The Trade-Creating Effect of Emigrant Networks: A Global Approach

by

Krisztina Horváth

Submitted to

Central European University

Department of Economics

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Professor Ádám Szeidl

Budapest, Hungary

2013

Abstract

In this thesis I investigate the effect of emigration on international export flows. The main contribution of this paper is that it puts the country-specific findings of the corresponding literature into a new comprehensive cross-country perspective using a world-wide dataset. A further development of this paper is that it ensures the consistency of its results by exploiting information about the extensive margin of trade as well: contrary to previous papers I pay special attention to the large proportion of observations with zero trade flows between countries, and I implement a two-step estimation procedure recently developed by Helpman, Melitz and Rubinstein (2008) that is able to simultaneously correct for the potential selection bias and the bias from asymmetric trade flows between pairs of countries. My findings confirm the previously documented positive effect of migration on trade flows for the whole world as well; however I show that it is subject to a critical level of development in both the origin and the destination countries. My results also reveal that the main source of this positive relationship is the trade facilitating role of the information networks maintained by emigrant diasporas.

Contents

1 Introduction

The process of globalization seriously accelerated in the second part of the last century, resulting in challenging economic and social consequences that have reshaped many aspects of daily life. Therefore it has become more important than ever for responsible policy makers to understand the underlying mechanisms and relationships between global phenomena. Recognizing this need, my thesis links two major consequences of global integration: the unprecedented level of mobility of people and goods between countries.

The constant increase of international migration is a striking global challenge of our days: just over the last decade the total stock of migrants increased by more than 40% (International Organization for Migration, 2013). As a result, in 2010 more than 215 million people (3 percent of the world's population) lived outside of their country of birth (The World Bank, 2011). These numbers are even expected to rise in the future due to demographic and environmental changes. Such enormous relocation of people clearly leads to serious and contradictory consequences both for the origin and the destination countries. Immigrants are often blamed for decreasing wage levels in their destination countries (Grossman, 1982; Borjas, 2003) and for social conflicts during the assimilation. On the other hand, relocation of the labor force towards more productive countries could contribute to the economic growth of the destination countries of migrants (Friedbergs and Hunt, 1995). In addition, emigration is associated with welfare transfers through the enormous amount of remittances sent from developed to developing countries (The World Bank, 2011); however the loss of skilled labor might have serious negative long-run consequences for low income origin countries, contributing to the maintenance of poverty-traps.

A tremendous rise in the volume of trade over the last century was another significant consequence of globalization that resulted in an average export-GDP ratio of 28% in 2010 (United Nations, 2013). This process has also led to serious consequences on the world economy: international trade is often thought to be welfare improving, but at the same time it is generally associated with a leading role in the increase of income inequality. The phenomenon of increasing international trade volumes is often explained by the significant reduction of formal trade barriers, such as the decline of transport costs (Jacks and Pendakur, 2010), the formation of trade agreements between countries and the decrease of protectionist trade policies (Anderson and van Wincoop, 2004). However, the decline of formal obstacles of trade just emphasizes even more the relative importance of informal trade barriers - such as information asymmetries and cultural differences - whose effect on trade is just as large as that of formal impediments, but they are much more difficult to be influenced by policy making. Therefore the investigation of informal trade barriers became an area of interest in the empirical literature of international economics.

In this thesis I investigate how these parallel processes in international migration and trade can be linked together. This paper aims to answer the question whether there is a relationship between emigrant stocks coming from a given origin, living in a certain destination country and export flows from the emigrants' origin country to their destination country.

Most studies investigating the migration-trade link have found a significant positive relationship between migrant stocks and trade flows (Gould, 1994; Girma and Yu, 2002; Wagner, Head and Ries, 2002; Bastos and Silva, 2012), but the there is still large room

for further research. The contribution of this thesis is twofold.

First, I explore the question from a global perspective using a new country-pair level dataset from the year 2000 that covers 224×223 country combinations of the world. Previous papers investigated the link between migration and international trade from the perspectives of certain developed countries. However, as both international migration and trade are global phenomena, examining the relationship between them from the perspectives of single countries might deliver only country-specific results. This concern is enlarged by the fact that these focus countries were developed ones, but no work has investigated yet this relationship from the perspectives of less developed countries, although they are hugely affected by emigration. For investigating this question I use the traditional gravity equation with country fixed-effects, augmented with emigrant stocks in the destination countries of export flows as a baseline specification. Focusing on the whole world, I find evidence for the previously documented positive relationship. According to my results, a 1% rise in the emigrant population in a given destination country increases exports from the origin country of emigrants to that country by 0.206%. However, the subsample estimates reveal that this link is limited only to developed countries, and there is no significant relationship between emigration and export flows if at least one member of the pair is a low income country. This result implies that the trade-facilitating role of emigrant networks is subject to a critical level of development in both the origin and the destination countries.

Second, I develop substantially the identification strategy to rule out the potential sources of estimation biases that exist in previous works. A serious shortcoming of the traditional gravity equation is that it fails to predict zero and asymmetric trade flows

between countries. Contrary to the usual approach in the literature that restricts the empirical analysis only to observations with positive trade flows for exploring the migrant-trade link, I keep the substantial part of the sample with zero exports between pairs of countries, because this is a general characteristic of trade datasets and it contains important information about the extensive margin (Helpman, Melitz and Rubinstein, 2008; Santos Silva and Tenreyro, 2006). The omission of this information causes a potential source of estimation biases arising due to the existence of unobserved country-pair level factors. Therefore, I use two-step estimation procedures that are able to solve this problem. I implement an estimation method developed by Helpman, Melitz and Rubinstein (2008) that uses the predicted components of a first stage selection equation as additional explanatory variables to yield consistent estimates of the gravity equation by correcting simultaneously for the potential selection bias and the bias coming from asymmetric trade flows between pairs of countries. Performing this method, the estimated coefficient of emigrants drops by $\frac{1}{3}$, from 0.206 to 0.140. According to my results, the main source of the bias of the traditional gravity estimates is the unobserved variation in the proportion of exporters between directions, not the self-selection of countries into trading partners.

I also pay special attention to exploring the underlying channels of the link between emigration and export flows: theoretically migrants might affect trade through the reduction of information asymmetries and by creating extra demand for goods produced in their country of origin. To separate these channels, I use the information content of interaction terms of emigrants with other explanatory variables that might influence the strength of the emigrant-export link. To help a deeper understanding of the information mechanism, based on White and Tadesse (2008) I introduce a new explanatory variable

that captures cultural differences between countries. I show that the relationship between emigrant networks and export flows increases with the extent of cultural dissimilarity between countries. Finally, to test the existence of the preference diffusion channel, I compare the coefficient estimates of migration to the same and opposite directions to export flows, but I find contradictory results. My results suggest that the main source of the trade facilitating role emigrants is the information channel.

Throughout this thesis I emphasize the need for caution for interpreting these results causally, because of the potential existence of reverse causality that is extremely difficult to rule out econometrically in this case.

The paper is structured as follows. Chapter 2 summarizes the literature about the channels that explain the link between emigration and export flows, and presents the most important findings of the previous empirical works. Chapter 3 describes the data and Chapter 4 provides some descriptive statistics. Chapter 5 details the estimation strategies and results. Chapter 6 concludes.

2 Literature Review

There are different theoretical explanations on how emigrants might influence export flows between countries. The two main mechanisms usually mentioned in the literature are the information channel and the preference diffusion channel (Combes, Lafourcade and Mayer, 2005; Head and Ries, 1998; Rauch and Trindade, 2002).

First, emigrant diasporas are associated with an information network that facilitates the international flow of country-specific knowledge: they possess unique information about social norms, culture specific business practices, informal procedures, language and market information connected to both of their country of birth and their new destination country. Since a large pool of recent work on heterogeneous firms trade models emphasized the role of fixed costs of exporting when entering new markets (Melitz, 2003), emigrants might affect trade by reducing information asymmetries associated with different business environments of different cultures, thus reducing the fixed costs of exporting and facilitate the formation of trade relationships between countries (Combes, Lafourcade and Mayer, 2005). Therefore, as Figure 1 shows, emigrants as human information carriers can facilitate trade in both directions between two countries by reducing the market failures originated from information asymmetries. The importance of information networks maintained by emigrants might be especially emphasized when the level of dissimilarity is large between the countries, since extra information about a very distant country is more precious (Girma and Yu, 2002). This role may also depend on the characteristics of the institutional environment of the countries: additional information about the parties is more important where the extent of market failure is greater and less restored by policy interactions. Finally, the information link associated with emigrants might also help

the matching between potential sellers and buyers. This additional information is especially important in case of differentiated products, where the role of information is more emphasized in trade (Rauch, 1999).

