247                                        max =       256

                                                Wald chi2(4)       =  12090.81
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

----------------------------------------------------------------------------------
         lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+----------------------------------------------------------------
          lngdp1 |   .1938472   .0087958    22.04   0.000     .1766078    .2110866
          lngdp2 |   .3475495   .0039349    88.32   0.000     .3398372    .3552618
          lndist |  -1.400444   .0229108   -61.13   0.000    -1.445348   -1.355539
lnmultiindexsum  |   -2.27409   .1399689   -16.25   0.000    -2.548424   -1.999756
           _cons |   27.81234   .8869491    31.36   0.000     26.07395    29.55073
-----------------+----------------------------------------------------------------
         sigma_u |  .42875142
         sigma_e |  .71014869
             rho |  .26713759   (fraction of variance due to u_i)
----------------------------------------------------------------------------------
```

Figure 7.26: MI sum in new dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8622                         Obs per group: min =        60
       between = 0.8976                                        avg =     416.3
       overall = 0.8415                                        max =       551

                                                F(8,6636)          =   5191.62
corr(u_i, Xb)  = -0.5801                        Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	1.588838	.0185049	85.86	0.000	1.552562	1.625114
lngdp2	1.153635	.0108264	106.56	0.000	1.132412	1.174859
lndist	-1.08376	.023593	-45.94	0.000	-1.13001	-1.03751
lnmultiindexsum	5.990248	.1376129	43.53	0.000	5.720482	6.260013
border	.2543949	.0307951	8.26	0.000	.1940266	.3147632
colony	.266204	.0694766	3.83	0.000	.1300075	.4024006
comcol	2.292884	.1744896	13.14	0.000	1.950829	2.63494
comctry	0	(omitted)				
landl	.2600387	.0287675	9.04	0.000	.2036453	.3164322
_cons	-60.28907	.930876	-64.77	0.000	-62.11389	-58.46426
sigma_u	.54225441					
sigma_e	.66652528					
rho	.39826835	(fraction of variance due to u_i)				

Figure 7.27: MI sum in Rose dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7562                         Obs per group: min =       256
       between = 0.7852                                        avg =     256.0
       overall = 0.7295                                        max =       256

                                                F(7,4073)          =   1804.67
corr(u_i, Xb)  = 0.2461                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1859432   .0088714    20.96   0.000     .1685505    .2033359
      lngdp2 |   .3451857   .0040344    85.56   0.000     .3372762    .3530953
      lndist |   -1.16575   .0334992   -34.80   0.000    -1.231427   -1.100073
lnmultiindexsum |  -2.0838   .138817   -15.01   0.000    -2.355957   -1.811643
      border |   .4723085   .0414969    11.38   0.000     .3909518    .5536652
      colony |  -.3215998   .0650422    -4.94   0.000     -.449118   -.1940815
      comcol |          0  (omitted)
     comctry |   .1407533   .0577425     2.44   0.015     .0275464    .2539602
       _cons |   25.33529   .9048354    28.00   0.000     23.56131    27.10926
-------------+----------------------------------------------------------------
     sigma_u |  .61709532
     sigma_e |  .69838019
         rho |  .43844378   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 4073) =    90.09            Prob > F = 0.0000
```

Figure 7.28: MI sum in new dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8799                         Obs per group: min =        60
       between = 0.2259                                        avg =     416.3
       overall = 0.7654                                        max =       551

                                                F(60,6584)         =    803.75
corr(u_i, Xb)  = -0.0192                        Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.5050143	.0578614	8.73	0.000	.3915871	.6184414
lngdp2	1.122985	.0102939	109.09	0.000	1.102805	1.143164
lndist	-1.085701	.0221651	-48.98	0.000	-1.129152	-1.04225
lnmultiindexsum	5.641114	.1301095	43.36	0.000	5.386057	5.896171
border	.2582503	.0288726	8.94	0.000	.2016507	.3148499
colony	.2825447	.0651437	4.34	0.000	.154842	.4102475
comcol	2.307329	.1635929	14.10	0.000	1.986633	2.628024
comctry	0	(omitted)				
landl	.1851043	.0271673	6.81	0.000	.1318476	.238361

Figure 7.29: MI sum in Rose dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7697                         Obs per group: min =       256
       between = 0.5292                                        avg =     256.0
       overall = 0.6537                                        max =       256

