Erasmus University Rotterdam

Erasmus School of Economics

THE IMPACT OF INTERNATIONALLY DISPERSED ETHNIC NETWORKS ON FOREIGN DIRECT INVESTMENTS

Abstract:

This study examines the role internationally dispersed ethnic networks have on countries' foreign investment stock by estimating a gravity model using cross sectional data from 2010. Although the study focusses mainly on the impact of ethnic Chinese networks, 86 ethnicities have been examined. Ethnic network strength between countries was measured by using a proxy variable, created by taking the product of two countries' migrant population from a respective ethnicity. Using a Poisson Pseudo Maximum Likelihood estimation, the likely existence of 30 ethnic networks has been distinguished, with the South Korean network being the most impactful. The Chinese ethnic network ranks 23d in terms of its impact on FDI.

Author: J.P. Meij

Thesis supervisor: Prof. dr. J. M. A. R. G. Viaene

Table of contents

1) Introduction

According to a report published by the United Nations (2017), the international migrant stock has grown more rapidly than the world's population over the past 17 years. Where in 2000 2.8% of the total world population was a migrant, this percentage had risen to 3.4% in 2017. The report states that migration has accounted for 42% of the population growth in Northern America, and that the European population would have decreased over the same time span were it not for migrants moving to European countries.

Several large economies have taken measures to restrict the flow of migrants entering the country. The Trump administration is currently constructing a wall along the southern border of the United States that is meant to repel migrants coming from Central and South America (The Economist, 2017). In Europe, the United Kingdom has decided to leave the European Union, partly because its citizens want to regain control over who enters and leaves the country (The Economist, 2016). In various European nations, including the Netherlands, this discourse has been adopted by politicians and is gaining support of parts of their respective populations.

Restricting the access to migrants can have severe consequences on a country's economy. Although the net effect of migration remains uncertain, much has been written on the social and economic impact of migration. Migration has been integrated in theoretical models as a prerequisite for a general equilibrium, and empirical papers have studied its determinants and implications.

This paper extends this field of the literature by studying the effect migration has on inward foreign direct investments (FDI). To be more concise, research has been conducted to demonstrate the impact internationally dispersed ethnic networks have on countries' inward FDI stock. Using cross sectional data, a gravity model is estimated using a Poisson pseudo maximum likelihood (PPML) estimation. Although this study mainly focusses on the impact of a Chinese network, the likely existence of thirty ethnic networks has been distinguished. Apart from the Chinese ethnic network, which has been distinguished in earlier studies, this study has distinguished, among others, a Korean, French, and a Venezuelan ethnic network.

This paper starts with an overview of the theory and literature regarding the impact migrants have on international economic exchanges, which leads to the derivation of the hypotheses. Afterwards, the data used, and the methods performed to test the hypotheses will be distinguished. This study finishes with the results and the conclusions derived from the estimated parameters. Lastly, implications for further research on the topic are provided.

2) Theoretical framework

2.1) Theory and literature

An extensive amount of literature has been written regarding the motives, determinants and impediments of international capital movements. Literature regarding FDI and trade are closely related. Both economic exchanges are subject to certain impediments that negatively affect the transaction. Some of these impediments, like transportation costs and tariff rates, are easily distinguishable. Multiple studies however have shown that ethnic and cultural dissimilarities can also complicate economic exchanges. When social, institutional, or cultural values like legislation or language differ between two parties, communication and mutual understanding requires more effort, thus costing more time and resources compared to interaction between parties with a similar background (Flisi and Murat, 2011).

Previous literature suggests that international social and business networks facilitate in overcoming cultural impediments in international economic exchanges. Rauch and Trindade (2002) and Rauch and Casella (2003) noted that incomplete information is a reason for international markets to not function optimally. They state that internationally dispersed ethnic minorities overcome this impediment by forming social and business networks. These networks act as channels where information is shared. Apart from serving as a channel for information these networks furthermore enhance contract enforcement. This can be beneficial when the jurisdictional systems of actors in a transaction vary, as it causes the penalties for not following a contract to differ (Ammon and Baiardi, 2016). Both the sharing of information and the enforcement of contracts facilitate international transactions.

This theory regarding ethnic networks in international economic exchanges is closely related to literature regarding the effect of migrants in bilateral trade and FDI. Migrants possess knowledge of both their country of origin and their country of residence. They are familiar with both markets, often have an international social network, and are multilingual. These characteristics make migrant employees valuable for internationally operating firms since it makes them less dependent on intermediaries to gather information (Javorcik et al., 2011). Migrants furthermore might prefer products originating from their home country, thereby increasing demand for imported goods.

Studies that have verified that migrants increase trade flows between their country of origin and their country of residence include, among others, Gould (1994), Head and Ries (1998), Girma and Yu (2002) and Genc et al. (2012). Studies on the enhancing effect of migrants regarding

FDI include Gao (2003), Murat and Pistoresi (2009), and Flisi and Murat (2011). Each of these studies used a gravity model, as introduced by Tinbergen (1962), as their base model to estimate the effects of migration on trade or FDI.