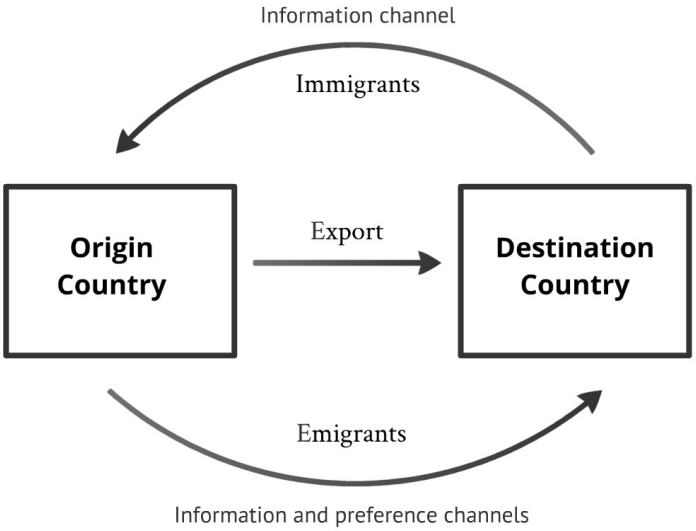

Figure 1: Mechanisms behind the emigration-export link

Second, emigrants also bring their origin country-specific preferences with them to their destination countries; thus they facilitate the world-wide diffusion of preferences (Combes, Lafourcade and Mayer, 2005). This fact should have a positive effect on the volume of exports on the same direction as migration flows (see Figure 1), since emigrants may create a substantial market in the destination country for the goods produced in their origin countries. Again, this link might be more emphasized when considering countries with a higher level of dissimilarity, because of the more limited availability of substitutes. However, it is also possible that after a critical number of immigrants from a certain origin country the local production also starts in the destination country: for example many traditional Russian food items are produced in Germany. This consideration works in the opposite direction and might decrease the role of the preference diffusion mechanism in international trade flows.

Since the influential works of Rauch (1999) and Gould (1994) in the 1990s, the effect of migrant networks on international trade has become an area of interest in the empirical literature; hence I summarize the main qualitative findings of the seminal works. This rich literature of the migration-trade link can be explained by the diversity of possible samples and the difficult identification of the true effect of migrants on trade flows.

Early studies mostly investigated the question using country level data and single-country perspectives. Gould (1994) wrote the first paper that documented a positive link between migration and trade. Using data on US country-level bilateral trade flows and information about the country of birth of US immigrants he found an early evidence for the preference diffusion channel. Rauch (1999) also provided a leading work corresponding to this topic by investigating the effect of cultural ties (common language and colonial ties) on international trade flows using a cross-country approach and a country-level dataset. He found that cultural networks have significantly higher effect on the trade of differentiated products than that of homogenous goods which are traded on organized exchanges, thus information asymmetries are lower. Rauch (1999) was thus the first who provided empirical evidence for the existence of the information channel mechanism, because the information cost of matching sellers and buyers makes information carried by cultural networks more valuable in case of differentiated products.

This evidence was confirmed by many later studies using migration data instead of cultural ties. A classic study by Rauch and Trindade (2002) investigated the indirect effect of migrant networks proxied by Chinese ethnics in each country on international trade flows.

They found that countries with higher concentration of residents of Chinese origin trade significantly more intensively with each other, especially in case of differentiated products. The role of dissimilarity between countries in the relationship of migrant networks and international trade was first emphasized by Girma and Yu (2002) who distinguished between immigrants coming from Commonwealth and Non-Commonwealth member countries when investigating the question using UK country-level bilateral trade data. They found that only immigrants from Non-Commonwealth countries have significant effect on trade flows between the UK and the origin country of the immigrants. Since the Commonwealth membership was used as an indicator for the level of general similarity to the UK, this result can be interpreted as an evidence supporting the existence of the information network mechanism. Thus only migration between sufficiently different countries carries enough information to facilitate international trade. Wagner, Head and Ries (2002) also emphasized the role of the information channel maintained by migrants, by using Canadian province level immigration and international trade data.

After the work of Wagner, Head and Ries (2002) the literature became dominated by studies that still investigate the question from the perspective of a single destination country of migrants, but using regional level international trade and migration data instead of using country level ones (Combes, Lafourcade and Mayer, 2005; Peri and Requena-Silvente, 2010; Dunlevy, 2006). Combes, Lafourcade and Mayer (2005) have examined the link between migration and trade flows at the sub-national level in France finding similar results to studies using international data.

3 Data

In order to investigate the relationship between emigration and export flows I use a world-wide country-pair level dataset that contains information on all possible combinations of countries. I created this database by merging information from several sources.

The empirical analysis of this topic is substantially constrained by the limited availability of migration data. Most of the available information on migration comes from developed countries who record their immigrant population. This explains why the empirical literature usually investigates the migration-trade link only from the perspectives of some single developed countries instead of doing cross-country studies. However, I was able to find a migration dataset that allows me to perform a cross-sectional analysis from a global point of view. Even though the cross-sectional nature of the data limits the dimensions of information I can analyze, the uniqueness of this new perspective is worth the trade-off.

The migration data used in this thesis came from the Global Migrant Origin Database of the University of Sussex's Development Research Centre on Migration, Globalisation and Poverty (Development Research Centre on Migration, Globalisation and Poverty, 2013), which is currently the most complete dataset of international migration. This database contains information of a 226-226 matrix of origin and destination country foreign born migrant stocks. The data were collected from the 2000 round of censuses in each destination countries, using international migration databases from different sources (United Nations, World Bank, OECD) and applying a forecasting technique to derive missing values from available historical data (Development Research Centre on Migration, Globalisation and Poverty, 2013).

Data on bilateral export flows between each country pairs was obtained from the United Nations Commodity Trade Statistics Database (United Nations, 2013). I combined the migration and export data with further information. Since I use the gravity equation for estimation, I merged these data with a pair-wise gravity database of Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) containing country-pair level information on common gravity controls, such distance, common border, common language, colonial ties etc (Head, Mayer and Ries, 2010). I also used GDP data obtained from the World Development Indicators (WDI) database of the World Bank (The World Bank, 2013*b*). I collected further country-specific information from the WDI on the institutional and regulatory environment, such as the average number of documents required for exporting/importing, the average number of procedures needed to enforce contracts, the average number of startup procedures for registering a firm and an index for the ability of controlling corruption. Finally I merged my dataset with information about measures of cross-cultural variation between countries. I obtained these data from a cultural map of the world in 2000 by Inglehart and Welzel (2005), who used the World Values Survey and the European Values Survey (Inglehart et al., 2004; Hagenaars, Halman and Moors, 2003) to score 80 countries on the two dimensions of traditional values/ secular-rational values and survival values/ self expression values.

The final dataset contains information on 224×223 origin-destination country-pairs from the year 2000.

4 Descriptive statistics

Investigating the export flows between country-pairs in my dataset reveals a common feature of trade data: a large proportion of the dataset (62%) consists of observations with zero trade flows between pairs of countries. Exploring further trade flows between country-pairs highlights another important characteristic of the data that is often left out of consideration in the literature that analyzes the migration-trade link: as Figure 2 shows, 26% of the country-pairs trades in both directions, while 23% trades only in a single direction. In other words, the exporting relationship between pairs of countries is often asymmetric. These facts contradict the simple gravity equation and reveal the need for new underlying models that are able to reproduce these regularities of the actual data (Helpman, Melitz and Rubinstein, 2008).