                                                F(22,4058)         =    616.40
corr(u_i, Xb)  = 0.1351                         Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.1060985	.0107709	9.85	0.000	.0849816	.1272154
lngdp2	.3336924	.0040355	82.69	0.000	.3257805	.3416042
lndist	-1.156757	.0326276	-35.45	0.000	-1.220725	-1.092789
lnmultiindexsum	-2.162488	.1353199	-15.98	0.000	-2.427789	-1.897187
border	.4886435	.0404287	12.09	0.000	.4093811	.5679059
colony	-.31707	.0633354	-5.01	0.000	-.4412422	-.1928978
comcol	0	(omitted)				
comctry	.0984605	.056329	1.75	0.081	-.0119753	.2088963

Figure 7.30: MI sum in new dataset pooled with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      3279
Group variable: pair                            Number of groups   =       119

R-sq:  within  = 0.8916                         Obs per group: min =         4
       between = 0.8483                                        avg =      27.6
       overall = 0.8740                                        max =        53

                                                Wald chi2(8)       =  26281.60
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	1.665842	.0312754	53.26	0.000	1.604543	1.72714
lngdp2	1.17548	.0280332	41.93	0.000	1.120536	1.230424
lndist	-1.150497	.1346902	-8.54	0.000	-1.414485	-.8865094
lnmultiindexsum	8.511414	.5161404	16.49	0.000	7.499797	9.52303
border	.3489303	.187298	1.86	0.062	-.0181669	.7160276
colony	.3652453	.4374579	0.83	0.404	-.4921565	1.222647
comcol	1.921269	.6105404	3.15	0.002	.7246323	3.117907
comctry	0	(omitted)				
landl	.6776097	.1036236	6.54	0.000	.4745112	.8807082
_cons	-76.66711	3.224547	-23.78	0.000	-82.98711	-70.34712
sigma_u	.55318673					
sigma_e	.44809282					
rho	.60381592	(fraction of variance due to u_i)				

Figure 7.31: MI sum in Rose dataset country pair random effects regression with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      4096
Group variable: pairid                          Number of groups   =       256

R-sq:  within  = 0.5713                         Obs per group: min =        16
       between = 0.7985                                        avg =      16.0
       overall = 0.7678                                        max =        16

                                                Wald chi2(7)       =   6129.45
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.2297502	.0051268	44.81	0.000	.2197019	.2397985
lngdp2	.2393475	.0052754	45.37	0.000	.2290079	.249687
lndist	-1.049722	.1197584	-8.77	0.000	-1.284444	-.8149999
lnmultiindexsum	-3.089091	.3870315	-7.98	0.000	-3.847659	-2.330523
border	.8516954	.1543921	5.52	0.000	.5490925	1.154298
colony	-.3880662	.2511284	-1.55	0.122	-.8802688	.1041364
comcol	0	(omitted)				
comctry	-.664511	.2047643	-3.25	0.001	-1.065842	-.2631803
_cons	32.22499	2.445934	13.17	0.000	27.43105	37.01893
sigma_u	.66782866					
sigma_e	.40588544					
rho	.73025592	(fraction of variance due to u_i)				

Figure 7.32: MI sum in new dataset country pair random effects regression with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      3279
Group variable: pair                            Number of groups   =       119

R-sq:  within  = 0.9308                         Obs per group: min =         4
       between = 0.7271                                        avg =      27.6
       overall = 0.8104                                        max =        53

                                                Wald chi2(60)      =  39910.20
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.8407085	.0413094	20.35	0.000	.7597435	.9216735
lngdp2	.4850112	.0362291	13.39	0.000	.4140034	.556019
lndist	-.8815326	.1202164	-7.33	0.000	-1.117152	-.6459128
lnmultiindexsum	.8592522	.5361074	1.60	0.109	-.1914991	1.910003
border	.8044314	.1674735	4.80	0.000	.4761893	1.132673
colony	1.16843	.3898926	3.00	0.003	.4042543	1.932605
comcol	1.691097	.5423802	3.12	0.002	.6280515	2.754143
comctry	0	(omitted)				
landl	-.6182669	.1035315	-5.97	0.000	-.8211849	-.415349

Figure 7.33: MI sum in Rose dataset country pair random effects regression with other country specific variables and time dummies included

Random-effects GLS regression		Number of obs =	4096
Group variable: pairid		Number of groups =	256
R-sq: within = 0.7508		Obs per group: min =	16
between = 0.6307		avg =	16.0
overall = 0.6382		max =	16
		Wald chi2(22) =	11424.88
corr(u_i, X) = 0 (assumed)		Prob > chi2 =	0.0000