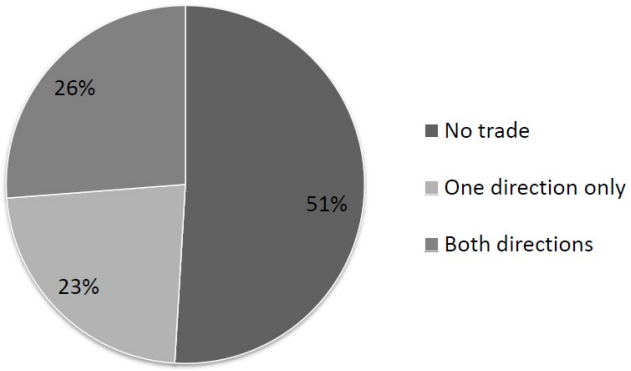

Figure 2: Distribution of the direction of trade between country pairs

Table 1 presents a correlation matrix of the main variables.The correlation patterns suggest that there is a positive relationship between exports and emigrants. Other correlation coefficients also match the expectations: emigrants are negatively correlated with the GDP per capita of their origin countries, but positively with the GDP per capita of their destination countries. This observation is in line with the fact that migration flows from developing countries to developed ones (International Organization for Migration,

2013). Distance is negatively correlated both with exports and emigrants that is also quite intuitive. Exports are positively correlated with both the GDP per capita of the origin and the destination countries, suggesting that richer countries tend to trade more.

	l export	l emigrants	l gdp o.	l gdp/cap o.	l gdp d.	l gdp/cap d.	l distance	l cult diff
l export	1							
l emigrants	0.5723	1						
l gdp o.	0.5557	0.2138	1					
l gdp/cap o.	0.3678	-0.0409	0.5388	1				
l gdp d.	0.4574	0.5052	-0.0658	-0.0618	1			
l gdp/cap d.	0.3137	0.3355	-0.0471	-0.0477	0.5663	1		
l distance	-0.2290	-0.1536	0.1614	-0.0821	0.1500	-0.0827	1	
l cult diff	-0.0987	-0.0963	0.0206	0.0640	0.0185	0.0477	0.1873	1

Table 1: Correlation matrix of the main variables

Figure 3 shows the cultural map of the world in 2000 (Inglehart and Welzel, 2005) that I used as an information source to construct a variable that measures cultural differences between the pairs of countries.

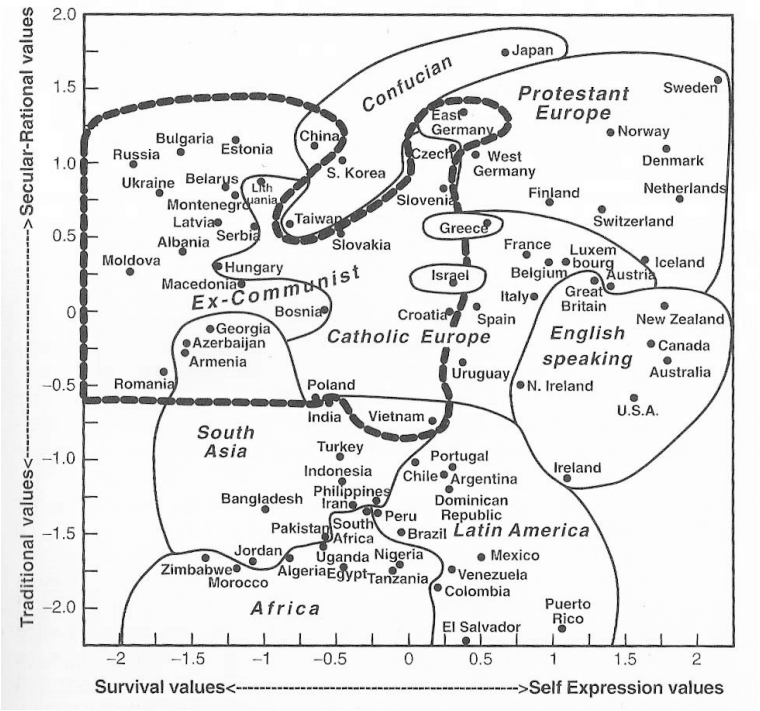

Figure 3: Cultural map of the world around 2000 (source: Inglehart and Welzel (2005))

This graph can be interpreted as follows. Low scores on the traditional values/ secular-rational (TSR) axis mean the dominance of traditional values, such as the importance of God, respect of the authority and national pride. High scores on this axis indicate the opposite. Considering the other dimension (SSE), low scores represent the dominance of survival values over self expression values: for example the priority of physical security over the quality of life, most of the citizens of these countries would never sign any petition, suspicion instead of trusting other people, etc. High scores on this axis indicate the opposite (Inglehart and Welzel, 2005). The map shows some well identifiable groups of countries, and it can also be observed that income level also strongly corresponds to the country's basic cultural values: countries with lower incomes are in the lower left-hand corner, while rich countries are located around the upper right-hand corner.

I translated the country level information of this cultural map into country-pair specific measures of cultural differences by using a formula developed by White and Tadesse (2008):

$$cultural\ distance_{ij} = \sqrt{(TSR_i - TSR_j)^2 + (SSE_i - SSE_j)^2}$$

I used this cultural difference variable as an additional explanatory variable in the gravity equation to get a bet better insight about the role of emigrant networks in international trade.

5 Estimation

5.1 Empirical strategy

In this part I analyze the effect of emigrant networks on export flows between countries using my global dataset and test hypotheses derived from theoretical considerations about the potential underlying mechanisms. This analysis helps to provide a deeper understanding of the role of social and business networks-effects in international trade.

5.1.1 Hypotheses

Hypothesis 1: *The tightness of the emigrant-export link differs across country income groups.*

It is reasonable to assume that the trade-creating role of emigrants is not the same when considering countries of different income levels, since both the characteristics of emigrant people and trade varies with the level of development. For example refugees from poor countries are less likely to utilize their culture-specific knowledge in the business sector of the destination country, thus lowering the effect of the information-network mechanism. Therefore I expect a weaker relationship between emigration and exports flows for low-income origin countries. Evidence for this hypothesis would provide a new insight about the effect migrant networks on international trade that, to the best of my knowledge, was not investigated yet by other authors.

Hypothesis 2: *The role of emigrant networks increases with the level of dissimilarity between countries.*

The existence of the information-network mechanism would imply that the role of extra information carried by emigrant networks increases with the extent of information

asymmetry across countries. For this reason I expect a tighter relationship between emigration and exports flows for example for countries with different cultural values.

Hypothesis 3: *The effect of emigrants from a given origin country and living in the destination country of trade flows should have larger effect on exports than immigrants living in the origin country of the export flows and coming from the destination country of exports.*

The underlying argument for this hypothesis is that the preference diffusion channel should exist only in case of migration and trade flowing to the same direction. Therefore I expect a higher coefficient estimate for emigrants than for immigrants.

Hypothesis 4: *Emigrants have a different role in supporting the extensive margin and the intensive margin of trade.*

It is possible that the information network-effect is more likely to affect the fixed costs of exporting than the variables. Therefore I expect the role of emigrants to be grater when considering the probability of an exporting relationship between two countries.

5.1.2 Methodology

I use the gravity equation as an empirical framework to estimate the relationship between emigrant populations and export flows. This is a widely used method in the empirical literature to predict bilateral trade flows. This powerful empirical tool originated with Tinbergen (1962), and it relies on a stable observed relationship between the volume of trade, the size of economies and their distances:

$$X_{ij} = c \, \frac{GDP_i \cdot GDP_j}{D_{ij}}$$

The usual procedure to estimate the gravity equation is to use its log-linear form. Accordingly, I estimate the following gravity equation augmented with emigrant stocks born in the origin country of exports, living the destination country:

$$log\, export_{ij} = \alpha + \beta_1 \mathbf{X} + \beta_2 log\, emigrants_{ij} + \beta_3 log\, gdp_i + \beta_4 log\, gdp_j + \varepsilon_{ij},$$

where $\mathbf{X}$ is a vector of country and country-pair level control variables that are commonly used as proxies for trade impediments: such as distance, contiguity, historical ties or the existence of trade agreements between the members of the country-pair.

The identification of the true effect of emigrants on export flows is difficult, because of the endogeneity of emigrants in the trade flow equations. This problem arises for two reasons: one is reverse causality and the other is the existence of unobserved variables that affect both emigration and export flows between countries.

Reverse causality means that it is possible that the relationship between emigrant stocks and export flows is not caused by emigrants, but the other way around. In other words, an intense trade relationship may act as a driver of migration between countries, for example because of business delegates sent to the country of export destination. In this case this concern is extremely difficult to overcome by econometric methods, since usually the same factors affect both migration and trade at the same time. Thus, the exclusion restriction is violated that would be necessary to be able to rule out the possibility of reverse causality: I was not able to find such factors or events that create exogenous variation in migration but do not affect trade directly. Although there were some attempts in the literature to use historical migrant stocks as an instrumental variable (Bratti, M. and De Benedictis, L. and Santoni, G., 2013; Peri and Requena-Silvente, 2010), it is easy

to see that historical emigrant stocks also have a direct effect on current trade, not only through current migrant stocks. However, there is a general opinion in the literature that migration is much more strongly driven by long term factors, than by current trade between countries (Genc, M. and Gheasi, M. and Nijkamp, P. and Poot, J., 2011). This argument helps to rule out that causality goes from export flow to migration.

In this thesis I present the process of identifying the effect of emigrants on export flows that is not biased due to the unobserved heterogeneity that affects both. Therefore first I use a simple OLS estimation of the gravity equation augmented by emigrant stocks in the destination country of the export flows to replicate the findings of the most general approach used to study the link between migration and trade. Then, using origin and destination country fixed effects I implement the estimation method of the gravity equation proposed by Anderson and van Wincoop (2003) to correct for unobserved country-level heterogeneity that may bias the OLS estimates. A disappearing effect of emigrants on export flows after the inclusion of country fixed-effects would imply that the effects yielded by the OLS estimation are driven by omitted unobserved country-specific variables, such as the general attractiveness of countries. However, a remaining relationship, even after controlling for country fixed-effects, still does not necessarily mean the consistency of the estimates, since the country-pair level observations imply that there might exist even country-pair level unobserved variables that biases further the estimates.

A further problem is that, as I mentioned in Chapter 4, 62% of the dataset consists of observations with zero export flows that is a common characteristic of even the country-level trade data. Since the traditional gravity equation takes a log-linear form,

its estimation relies only on positive trade flows. In addition, as I also presented, trade flows between two countries are often asymmetric (23% here): there is positive trade into one direction, but zero to the other. This feature of the actual data also contradicts the traditional gravity equation that predicts symmetric trade flows between pairs of countries.

These facts provide challenges for the consistent estimation of the log-linear gravity equation. Thus, estimation methods are needed that take into account the limited dependent variable with frequent zeros, because omitting the observations with zero exports between pairs of countries may result in biased OLS estimates of the gravity equation. This problem occurs because it is likely that trade occurs between countries as an outcome of other variables. Since the existence of trade relationship between countries may be systematically correlated with unobserved variables that affect the intensity of trade as well, using only country-pairs with positive export flows might produce biased coefficient estimates of the log-linear gravity equation (Wooldridge, 2003). In other words, there is a potential selection bias in this case that arises due to the self selection of countries into trade partners. In addition, the country-pair level unobserved heterogeneity may also affect the asymmetry of trade flows between two countries through the fraction of exporting firms between a origins and destinations, resulting in an additional potential source of bias.

Therefore I implement estimation methods that are able to take into account the prevalence of zero trade flows and asymmetries by exploiting the information about the extensive margin of trade as well. Hence I estimate the effect of emigrants on the probability of an existing exporting relationship between countries using a Probit model. Then I perform two-step estimation procedures that use the information about the extensive

margin coming from the Probit model to construct consistent estimates of the gravity equation. First I use a Heckman selection model that is able to account for the self selection of countries into trading partners by including the inverse Mills ratio as additional explanatory variable in the gravity equation. Then I implement empirically the recently developed model of Helpman, Melitz and Rubinstein (2008) that is able to correct not only for the potential selection bias but also for the omitted variable bias arising due to country pair-level unobserved heterogeneity that causes asymmetric trade flows between countries. Helpman, Melitz and Rubinstein (2008) build an underlying Melitz-type heterogeneous firm trade model for deriving an extended gravity equation that is able to reproduce both the frequent zeros and asymmetry present in the data. This model makes predictions about the extensive margin as well by endogeneizing the number of exporting firms. Although the model relies on firm level heterogeneity, it is linked to country level trade flows through marginal exporters in each country, therefore the estimation does not require firm-level data. The empirical implementation of the Helpman, Melitz and Rubinstein (2008) model is similar to an extended Heckman selection model, since it uses the predicted components of the first stage Probit as additional explanatory variables in the second stage gravity equation:

$$x_{ij} = \beta_0 + \alpha_j + \chi_i - \gamma d_{ij} + \ln\left\{\exp[\delta(\hat{z_{ij}}^* + \hat{\overline{\eta}}_{ij}^*)] - 1\right\} + \beta_{u\eta}\,\hat{\overline{\eta}}_{ij}^* + \varepsilon_{ij},$$

where d_{ij} is the trade barrier of interest, α_j and χ_i are origin and destination dummies, $\hat{\overline{\eta}}_{ij}^*$ is the inverse Mills ratio that corrects for the selection bias and the non-linear term captures the proportion of exporting firms from a certain origin to a given destination to correct for the omitted variable bias arising due to the asymmetry in trade flows between pairs of countries. Helpman, Melitz and Rubinstein (2008) argue that this estimation procedure is able to produce consistent estimates of the gravity equation.

5.2 Results

First I present the OLS estimation results, then I introduce origin and destination country fixed-effects, and finally I implement two-step estimation procedures for analyzing the emigration-export link.

5.2.1 OLS

Column (1) in Table 2 reports the OLS estimates yielded by the baseline specification of the traditional gravity equation using all the 49 952 country-pairs of the sample. The coefficient estimate of emigrants is the in the focus of interest of this thesis. The estimated effect is positive and statistically significant at 1%. The elasticity suggests that *ceteris paribus* a 1% increase in the emigrant population in a given destination country increases exports from the origin country of the emigrants to that country by 0.129%. This magnitude matches the usual findings of the literature investigating the migration-trade link. In addition, all the estimated coefficients of the usual gravity control variables have the expected sign: bilateral exports flows indeed depend positively on GDP and GDP per capita of both countries and negatively on distance. The estimate on distance is close to -1, which coincides with the usual finding of the literature using this method. Export flows are also predicted to be higher for contiguous countries and in case of existing trade agreement, common legal origin, official language and colonial ties. All these explanatory variables are statistically significant at 1%. However, using the same currency seems to have no effect on export flows between countries after controlling for the previously mentioned variables. Based on these results one can claim that a dataset that covers all country-pair combinations of the world is able to reproduce the major findings of the literature exploring the migration-trade link from the perspectives of some single developed

countries.

As a next step I divided the global dataset into subsamples based on the World Bank country income classification. This classification system characterizes each country by its GNI per capita and groups them into one of the following income categories: low income countries ($1,025 or less), middle income countries ($1,026-$12,475) and high income countries ($12,476 or more) (The World Bank, 2013*a*). Based on this information I constructed 9 subsamples that cover each possible combination of origin and destination country income groups. Then I re-estimated the baseline OLS specification for all the 9 subsamples separately. The results are presented in columns (2) to (10) of Table 2. Investigating the reported coefficient estimates reveal interesting facts about the emigrant-export relationship.