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.1105367	.0049117	22.50	0.000	.1009101	.1201634
lngdp2	.1174966	.0050472	23.28	0.000	.1076044	.1273889
lndist	-.9697294	.1237994	-7.83	0.000	-1.212372	-.727087
lnmultiindexsum	-3.842923	.4002504	-9.60	0.000	-4.6274	-3.058447
border	1.104319	.1596479	6.92	0.000	.7914146	1.417223
colony	-.3193366	.2596155	-1.23	0.219	-.8281737	.1895005
comcol	0	(omitted)				
comctry	-1.422334	.2119544	-6.71	0.000	-1.837757	-1.006911
t97	-.9346528	.031028	-30.12	0.000	-.9954665	-.8738391
t98	-.8256641	.0308836	-26.73	0.000	-.8861948	-.7651334
t99	-.8326884	.030906	-26.94	0.000	-.8932631	-.7721137
t00	-.7430762	.0310715	-23.92	0.000	-.8039752	-.6821772
t01	-.6936396	.0309832	-22.39	0.000	-.7543654	-.6329137
t02	-.6136886	.0306932	-19.99	0.000	-.6738461	-.5535311
t03	-.4393404	.0301717	-14.56	0.000	-.4984759	-.3802049
t04	-.3858479	.0289132	-13.35	0.000	-.4425167	-.329179
t05	-.3003431	.0288344	-10.42	0.000	-.3568574	-.2438287
t06	-.2142039	.0285917	-7.49	0.000	-.2702425	-.1581652
t07	-.0219479	.0285605	-0.77	0.442	-.0779255	.0340298
t08	.0966619	.0285697	3.38	0.001	.0406664	.1526574
t09	-.2136576	.0285611	-7.48	0.000	-.2696362	-.1576789
t10	-.0876473	.0285608	-3.07	0.002	-.1436254	-.0316692
t11	.071513	.0285643	2.50	0.012	.015528	.1274981
t12	0	(omitted)				
_cons	43.69709	2.534311	17.24	0.000	38.72993	48.66425
sigma_u	.67106102					
sigma_e	.30902867					
rho	.82503676	(fraction of variance due to u_i)				