Table 2: OLS estimates across different income groups

Inc. Orig.	All	Low			Middle			High		
Inc. Dest.	All	Low	Middle	High	Low	Middle	High	Low	Middle	High
dep var: log export	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log emigrants	0.129***	-0.0504	0.0382	0.0845*	-0.0514	0.121***	0.163***	0.00633	0.128***	0.0808***
	(0.00847)	(0.0637)	(0.0428)	(0.0444)	(0.0453)	(0.0178)	(0.0204)	(0.0343)	(0.0178)	(0.0210)
log gdp_o	1.088***	0.906***	0.645***	1.197***	1.073***	1.096***	1.150***	1.066***	1.134***	1.145***
	(0.0118)	(0.183)	(0.0944)	(0.0862)	(0.0566)	(0.0234)	(0.0280)	(0.0474)	(0.0273)	(0.0308)
log gdp per cap_o	0.0840***	0.453	0.974***	0.614**	-0.228**	-0.249***	0.0226	0.272**	0.00642	0.101
	(0.0145)	(0.485)	(0.257)	(0.257)	(0.107)	(0.0481)	(0.0530)	(0.111)	(0.0596)	(0.0658)
log gdp_d	0.739***	0.679***	0.651***	0.881***	0.466***	0.639***	0.880***	0.748***	0.780***	0.889***
	(0.0113)	(0.147)	(0.0559)	(0.0718)	(0.0756)	(0.0230)	(0.0310)	(0.0602)	(0.0211)	(0.0275)
log gdp per cap_d	0.0949***	-0.00241	-0.166	0.188	0.853***	0.0699	0.0311	0.316**	0.190***	0.101
	(0.0133)	(0.351)	(0.111)	(0.189)	(0.204)	(0.0435)	(0.0831)	(0.142)	(0.0379)	(0.0749)
log distance	-1.115***	-1.436***	-0.884***	-0.906***	-1.620***	-1.136***	-1.080***	-1.395***	-1.215***	-0.993***
	(0.0251)	(0.246)	(0.173)	(0.248)	(0.150)	(0.0542)	(0.0629)	(0.165)	(0.0480)	(0.0494)
contiguity	0.694***	2.229***	1.776***		0.247	0.662***	0.222		0.981***	0.0452
	(0.110)	(0.514)	(0.446)		(0.615)	(0.158)	(0.424)		(0.372)	(0.156)
trade agreement	0.823***	0.0413	1.627***	1.528	1.567***	1.685***	0.513***	3.942***	0.530***	0.282***
	(0.0618)	(0.402)	(0.359)	(0.955)	(0.411)	(0.142)	(0.143)	(0.288)	(0.114)	(0.104)
common legal o.	0.317***	0.190	0.270	0.577***	-0.156	0.218***	0.744***	0.457***	0.239***	0.250**
	(0.0408)	(0.345)	(0.207)	(0.217)	(0.179)	(0.0825)	(0.0962)	(0.147)	(0.0782)	(0.108)
colonial ties	0.680***	-3.233***	3.180***	1.540***	1.910***	0.156	0.924***	1.459***	0.139	0.216
	(0.0963)	(0.695)	(0.886)	(0.409)	(0.411)	(0.353)	(0.237)	(0.254)	(0.155)	(0.171)
common language	0.586***	0.392	0.548**	0.306	0.448*	0.480***	0.520***	0.108	0.761***	0.810***
	(0.0536)	(0.358)	(0.252)	(0.283)	(0.255)	(0.102)	(0.143)	(0.181)	(0.118)	(0.145)
common currency	0.103	0.0200	-0.546	-0.400	0.279	0.972**	3.458***	-0.617	1.414***	-0.226**
	(0.138)	(0.429)	(0.413)	(0.594)	(0.578)	(0.423)	(0.635)	(0.724)	(0.541)	(0.0979)
Observations	16,643	266	755	698	1,053	3,968	2,924	1,111	3,686	2,088
R-squared	0.654	0.478	0.341	0.441	0.363	0.546	0.640	0.545	0.695	0.781

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The first striking observation is that the estimated coefficients of emigrants coming from low income countries and living in destination countries of any income levels are not statistically significant. (The coefficient of emigration between low origin and high income destination countries is statistically significant at 10% level, but considering the large size of the dataset this significance is still negligible.) This finding implies that emigrants from low income origin countries have no effect on export flows from their origin to their destination country. There are many explanations for this intuitive result. Countries in the low income subsample mostly include Sub-Saharan ones, and some Asians such as Bangladesh and Afghanistan. Since the quality of education is very poor in these countries, people who emigrate from there are mostly unskilled persons who live as refugees in the destination countries. The information carrier mechanism is less likely to work in this case, because these people are rarely employed in the business sector of the destination countries. Thus they have no significant role in the flow of informal business practices, they do not bear important market-related information or do not facilitate contract enforcement. Furthermore, the preference diffusion channel is also unlikely to work in this case, because these people with very low income do not create considerable demand for the cultural goods produced in their home country. In addition, since the exports of these low income countries are mostly limited to agricultural goods that are homogenous goods, the role of extra information carried by emigrant networks is less emphasized in their trade.

The second observation is also related to low income countries: emigrants from origin countries with any level of development who live in low income countries have no significant effect on exports to low income countries. A possible explanation for this result could be that low income countries mostly import necessity goods, and as mentioned before the

trade of such homogenous goods does not require much information to match sellers and buyers. Furthermore, since these countries are not attractive destinations for emigrants, most people who were born in higher income countries and live in low income countries are volunteer workers, employees of humanitarian organizations, who do not have business contacts that might be linked to trade. However, these higher income foreigners could provide a potential demand for exporting goods from their origin countries to low income countries, but this preference diffusion channel neither seems to work in this case. A possible reason for this finding is the small size of the immigrant population of low income countries.

The third finding is that the relationship between emigrants and export flows is stronger between countries of different income levels: the coefficient of emigrants is larger from middle origin to high income destination countries than to similar middle income destination countries, and larger from high income origin countries to middle income destination countries than to similar high income destination countries. The importance of the cross-country differences in the emigrant-trade link emphasizes again the role of the information network maintained by migrants in trade relationships.

These results overall imply that most part of the relationship between emigrants and export flows is caused by people who move from middle income countries to high income countries. This can also explain the similar findings of the literature considering the relationship only from single developed country perspectives. In addition, the role of the migrant information network is enlarged by differences between countries, but this relationship to exist requires a critical level of development in both the origin and destination countries.

5.2.2 Country fixed-effects

It is important to emphasize that the OLS estimates yielded by the previous specification might be biased because of the endogeneity of emigrant stocks in the export flow gravity equation. Thus the OLS estimates might not reflect the true relationship because of the potential existence of unobserved heterogeneity. In this case the relationship between emigration and exports would be driven by a third unobserved factor that affects both, such as a general attractiveness of the destination countries. Therefore other econometric methods are needed to properly identify the effect of emigrant stocks on trade flows.

For this reason I added origin and destination country fixed-effects to the previous model to control for country-level unobserved factors.The estimated equation became the following:

$$log\, export_{ij} = \alpha + \beta_1 \mathbf{X}_{ij} + \beta_2 log\, emigrants_{ij} + \beta_3 D_i + \beta_4 D_j + \varepsilon_{ij},$$

where $\mathbf{X}_{ij}$ is a vector of country-pair level explanatory variables, D_i and D_j are country specific dummy variables that take the value 1 for a given destination and origin country. The use of this specification is also justified by the augmented gravity model proposed by Anderson and van Wincoop (2003). From now on I discuss only estimation methods that correspond to the whole sample, because the previously used development level is a country specific variable that is now controlled by the origin and destination country fixed-effects.

Column (1) in Table 3 reports the results of the OLS estimation of this specification. Compared to the estimates without fixed effects (column (1) of Table 2), all the coefficients have the same signs, only the magnitudes of the elasticities have changed. The

coefficient estimate of emigrants grew from 0.129 to 0.206. The estimates of distance and colonial ties also increased in absolute value. The coefficients of the other country-pair specific gravity controls decreased.

Table 3: Country fixed-effects

VARIABLES	(1) log export	(2) log export	(3) log export	(4) log export	(5) log export
log emigrants	0.206***	0.246***	0.172***	0.216***	0.162***
	(0.0108)	(0.0218)	(0.0199)	(0.0224)	(0.0111)
log distance	-1.400***	-1.396***	-1.448***	-1.448***	-1.271***
	(0.0283)	(0.0285)	(0.0481)	(0.0473)	(0.0291)
contiguity	0.391***	0.443***	0.225*	0.239*	0.170
	(0.106)	(0.106)	(0.129)	(0.126)	(0.104)
trade agreement	0.431***	0.451***	-0.0852	-0.0652	0.344***
	(0.0597)	(0.0602)	(0.0766)	(0.0776)	(0.0592)
common legal o.	0.315***	0.316***	0.548***	0.560***	0.284***
	(0.0370)	(0.0370)	(0.0601)	(0.0598)	(0.0368)
colonial ties	0.711***	0.741***	0.327***	0.342***	0.472***
	(0.0984)	(0.0987)	(0.124)	(0.118)	(0.0997)
common language	0.518***	0.516***	-0.0809	-0.103	0.449***
	(0.0545)	(0.0545)	(0.112)	(0.110)	(0.0545)
common currency	0.0274	0.0239	-0.859***	-0.759***	-0.0393
	(0.134)	(0.134)	(0.114)	(0.111)	(0.132)
logemigrants^2		-0.00376**			
		(0.00167)			
log cultural dif			-0.326***	-0.278***	
			(0.0953)	(0.0952)	
log emigrants_#log cultural dif			0.0384***	0.0298**	
			(0.0125)	(0.0124)	
low corruption				0.669***	
				(0.185)	
low corruption#log emigrants				-0.110***	
				(0.0201)	
log immigrants					0.169***
					(0.0109)
Observations	18,474	18,474	4,837	4,837	18,474
R-squared	0.735	0.736	0.823	0.824	0.739
Origin FE	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As a next step I extended this baseline fixed-effects specification with additional control variables that help to get a better insight into the underlying mechanisms of the emigrant-trade link.