Figure 7.34: MI sum in new dataset country pair random effects regression with other country specific variables and time dummies included

```
Fixed-effects (within) regression               Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8545                         Obs per group: min =        60
       between = 0.7498                                        avg =     416.3
       overall = 0.7578                                        max =       551

                                                F(4,6640)          =   9747.83
corr(u_i, Xb)  = -0.6446                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.631518    .018126    90.01   0.000     1.595985    1.667051
      lngdp2 |   1.090815   .0086019   126.81   0.000     1.073952    1.107677
      lndist |  -1.261085   .0181989   -69.29   0.000    -1.296761   -1.225409
lnmultiindex1|          0  (omitted)
lnmultiindex2|   2.659994   .0589252    45.14   0.000     2.544482    2.775507
       _cons |  -37.54495   .4910297   -76.46   0.000    -38.50752   -36.58237
-------------+----------------------------------------------------------------
     sigma_u |  .95261382
     sigma_e |  .68480819
         rho |  .65929221   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 6640) =   269.24            Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lnmultiindex1 lnmultiindex2, re i(cty1)

Random-effects GLS regression                   Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8544                         Obs per group: min =        60
       between = 0.8841                                        avg =     416.3
       overall = 0.8283                                        max =       551

                                                Wald chi2(5)       =  38796.20
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.602975    .017728    90.42   0.000     1.568229    1.637721
      lngdp2 |   1.095472    .008608   127.26   0.000     1.078601    1.112344
      lndist |  -1.262473   .0182307   -69.25   0.000    -1.298204   -1.226741
lnmultiindex1|   4.284882   .4509602     9.50   0.000     3.401016    5.168747
lnmultiindex2|   2.672345   .0591158    45.21   0.000      2.55648     2.78821
       _cons |  -58.73683   2.376008   -24.72   0.000    -63.39372   -54.07994
-------------+----------------------------------------------------------------
     sigma_u |  .28627775
     sigma_e |  .68480819
         rho |  .14876092   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.35: MI single in Rose dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7470                         Obs per group: min =       256
       between = 0.6938                                        avg =     256.0
       overall = 0.6863                                        max =       256

                                                F(4,4076)          =   3008.16
corr(u_i, Xb)  = 0.2011                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1846518   .0090262    20.46   0.000     .1669555    .2023482
      lngdp2 |   .3492334   .0039303    88.86   0.000     .3415278    .3569389
      lndist |  -1.405946   .0229273   -61.32   0.000    -1.450896   -1.360996
lnmultiindex1 |          0  (omitted)
lnmultiindex2 |  -1.110649   .0711029   -15.62   0.000    -1.250049   -.9712481
       _cons |   20.57918   .4898378    42.01   0.000     19.61884    21.53953
-------------+----------------------------------------------------------------
     sigma_u |  .70585519
     sigma_e |  .71120503
         rho |  .49622475   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 4076) =   129.65            Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lnmultiindex1 lnmultiindex2, re i(cty1)

Random-effects GLS regression                   Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7469                         Obs per group: min =       256
       between = 0.7297                                        avg =     256.0
       overall = 0.7268                                        max =       256

                                                Wald chi2(5)       =  12045.99
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1930853   .0088175    21.90   0.000     .1758033    .2103672
      lngdp2 |   .3485655   .0039347    88.59   0.000     .3408537    .3562773
      lndist |  -1.403443   .0229422   -61.17   0.000    -1.448409   -1.358477
lnmultiindex1 |  -2.008158   .6662561    -3.01   0.003    -3.313996   -.7023197
lnmultiindex2 |   -1.11288   .0712381   -15.62   0.000    -1.252504   -.9732559
       _cons |    30.5318   3.430539     8.90   0.000     23.80806    37.25553
-------------+----------------------------------------------------------------
     sigma_u |  .42928162
     sigma_e |  .71120503
         rho |  .26703951   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.36: MI single in new dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8630                         Obs per group: min =        60
       between = 0.7394                                        avg =     416.3
       overall = 0.7725                                        max =       551

                                                F(8,6636)          =   5224.23
corr(u_i, Xb)  = -0.6159                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.589525   .0184407    86.20   0.000     1.553375    1.625675
      lngdp2 |    1.15294   .0107437   107.31   0.000     1.131879    1.174001
      lndist |  -1.055493   .0235688   -44.78   0.000    -1.101696   -1.009291
lnmultiindex1 |         0  (omitted)
lnmultiindex2 |  3.049111    .069208    44.06   0.000     2.913441     3.18478
      border |   .2608635   .0307129     8.49   0.000     .2006564    .3210707
      colony |   .2529742   .0693147     3.65   0.000      .117095    .3888534
      comcol |   2.307857   .1740195    13.26   0.000     1.966723    2.648992
     comctry |          0  (omitted)
       landl |   .2653978    .028674     9.26   0.000     .2091875    .3216081
       _cons |  -41.38443   .5544627   -74.64   0.000    -42.47136    -40.2975
-------------+----------------------------------------------------------------
     sigma_u |  .90240988
     sigma_e |  .66472773
         rho |  .64825613   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 6636) =   287.82          Prob > F = 0.0000
```

Figure 7.37: MI single in Rose dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7554                         Obs per group: min =       256
       between = 0.7370                                        avg =     256.0
       overall = 0.7021                                        max =       256

                                                F(7,4073)          =   1796.48
corr(u_i, Xb)  = 0.2199                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1857114   .0088864    20.90   0.000     .1682891    .2031336
      lngdp2 |   .3459124   .0040366    85.69   0.000     .3379984    .3538264
      lndist |  -1.170708   .0335821   -34.86   0.000    -1.236547   -1.104869
lnmultiindex1|          0  (omitted)
lnmultiindex2|  -1.021306   .0703793   -14.51   0.000    -1.159288   -.8833242
      border |   .4721145   .0415813    11.35   0.000     .3905925    .5536365
      colony |  -.3160852   .0651787    -4.85   0.000    -.4438712   -.1882993
      comcol |          0  (omitted)
     comctry |   .1310673   .0578729     2.26   0.024     .0176048    .2445297
       _cons |   18.50372   .5286359    35.00   0.000      17.4673    19.54014
-------------+----------------------------------------------------------------
     sigma_u |  .68370685
     sigma_e |  .69958169
         rho |  .48852536   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 4073) =   119.72            Prob > F = 0.0000
```

Figure 7.38: MI single in new dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      6660
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.8806                         Obs per group: min =        60
       between = 0.3257                                        avg =     416.3
       overall = 0.7855                                        max =       551

                                                F(60,6584)         =    809.50
corr(u_i, Xb)  = -0.0265                        Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.5044857	.0576799	8.75	0.000	.3914144	.6175571
lngdp2	1.122741	.0102098	109.97	0.000	1.102726	1.142755
lndist	-1.058996	.0221321	-47.85	0.000	-1.102382	-1.01561
lnmultiindex1	0	(omitted)				
lnmultiindex2	2.875248	.0653959	43.97	0.000	2.74705	3.003445
border	.2643366	.0287828	9.18	0.000	.207913	.3207602
colony	.2696749	.0649636	4.15	0.000	.1423252	.3970246
comcol	2.321515	.1630808	14.24	0.000	2.001823	2.641206
comctry	0	(omitted)				
landl	.1909804	.0270643	7.06	0.000	.1379255	.2440353

Figure 7.39: MI single in Rose dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      4096
Group variable: cty1                            Number of groups   =        16

R-sq:  within  = 0.7688                         Obs per group: min =       256
       between = 0.4458                                        avg =     256.0
       overall = 0.6199                                        max =       256

                                                F(22,4058)         =    613.28
corr(u_i, Xb)  = 0.1084                         Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.1061482	.010792	9.84	0.000	.08499	.1273064
lngdp2	.3344871	.0040385	82.82	0.000	.3265695	.3424048
lndist	-1.161937	.0327157	-35.52	0.000	-1.226077	-1.097796
lnmultiindex1	0	(omitted)				
lnmultiindex2	-1.059812	.0686181	-15.45	0.000	-1.194341	-.9252832
border	.4883786	.0405203	12.05	0.000	.4089366	.5678207
colony	-.3113603	.0634834	-4.90	0.000	-.4358227	-.1868979
comcol	0	(omitted)				
comctry	.08856	.0564711	1.57	0.117	-.0221543	.1992744

Figure 7.40: MI single in new dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      2779
Group variable: cty1                            Number of groups   =        14

R-sq:  within  = 0.8316                         Obs per group: min =        16
       between = 0.7781                                        avg =     198.5
       overall = 0.7404                                        max =       486

                                                F(4,2761)          =   3408.57
corr(u_i, Xb)  = -0.7433                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.789704   .0293197    61.04   0.000     1.732213    1.847195
      lngdp2 |   .8318292   .0127824    65.08   0.000     .8067651    .8568933
      lndist |   -1.20759   .0315005   -38.34   0.000    -1.269357   -1.145823
   lnldndcci |    .847173   .0278813    30.39   0.000     .7925028    .9018433
       _cons |  -29.26126   .6609137   -44.27   0.000    -30.55719   -27.96532
-------------+----------------------------------------------------------------
     sigma_u |  1.0709135
     sigma_e |  .71499428
         rho |  .69168005   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(13, 2761) =   101.20            Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lnldndcci, re i(cty1)

Random-effects GLS regression                   Number of obs      =      2779
Group variable: cty1                            Number of groups   =        14

R-sq:  within  = 0.8313                         Obs per group: min =        16
       between = 0.7854                                        avg =     198.5
       overall = 0.7496                                        max =       486

                                                Wald chi2(4)       =  13277.73
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.715227   .0285124    60.16   0.000     1.659343     1.77111
      lngdp2 |   .8422006   .0129121    65.23   0.000     .8168933    .8675078
      lndist |  -1.207175   .0318343   -37.92   0.000    -1.269569   -1.144781
   lnldndcci |   .8311088   .0280168    29.66   0.000      .776197    .8860207
       _cons |  -27.64208   .6541651   -42.26   0.000    -28.92422   -26.35994
-------------+----------------------------------------------------------------
     sigma_u |  .37666468
     sigma_e |  .71499428
         rho |  .21723746   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.41: LDND and CCI in Rose dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      2128
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.7571                         Obs per group: min =        16
       between = 0.7114                                        avg =     141.9
       overall = 0.6647                                        max =       240

                                                F(4,2109)          =   1643.75
corr(u_i, Xb)  = 0.1302                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1791262   .0118208    15.15   0.000     .1559444    .2023079
      lngdp2 |   .3624215   .0051081    70.95   0.000     .3524041    .3724389
      lndist |  -1.515268   .0338807   -44.72   0.000    -1.581711   -1.448825
   lnldndcci |  -.0811974   .0284216    -2.86   0.004    -.1369347   -.0254601
       _cons |   16.15361   .4899573    32.97   0.000     15.19276    17.11446
-------------+----------------------------------------------------------------
     sigma_u |  .78038933
     sigma_e |  .70320891
         rho |  .55188205   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(14, 2109) =   102.70            Prob > F = 0.0000

. xtreg lntrade lngdp1 lngdp2 lndist lnldndcci, re i(cty1)

Random-effects GLS regression                   Number of obs      =      2128
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.7570                         Obs per group: min =        16
       between = 0.7317                                        avg =     141.9
       overall = 0.6738                                        max =       240

                                                Wald chi2(4)       =   6570.55
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   .1911201   .0115195    16.59   0.000     .1685423    .2136979
      lngdp2 |   .3614777     .00512    70.60   0.000     .3514427    .3715127
      lndist |  -1.512278   .0339331   -44.57   0.000    -1.578785    -1.44577
   lnldndcci |  -.0769131   .0283468    -2.71   0.007    -.1324717   -.0213544
       _cons |   15.76962   .4959869    31.79   0.000     14.79751    16.74174
-------------+----------------------------------------------------------------
     sigma_u |  .50388919
     sigma_e |  .70320891
         rho |  .33925964   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 7.42: LDND and CCI in new dataset pooled without other country specific variables