First I wanted to investigate whether the increase of the emigrant population has a constant marginal effect on export flows between countries to test both the information and the preference diffusion channels. The results are presented in column (2) of Table 3. The estimated coefficient of the quadratic term is statistically significant and negative, thus the marginal effect of emigrant stocks on export flows from the origin to the destination of emigrants is decreasing. This observation can be explained by the information network effect of emigrants, since only a limited number of emigrants are able to provide additional country related information that lowers informal trade barriers, and after a critical amount of emigrants this channel becomes saturated. At the same time, the diminishing marginal effect has implication about the preference channel. If this mechanism worked as expected (more emigrants from a certain country create demand in their destination country for the goods produced in their origin country), then the marginal effect of emigrants on exports would not be diminishing, since more emigrants create more demand. However, this is not true here, so either the preference channel does not exist and the trade facilitating role emigrants is limited to the information network effect only, or the preference channel works differently in reality: it is possible that after the number of immigrants from a given country reaches a critical level, local production of traditional goods starts in the destination countries, providing a substitution for importing them from the origin countries of migrants.

Then, based on White and Tadesse (2008), I included a variable that measures cultural differences between countries and its interaction term with emigrants. The estimates yielded by this specification are presented in column (3) of Table 3. The estimated coefficient of cultural difference is statistically significant and it has an intuitive negative

sign: cultural differences imply informal barriers for trade. The estimate of emigrants and its interaction term with the cultural difference variable mean that exports from a given origin country to a certain destination country are significantly increased by the amount of its emigrants living in the destination country, and this emigrant-trade link is higher when the extent of cultural difference between the two countries is larger. This finding provides additional evidence for the existence of the information network effect of emigrants on export flows, since cultural difference is a source of information asymmetry that creates additional informal trade barrier between countries. Adding cultural differences to the model also changed the significances and signs of the other gravity controls: contiguity, common language and trade agreement do not seem to matter when controlling for cultural differences, but the coefficient of common currency became significant with a counterintuitive negative sign.

I developed further the model by adding a variable that captures the quality of the institutional environment in both countries. The dummy variable *low corruption* takes value 1 if the ability of controlling corruption is high both in the origin and in the destination country. This variable was constructed using country level data on corruption control, then creating dummy variables indicating low corruption levels if the value of the corruption control index is above its median in the given country. Then I transformed the origin and destination country-specific variables to a country-pair level one by constructing the *low corruption* variable by multiplying the country-level dummy variables. As column (4) shows, the results are intuitive again. The individual term of the variable indicating low corruption in both countries is statistically significant and positive, meaning that the well functioning institutional framework facilities trade. The interaction term

of *low corruption* with emigrants shows that the emigrant-trade relationship diminishes when considering less corrupt countries. This result provides evidence for the role of social networks maintained by emigrants in contract enforcement that is more highlighted in trade between more corrupt countries. The other variables have similar signs and magnitudes as in the previous specification.

Finally I tested the existence of the preference diffusion mechanism by adding immigrants as a control variable to the baseline fixed-effects specification. As mentioned earlier, the preference link should work only in case of migration and trade flowing to the same directions, but the information channel works in both directions. Therefore the coefficient of emigrants should be higher than that of immigrants. The estimation results are presented in column (5) of Table 3. However, the coefficient of immigrants is higher than that of emigrants, so I am not able to find evidence for the classical preference diffusion mechanism. This result is in line with the studies of other authors. An alternative explanation could be again that this link is distorted by substation with local production.

The results of the fixed-effects estimation still support the existence of the emigrant-trade relationship and the underlying role of the information network effect. Although the inclusion of country fixed-effects is able to control for country level unobserved factors that affect both emigration and trade, there still exists some sort of unobserved heterogeneity that may bias the results. The level of this remaining unobserved heterogeneity corresponds to country-pairs, such as a historically tight relationship and current or historical conflicts between two countries. Such unobserved country-pair level factors might affect the intensity of trade and the formation of trading relationships between countries as well.

Thus, as explained in Chapter 5.1.2., there is a potential self-selection of countries into trade partners. Furthermore, it has to be highlighted that all the results presented until now were yielded by estimating a log-linear equation using observations only with positive export flows between countries. However, as mentioned above 62% of the dataset consists of observations with zero trade flows between countries, and there are many country pairs (23%) with positive exports in one direction only, which asymmetry cannot be explained by a simple gravity equation. Therefore the previous estimation methods omitted important information about the role of emigrants in the formation trade partnerships between countries. As a result, the traditional estimates of the gravity equation might be biased because of the country-pair level unobserved heterogeneity and the omission of information about the extensive margin. These problems have to be solved econometrically to identify the true relationship between emigrants and export flows.

5.2.3 Two-step estimation procedures

There are some two-step estimation methods that can get rid of these potential sources of bias. These estimation procedures first exploit information about the extensive margin by estimating the probability of trade between countries with a Probit model. Then the predicted components of the first stage selection equation are used as additional explanatory variables in the second stage gravity equation to correct for the possible biases. Accordingly, in this subchapter I estimate first a Probit model, then use its results to implement a Heckman selection model that corrects for the potential selection bias, and finally I implement an extended version of it recently developed by Helpman, Melitz and Rubinstein (2008) that is able to deliver consistent estimates of the gravity equation by controlling also for factors that cause asymmetries in trade flows between pairs of countries.

5.2.3.1 The selection into trade partners: Probit

For implementing both two-step estimation methods, the identification of the second stage equations requires an exclusion restriction to hold: there should be some explanatory variables that affect only the probability of trade participation, but not intensity to be able to distinguish sample selection from a misspecification of functional form (Wooldridge, 2003). In other words, these excluded variables affect only the fixed costs of exporting, but do not enter into the second stage export flow equation by influencing also the variable costs. I use two possible explanatory variables for this purpose based on the recommendations of Helpman, Melitz and Rubinstein (2008). The first is the existence of a common official language and the alternative one is an indicator of high fixed costs connected to exporting in both countries of the pair. I constructed the latter variable by collecting data from the World Development Indicators (The World Bank, 2013*b*) database on country-level information about regulatory costs that affect the fixed costs of exporting, such as the number of documents needed for exporting and importing, and the average number of procedures needed to enforce a contract. Then I constructed a dummy variable that takes the value 1 if these measures are above their median. Then I used these country-level variables to construct a country-pair level one that indicates high fixed costs of exporting in both countries.

I estimated a Probit model of the baseline fixed-effect specification by adding also the variables to be excluded from the second stage:

$$P(export_{ij} = 1) = \phi(\alpha + \beta_1 \mathbf{X}_{ij} + \beta_2 log\, emigrants_{ij} + \varepsilon_{ij}),$$

where $export_{ij}$=1 if country i exports to country j and $\mathbf{X}_{ij}$ is a vector of country-pair specific gravity controls.

Column (1) and (2) of Table 4 report the estimated coefficients and average marginal effects yielded by the Probit model including common official language as an explanatory variable as well. The results reveal that the same factors affect the probability of trade between countries that influence also the intensity of trade. The estimated marginal effect of emigrants suggests that on average a 1% increase in the emigrant stock in a given country increases the probability of exporting to that country from the origin of the emigrants by 0.0343%. Most of the explanatory variables have the expected signs: it has a higher probability to export to closer countries, to countries having trade agreement with, that have common legal origin, language and currency. The estimate of historical colonial ties is not significant, meaning that colonial ties does not influence the probability of trade relationship between countries. However, the coefficient of contiguity is striking: it indicates that a common border decreases the probability of trade between countries. This finding is in line with the first stage results of Helpman, Melitz and Rubinstein (2008). A possible explanation could be that territorial conflicts are in the background of this negative coefficient estimate. Column (3) and (4) of Table 4 present the estimated coefficients and marginal effects after adding the alternative excluded variable indicating high regulatory fixed costs in both countries. Neither the coefficient estimate nor the marginal effect of the variable indicating high regulatory costs is significant. A possible explanation for this counterintuitive result can be that this variable is not able to capture the amount of fixed costs of exporting between countries, because a better measure than the number of procedures and documents are needed to quantify regulatory costs.

Table 4: Probit estimates of the selection equation

VARIABLES	(1) $export_{ij}$	(2) ME	(3) $export_{ij}$	(4) ME
log emigrants	0.0908***	0.0343***	0.0908***	0.0343***
	(0.00922)	(0.00346)	(0.00922)	(0.00346)
log distance	-0.894***	-0.338***	-0.894***	-0.338***
	(0.0247)	(0.00937)	(0.0247)	(0.00937)
contiguity	-0.608***	-0.239***	-0.608***	-0.239***
	(0.196)	(0.0764)	(0.196)	(0.0764)
trade agreement	0.886***	0.267***	0.886***	0.267***
	(0.108)	(0.0230)	(0.108)	(0.0230)
common legal o.	0.170***	0.0634***	0.170***	0.0634***
	(0.0260)	(0.00961)	(0.0260)	(0.00961)
colonial ties	-0.0533	-0.0203	-0.0533	-0.0203
	(0.247)	(0.0948)	(0.247)	(0.0948)
common language	0.494***	0.173***	0.494***	0.173***
	(0.0395)	(0.0126)	(0.0395)	(0.0126)
common currency	0.437**	0.150***	0.437**	0.150***
	(0.174)	(0.0521)	(0.174)	(0.0521)
regulatory costs			-0.0289	-0.0044
			(0.187)	(0.0288)
Observations	35,042	35,042	35,042	35,042
Origin FE	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.2.3.2 Heckman selection model

Then I re-estimated the baseline gravity equation with country fixed-effects by implementing the Heckman selection procedure. Table 5 contains the second stage OLS estimation results of the same specifications as in Table 3 with the inverse Mills ratio derived from the first stage Probit predicted values as an additional explanatory variable. I used the existence of common language between countries as an excluded variable from the second stage. The inverse Mills ratio is statistically significant at 1% in all specifications, meaning that there is significant selection bias in the estimates of the traditional gravity equation reported in Table 3, but the new estimates yielded by the Heckman

selection models are corrected for this self-selection. However, comparing the estimated coefficients of Table 3 and Table 5 reveals that the magnitude of the selection bias is quite small that is consistent with the finding of Helpman, Melitz and Rubinstein (2008).

Table 5: Heckman selection model

VARIABLES	(1) log export	(2) log export	(3) log export	(4) log export	(5) log export	(6) log export
log emigrants	0.224***	0.306***	0.170***	0.207***	0.227***	0.177***
	(0.0104)	(0.0214)	(0.0180)	(0.0193)	(0.0197)	(0.0107)
log distance	-1.503***	-1.508***	-1.397***	-1.405***	-1.377***	-1.354***
	(0.0297)	(0.0297)	(0.0473)	(0.0472)	(0.0474)	(0.0307)
contiguity	0.329***	0.426***	0.263**	0.270**	2.350***	0.115
	(0.103)	(0.105)	(0.128)	(0.127)	(0.418)	(0.103)
trade agreement	0.358***	0.389***	-0.0470	-0.0492	-0.0153	0.283***
	(0.0667)	(0.0671)	(0.0793)	(0.0816)	(0.0816)	(0.0663)
common legal o.	0.462***	0.466***	0.489***	0.505***	0.508***	0.408***
	(0.0346)	(0.0346)	(0.0554)	(0.0553)	(0.0551)	(0.0344)
colonial ties	0.895***	0.957***	0.329***	0.332***	0.386***	0.626***
	(0.112)	(0.113)	(0.142)	(0.141)	(0.141)	(0.112)
common currency	0.147	0.143	-0.798***	-0.726***	-0.683***	0.0614
	(0.122)	(0.122)	(0.165)	(0.165)	(0.165)	(0.121)
logemigrants^2		-0.00770***				
		(0.00175)				
log cultural dif			-0.351***	-0.302***	-0.241**	
			(0.0880)	(0.0893)	(0.0898)	
log emigrants#l cult dif			0.0394***	0.0321***	0.0216*	
			(0.0121)	(0.0122)	(0.0123)	
low corruption				0.629***	0.628***	
				(0.156)	(0.156)	
low corruption#l emigr				-0.0942***	-0.0943***	
				(0.0184)	(0.0183)	
contiguity#l emigr					-0.217***	
					(0.0415)	
log immigrants						0.172***
						(0.0103)
iMills	0.344***	0.409***	-0.793***	-0.655***	-2.671***	0.782***
	(0.0522)	(0.0541)	(0.138)	(0.143)	(0.121)	(0.112)
Observations	46,230	46,230	4,593	4,593	4,593	46,230
R-squared	0.716	0.716	0.812	0.812	0.812	0.732
Origin FE	yes	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes	yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The inclusion of the inverse Mills ratio does not change the main qualitative findings of Subchapter 5.2.2. The trade facilitating role emigrants is more important in case of larger cultural differences, and it is less emphasized in trade between less corrupt or contiguous countries. All these results support the existence of the underlying information network

effect. However, it can be observed in column (6) that the coefficient of emigrants is higher than that of immigrants, meaning that migration flowing in the same direction as export has a larger effect. This provides an evidence for the existence of the preference diffusion channel, since this works only in the same direction of trade, while the information network effect exists in both directions. Therefore the previous findings of the literature about the lack of the preference effect can be explained by the existence of a selection bias of matching countries into trade partners.

5.2.3.3 Helpman, Melitz and Rubinstein (2008)

As a next step I implemented empirically the Helpman, Melitz and Rubinstein (2008) model. Therefore I added an extra explanatory variable of a non-linear form to the second stage export flow equation of the previous Heckman selection model. This method is able to control for the unobserved country-pair level heterogeneity causing asymmetries in trade flows between two countries. Similarly to the inverse Mills ratio, this can be also derived from the predicted components of the first stage Probit selection equation.

The estimation of the nonlinear second stage gravity equation caused unexpected difficulties, since the `nl` command of Stata found the equation containing 2×224 origin and destination dummies to be too long to be able to estimate. To handle this problem, I decomposed the second stage equation into two parts: first I separated the nonlinear term, than I ran simple OLS regressions on the remaining part that are able to deal with the large number of independent variables. Therefore, I started with constructing two standard errors-wide confidence intervals around the estimated coefficient δ of the non-linear term by Helpman, Melitz and Rubinstein (2008), because the original article used a similar country-pair level word-wide dataset. Then to obtain potential values for δ in my

case, I divided this confidence interval into 10 equal parts. As a next step I substituted these values into the non-liner term and subtracted them from the log export variable to construct new dependent variables. Finally, using the logic of least squares, I manually iterated the linear equations by running OLS on the same set of independent variables, only changing the dependent variable that correspond to different possible values of δ. I chose the value of δ that minimized the sum of squared residuals of the OLS regressions. Then I repeated this procedure for all the previously discussed specifications. More precise standard errors could by obtained by using nonlinear least squares for example. I used again common language as an excluded variable in the selection equation.

The coefficient estimates yielded by the Helpman, Melitz and Rubinstein (2008) estimation method are presented in Table 6. A substantial decrease can be observed in the absolute value of the coefficients. This is in line with the results of Helpman, Melitz and Rubinstein (2008). Compared to the estimates of the baseline fixed-effects specification, for example as column (1) of Table 6 shows, the estimated coefficient of emigrants dropped from 0.206 to 0.140. These findings overall imply that the traditional estimates of the gravity equation are upward biased because of the unobserved country-pair level heterogeneity, but this bias can be corrected by the inclusion of a variable that measures the proportion of exporters from given origin to given destination countries.

Investigating the underlying channels, the estimates of the interaction terms does not change substantially compared to the previous estimation methods, therefore the previous interpretations are robust for all the estimation methods I use. However, testing the preference channel yields contradictory results. Column (2) shows that the quadratic term of emigrants becomes insignificant, meaning that the marginal effect emigrants on export flows is constant. This can be interpreted again as an evidence for the existence

of the preference diffusion channel, since as mentioned before it is reasonable to assume that the information network-effect alone becomes saturated after a certain number of emigrants. However, according to column (6) the estimated coefficient of emigrants who move to the same direction as export flows is smaller than that of immigrants who move to the opposite direction that seems as an evidence against the preference mechanism. These contradictory results suggest that the usual method of comparing the estimates of opposite migrants stocks is not enough alone to draw conclusion about the preference channel, but additional analysis is needed.