```
Fixed-effects (within) regression               Number of obs      =      2779
Group variable: cty1                            Number of groups   =        14

R-sq:  within  = 0.8344                         Obs per group: min =        16
       between = 0.7738                                        avg =     198.5
       overall = 0.7383                                        max =       486

                                                F(8,2757)          =   1736.32
corr(u_i, Xb)  = -0.7468                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     lntrade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lngdp1 |   1.798404   .0294551    61.06   0.000     1.740647     1.85616
      lngdp2 |   .8262033   .0134302    61.52   0.000     .7998691    .8525375
      lndist |  -1.166708   .0408459   -28.56   0.000    -1.246799   -1.086616
   lnldndcci |   .8084317   .0296085    27.30   0.000     .7503747    .8664887
      border |   .0565444   .0528126     1.07   0.284    -.0470119    .1601007
      colony |  -.2020782   .3714843    -0.54   0.587    -.9304938    .5263373
      comcol |   1.717822   .2693328     6.38   0.000     1.189708    2.245937
     comctry |          0  (omitted)
       landl |  -.1461029   .0552981    -2.64   0.008    -.2545329   -.0376729
       _cons |  -29.25408   .6842008   -42.76   0.000    -30.59568   -27.91248
-------------+----------------------------------------------------------------
     sigma_u |  1.0982907
     sigma_e |  .70955634
         rho |  .70552349   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(13, 2757) =    93.35           Prob > F = 0.0000
```

Figure 7.43: LDND and CCI in Rose dataset pooled with other country specific variables included

Fixed-effects (within) regression   Number of obs = 2128
Group variable: cty1   Number of groups = 15

R-sq: within = 0.7780   Obs per group: min = 16
between = 0.7189   avg = 141.9
overall = 0.6812   max = 240

F(7,2106) = 1054.24
corr(u_i, Xb) = 0.1364   Prob > F = 0.0000

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.1812639	.0113203	16.01	0.000	.1590637	.2034641
lngdp2	.3550465	.0051564	68.86	0.000	.3449344	.3651587
lndist	-1.116579	.046707	-23.91	0.000	-1.208176	-1.024982
lnldndcci	-.0363299	.0274199	-1.32	0.185	-.090103	.0174431
border	.7214024	.0564464	12.78	0.000	.6107058	.832099
colony	-.5415878	.0906681	-5.97	0.000	-.7193962	-.3637795
comcol	0	(omitted)				
comctry	.342154	.0808086	4.23	0.000	.183681	.500627
_cons	13.09485	.540191	24.24	0.000	12.03549	14.15421
sigma_u	.77576349					
sigma_e	.67283727					
rho	.57069527	(fraction of variance due to u_i)				

F test that all u_i=0:   F(14, 2106) = 108.93   Prob > F = 0.0000

Figure 7.44: LDND and CCI in new dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      2779
Group variable: cty1                            Number of groups   =        14

R-sq:  within  = 0.8561                         Obs per group: min =        16
       between = 0.8046                                        avg =     198.5
       overall = 0.8579                                        max =       486

                                                F(58,2707)         =    277.65
corr(u_i, Xb)  = 0.1877                         Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.6103288	.1048049	5.82	0.000	.404823	.8158346
lngdp2	.8175992	.0128046	63.85	0.000	.7924915	.8427068
lndist	-1.129656	.0385597	-29.30	0.000	-1.205265	-1.054046
lnldndcci	.8079931	.0279269	28.93	0.000	.7532329	.8627534
border	.0728605	.049703	1.47	0.143	-.0245992	.1703202
colony	-.1697299	.3495047	-0.49	0.627	-.855053	.5155932
comcol	1.769714	.2535178	6.98	0.000	1.272606	2.266822
comctry	0	(omitted)				
landl	-.1962185	.0522193	-3.76	0.000	-.2986123	-.0938247

Figure 7.45: LDND and CCI in Rose dataset pooled with other country specific variables included

```
Fixed-effects (within) regression               Number of obs      =      2128
Group variable: cty1                            Number of groups   =        15

R-sq:  within  = 0.7924                         Obs per group: min =        16
       between = 0.5037                                        avg =     141.9
       overall = 0.6196                                        max =       240

                                                F(22,2091)         =    362.76
corr(u_i, Xb)  = 0.0649                         Prob > F           =    0.0000
```

lntrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lngdp1	.0995541	.0138612	7.18	0.000	.0723709	.1267373
lngdp2	.3446844	.0051117	67.43	0.000	.3346598	.354709
lndist	-1.093393	.0453879	-24.09	0.000	-1.182403	-1.004383
lnldndcci	-.021471	.0266523	-0.81	0.421	-.0737387	.0307968
border	.7455399	.0548335	13.60	0.000	.638006	.8530738
colony	-.5438585	.087991	-6.18	0.000	-.7164176	-.3712995
comcol	0	(omitted)				
comctry	.3004179	.0785349	3.83	0.000	.1464033	.4544326

Figure 7.46: LDND and CCI in new dataset pooled with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      1719
Group variable: pair                            Number of groups   =        67

R-sq:  within  = 0.8743                         Obs per group: min =         4
       between = 0.7671                                        avg =      25.7
       overall = 0.8012                                        max =        51

                                                Wald chi2(8)       =  11290.89
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	1.63089	.0485537	33.59	0.000	1.535727	1.726054
lngdp2	1.100279	.0448441	24.54	0.000	1.012386	1.188171
lndist	-1.426783	.2581153	-5.53	0.000	-1.932679	-.9208861
lnldndcci	.6713288	.1710338	3.93	0.000	.3361087	1.006549
border	-.3994814	.3317043	-1.20	0.228	-1.04961	.250647
colony	.8273576	.8639246	0.96	0.338	-.8659035	2.520619
comcol	1.560545	.9044744	1.73	0.084	-.2121927	3.333282
comctry	0	(omitted)				
landl	.6139792	.2503626	2.45	0.014	.1232775	1.104681
_cons	-28.58214	2.287482	-12.50	0.000	-33.06553	-24.09876
sigma_u	.73586275					
sigma_e	.46997702					
rho	.71027492	(fraction of variance due to u_i)				

Figure 7.47: LDND and CCI in Rose dataset country pair random effects regression with other country specific variables included

Random-effects GLS regression		Number of obs =	2128
Group variable: pairid		Number of groups =	133
R-sq:	within = 0.5093	Obs per group: min =	16
	between = 0.7608	avg =	16.0
	overall = 0.7233	max =	16
		Wald chi2(7) =	2489.52
corr(u_i, X)	= 0 (assumed)	Prob > chi2 =	0.0000

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.2245113	.007736	29.02	0.000	.209349	.2396737
lngdp2	.2430482	.0084871	28.64	0.000	.2264137	.2596827
lndist	-1.026369	.1725066	-5.95	0.000	-1.364476	-.6882622
lnldndcci	.0181598	.0862068	0.21	0.833	-.1508024	.1871221
border	1.120201	.231967	4.83	0.000	.6655537	1.574848
colony	-.5511977	.3830106	-1.44	0.150	-1.301885	.1994892
comcol	0	(omitted)				
comctry	-.7843617	.3188	-2.46	0.014	-1.409198	-.1595252
_cons	14.10198	1.476529	9.55	0.000	11.20803	16.99592
sigma_u	.71154541					
sigma_e	.4518466					
rho	.71263025	(fraction of variance due to u_i)				

Figure 7.48: LDND and CCI in new dataset country pair random effects regression with other country specific variables included