Table 6: Helpman, Melitz, Rubinstein 2008

VARIABLES	(1) lhs18	(2) lhs17	(3) lhs54	(4) lhs61	(5) lhs60	(6) lhs26
log emigrants	0.140***	0.168***	0.166***	0.192***	0.212***	0.110***
	(0.0109)	(0.0228)	(0.0181)	(0.0195)	(0.0199)	(0.0112)
log distance	-1.064***	-1.053***	-1.187***	-1.193***	-1.166***	-1.014***
	(0.0328)	(0.0328)	(0.0475)	(0.0476)	(0.0477)	(0.0341)
contiguity	0.865***	0.910***	0.437***	0.443***	2.503***	0.578***
	(0.115)	(0.117)	(0.128)	(0.128)	(0.423)	(0.115)
trade agreement	0.222***	0.221***	-0.0884	-0.100	-0.0668	0.215***
	(0.0781)	(0.0785)	(0.0796)	(0.0822)	(0.0823)	(0.0779)
common legal o.	0.278***	0.276***	0.374***	0.387***	0.389***	0.259***
	(0.0358)	(0.0358)	(0.0556)	(0.0557)	(0.0556)	(0.0358)
colonial ties	0.852***	0.876***	0.457***	0.459***	0.512***	0.607***
	(0.124)	(0.125)	(0.142)	(0.142)	(0.142)	(0.124)
common currency	0.553***	0.539***	-0.780***	-0.732***	-0.689***	0.472***
	(0.153)	(0.153)	(0.166)	(0.166)	(0.166)	(0.153)
logemigrants^2		-0.00285				
		(0.00194)				
log cultural dif			-0.346***	-0.304***	-0.244***	
			(0.0892)	(0.0906)	(0.0911)	
log emigrants#l cult dif			0.0331***	0.0276**	0.0172	
			(0.0122)	(0.0123)	(0.0125)	
low corruption				0.490***	0.489***	
				(0.158)	(0.158)	
low corruption#l emigr				-0.0658***	-0.0659***	
				(0.0186)	(0.0185)	
contiguity#l emigr					-0.215***	
					(0.0420)	
log immigrants						0.152***
						(0.0107)
$\hat{\bar{\eta}}^*_{ij}$	0.751***	0.754***	0.385***	0.472***	0.470***	0.776***
	(0.0608)	(0.0621)	(0.145)	(0.148)	(0.148)	(0.0608)
δ	0.6336	0.6508	0.0144	0.0141	0.0141	0.496
Observations	17,548	17,548	4,837	4,837	4,837	17,548
R-squared	0.528	0.525	0.757	0.758	0.759	0.554
Origin and Destination FE	yes	yes	yes	yes	yes	yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6 Conclusion

In this thesis I investigated the relationship between emigrant stocks of a given origin, living in a certain destination country and export flows from the emigrants' origin country to their destination country. Contrary to the large pool of existing literature investigating the migration-trade link, this paper presented a twofold contribution by exploring the relationship from a global perspective looking also at developing countries and exploiting econometrically the information content of the prevalence of country-pairs with zero trade flows in the data. Accounting for these facts revealed interesting insights about the trade-creating role of emigrant networks.

My results confirmed the previous findings of the literature by showing a positive significant relationship between emigrant stocks and export flows when looking at the whole world, even after controlling for the sources of omitted variable bias. However, I showed that the trade facilitating role of emigrants is limited only to trade between countries above a certain level of development. In addition, I found that the main source of this positive relationship is the information network maintained by emigrant diasporas that contributes to the reduction of information asymmetries; while I got only contradictory results about the existence of the preference diffusion channel.

My results should be interpreted carefully, because I was not able to rule out econometrically the potential existence of reverse causality. However, this fact provides guidelines for further research, possibly by finding and exploring the effects of exogenous migration shocks on trade flows.

Since trade is thought to be welfare improving, the findings of this thesis may have interesting implications for immigration policy making in an era of deteriorated public opinion about the possibilities of assimilation.

References

Anderson, J., and E. van Wincoop. 2003. "Gravity with gravitas: a solution to the border puzzle." *American Economic Review*, 93: 170–192.

Anderson, J. E., and E. van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature*, 42: 691–751.

Bastos, P., and J. Silva. 2012. "Networks, firms, and trade." *Journal of International Economics*, 87: 352–364.

Borjas, G. J. 2003. "The Labor Demand Curve Is Downward Sloping: Reexamining The Impact Of Immigration On The Labor Market." *The Quarterly Journal of Economics*, 118(4): 1335–1374.

Bratti, M. and De Benedictis, L. and Santoni, G. 2013. "On the pro-trade effects of immigrants." Working Paper.

Combes, P., M. Lafourcade, and T. Mayer. 2005. "The trade-creating effects of business and social networks: evidence from France." *Journal of International Economics*, 66: 1–29.

Development Research Centre on Migration, Globalisation and Poverty. 2013. "Global Migrant Origin Database." http://www.migrationdrc.org/research/typesofmigration/global_migrant_origin_database.html Retrieved in March, 2013.

Dunlevy, J. A. 2006. "The Influence of Corruption and Language on the Protrade Effect of Immigrants: Evidence from the American States." *The Review of Economics and Statistics*, 88(1): 182–186.

Friedbergs, R. M., and J. Hunt. 1995. "The Impact of Immigrants on Host Country Wages, Employment and Growth." *The Journal of Economic Perspectives*, 9: 23–44.

Genc, M. and Gheasi, M. and Nijkamp, P. and Poot, J. 2011. "The Impact of Immigration on International Trade: A Meta-Analysis." IZA Discussion Paper Series.

Girma, S., and Z. Yu. 2002. "The link between immigration and trade: evidence from the United Kingdom." *Weltwirtschaftliches Archiv*, 138: 115–130.

Gould, D. 1994. "Immigrant links to the home country: empirical implications for U.S. bilateral trade flows." *Review of Economics and Statistics*, 76: 302–316.

Grossman, J. B. 1982. "The Substitutability of Natives and Immigrants in Production." *The Review of Economics and Statistics*, 64: 596–603.

Hagenaars, J. A., L. Halman, and G. Moors. 2003. *The Cultural Diversity of European Unity.* Boston, MA: Koninklijke Brill.

Head, K., and J. Ries. 1998. "Immigration and Trade Creation: Econometric Evidence from Canada." *Canadian Journal of Economics*, 31: 47–62.

Head, K., T. Mayer, and J. Ries. 2010. "The erosion of colonial trade linkages after independence." *Journal of International Economics*, 81: 1–14.

Helpman, E., M. Melitz, and Y. Rubinstein. 2008. "Estimating trade flows: trading partners and trading volumes." *Quarterly Journal of Economics*, 123: 441–487.

Inglehart, R., and C. Welzel. 2005. *Modernization, Cultural Change, and Democracy : The Human Development Sequence.* Cambridge University Press.

Inglehart, R., M. Basanez, J. Diez-Medrano, L. Halman, and R. Luijkx. 2004. *Human Beliefs and Values: A Cross-Cultural Sourcebook based on the 1999-2002 Values Surveys.* Siglo Veintiuno Editores.

International Organization for Migration. 2013. "Global Estimates and Trends About Migration." http://www.iom.int/cms/en/sites/iom/home/about-migration/facts–figures-1.html Retrieved in April, 2013.

Jacks, D. S., and K. Pendakur. 2010. "Global Trade and the Maritime Transport Revolution." *The Review of Economics and Statistics*, 92: 745–755.

Melitz, M. J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71: 1695–1725.

Peri, G., and F. Requena-Silvente. 2010. "The trade creation effect of immigrants: evidence from the remarkable case of Spain." *Canadian Journal of Economics*, 43: 1433–1459.

Rauch, J. 1999. "Networks versus markets in international trade." *Journal of International Economics*, 48: 7–35.

Rauch, J., and V. Trindade. 2002. "Ethnic Chinese networks in international trade." *Review of Economics and Statistics*, 84: 116–130.

Santos Silva, J. M. C., and S. Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

The World Bank. 2011. *Migration and Remittances Factbook.*

The World Bank. 2013*a*. "World Bank Country Classifications." http://data.worldbank.org/about/country-classifications/country-and-lending-groups Retrieved in March, 2013.

The World Bank. 2013*b*. "World Development Indicators." http://data.worldbank.org/data-catalog/world-development-indicators Retrieved in March, 2013.

Tinbergen, J. 1962. *Saping the World Economy: Suggestions for an International Economic Poloicy.* Twentieth Century Fund, New-York.

United Nations. 2013. "United Nations Commodity Trade Statistics Database." http://comtrade.un.org/ Retrieved in March, 2013.

Wagner, D., K. Head, and J. Ries. 2002. "Immigration and the trade of provinces." *Scottish Journal of Political Economy*, 49: 507–525.

White, R., and B. Tadesse. 2008. "Cultural Distance and the US Immigrant-Trade Link." *The World Economy*, 31(8): 1078–1096.

Wooldridge, J. M. 2003. *Introductory econometrics: a modern approach.* South-Western.