```
Random-effects GLS regression                   Number of obs      =      1719
Group variable: pair                            Number of groups   =        67

R-sq:  within  = 0.9191                         Obs per group: min =         4
       between = 0.7579                                        avg =      25.7
       overall = 0.8092                                        max =        51

                                                Wald chi2(58)      =  17665.66
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.7118834	.0553496	12.86	0.000	.6034002	.8203666
lngdp2	.4391953	.0487855	9.00	0.000	.3435775	.5348131
lndist	-1.102657	.1948925	-5.66	0.000	-1.484639	-.7206746
lnldndcci	.5304716	.1289422	4.11	0.000	.2777495	.7831937
border	.6219159	.2544349	2.44	0.015	.1232327	1.120599
colony	.1322775	.6548933	0.20	0.840	-1.15129	1.415845
comcol	1.356714	.6824236	1.99	0.047	.0191888	2.69424
comctry	0	(omitted)				
landl	-.5607074	.195553	-2.87	0.004	-.9439842	-.1774305

Figure 7.49: LDND and CCI in Rose dataset country pair random effects regression with other country specific variables and time dummies included

```
Random-effects GLS regression                   Number of obs      =      2128
Group variable: pairid                          Number of groups   =       133

R-sq:  within  = 0.6880                         Obs per group: min =        16
       between = 0.5717                                        avg =      16.0
       overall = 0.5772                                        max =        16

                                                Wald chi2(22)      =   4327.43
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lntrade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lngdp1	.1032638	.0077997	13.24	0.000	.0879766	.118551
lngdp2	.1161744	.0084383	13.77	0.000	.0996356	.1327131
lndist	-.8488533	.1786816	-4.75	0.000	-1.199063	-.4986438
lnldndcci	.2265756	.0894959	2.53	0.011	.0511669	.4019844
border	1.463779	.2404852	6.09	0.000	.9924364	1.935121
colony	-.6022954	.396909	-1.52	0.129	-1.380223	.1756319
comcol	0	(omitted)				
comctry	-1.465213	.3307798	-4.43	0.000	-2.113529	-.8168962

Figure 7.50: LDND and CCI in new dataset country pair random effects regression with other country specific variables and time dummies included

7.4 Zusammenfassung und Lebenslauf

Sprache und Handel sind zwei Aspekte menschlicher Interaktion, die in direkte Relation gesetzt werden können. Die kulturellen Überschneidungen von und durch Sprachen und die Fähigkeit einer Gesellschaft, in verschiedener Sprache kommunizieren zu können, erhöhen den bilateralen Handel zwischen Nationen und sind von hoher Wichtigkeit. In dieser Arbeit wird anhand von zwei verschiedenen Datensätzen gezeigt, dass noch immer ein negativer Einfluss von der Unterschiedlichkeit von Sprachen auf den Handel ausgeht. Weiter wird der positive Einfluss einer vielsprachigen Gesellschaft auf den Handel innerhalb der Europäischen Union gezeigt. Durch den Einsatz von Variablen, die Sprachdistanzen beschreiben, wird im Kontext eines klassischen Gravitationsmodells analysiert, wie Ähnlichkeiten zwischen Sprachen und der Einsatz für das Erlernen mehrerer Sprachen den bilateralen Handel erhöhen.

Jochen Löb

Lebenslauf

Persönliche Angaben

Geburt	**23.07.1987**, *Erbach im Odenwald*.
Nationalität	**Deutsch**.

Ausbildung und akademischer Werdegang

2000–2007	**Gymnasium**, *Gymnasium Michelstadt*, Michelstadt.
07/2007	**Abitur**.
2008–2009	**Studium der Volkswirtschaftslehre, Nebenfach Betriebswirtschaftslehre**, *Universität Mannheim*.
2009–2012	**Studium der Volkswirtschaftslehre**, *Universität Wien*.
06/2012	**Abschluss Bakk. soc. oec.**, *Universität Wien*.
seit 2012	**Magisterstudium der Volkswirtschaftslehre**, *Universität Wien*.
seit 2009	**Bachelorstudium der Mathematik**, *Universität Wien*.
2013-2015	**Organisation und Verwaltung**, *Innovationszentrum der Universität Wien GmbH*.

Wien, 20. Oktober 2015