Ammon and Baiardi (2016) extended this field of studies by examining the firm level impact of ethnic Cantonese residing in the United States by examining firms located in Guangdong. In their study they excluded the ethnic Cantonese that were born in China, and thus focused on the effects of a network formed by second, third and fourth generation migrants. They concluded that when firms were likely to have some connection to the ethnic Cantonese in the US, it positively impacted whether firms were exporting, and their amount exported.

The studies mentioned all provide empirical evidence for the impact of migrants and ethnic networks between the migrants' country of residence and their country of origin. Rauch and Trindade (2002) deviated from these studies by pointing out that ethnic minorities affect trade beyond trade between their country of residence and their country of origin. In their study they distinguished a significant increase in the bilateral trade of heterogeneous goods between any given pair of countries induced by the ethnic Chinese population within these countries. This indicates the existence of an internationally dispersed ethnic Chinese network.

Felbermayr et al. (2010) revisited Rauch and Trindade's (2002) study and corrected econometric issues which may have led to inconsistent estimates in their estimations. They found that even though Rauch and Trindade's elasticities were overestimated, the Chinese network still induces trade creation of about 15%. They furthermore extended Rauch and Trindade by taking multiple nationalities and respective ethnic networks into account. They found that next to Chinese, there are multiple other ethnicities that influence bilateral trade, including a Turkish, Pakistani, and Mexican network.

In a similar fashion as Rauch and Trindade (2002), Tong (2005) provided empirical evidence for the existence of a Chinese network that has an impact on bilateral FDI. She furthermore distinguished the network effect when the investing country is either an industrial or a developing country, and whether the quality of institutions in the investing country is high or low. Tong is cited up until this day, indicating that her findings are rather influential. However, since Tong wrote her study the econometric modelling of a gravity model has improved, leaving her results outdated.

Firstly, Tong did not properly control for multilateral resistance terms in her model. Multilateral resistance terms were originally derived in a theory-based gravity model for trade by

Anderson and van Wincoop (2003). It measures how trade impediments affect bilateral trade between two countries, relative to the effect of the average impediments the countries face with the rest of the world. When for example tariff rates between two countries increase, it becomes more expensive for these countries to trade with each other, thus trading with the rest of the world becomes relatively cheaper. Brouwer et al. (2008) note that the same logic applies for FDI estimations with a gravity model, where multilateral resistance can be interpreted as the relative attractiveness of alternative investment locations.

Tong (2005) used remoteness terms to proxy multilateral resistance as multilateral resistance tends to be low when a country is more remotely located. The remoteness variable as constructed by Tong is however not theoretically correct, as it only considers distance as a trade barrier whereas more factors ought to be considered (World Trade Organization, 2012). This is also acknowledged by Head and Mayer (2014), who are stating that remoteness terms are too weak to proxy multilateral resistance. Not dealing with multilateral resistance correctly can cause omitted variable bias, distorting the estimated elasticities. In a cross-section gravity model, the use of host and source country dummies deal with this issue (Feenstra, 2002; Baier and Bergstrand, 2007; Brouwer et al., 2008; Felbermayr et al., 2010).

A second issue stems from Tong's choice for using a Tobit estimation as the preferred estimation method. A Tobit estimation is a useful tool for estimating a gravity model when there are many zeroes observed in the dependent variable, as log-linearizing would give unusable values. However, a Tobit estimation fails in providing unbiased estimates when the assumptions of having normal and homoscedastic residuals are not met (Liu,2008).

Apart from the aforementioned econometric issues, Tong's (2005) study only provided estimates for the impact of ethnic Chinese networks, even though she mentions the existence of international associations formed by other ethnicities. According to Tong, ethnic Chinese are most famous for forming ethnic networks that actively engage in trade and international investments, but as Felbermayr et al. (2010) have shown in their study regarding ethnic networks in international trade, ethnic Chinese do not form the most influential network.

2.2) Derivation of the hypotheses

The points mentioned justify a reconstruction of Tong's (2005) study on ethnic Chinese networks. A reconstruction of her study would determine whether the outcome holds when alternative estimation techniques are used to correct several econometric issues. It furthermore gives opportunity to examine the existence of alternative ethnicities forming networks which have an impact on bilateral FDI. This study will address these prior topics by testing the following hypotheses against their alternative:

Hypothesis 1:

$$H_0:$$
$$Internationally\ dispersed\ ethnic\ Chinese\ networks\ have\ a\ positive\ impact\ on\ a\ country's\ inward\ foreign\ direct\ investment\ stock$$

Against:

$$H_1:$$
$$Internationally\ dispersed\ ethnic\ Chinese\ networks\ do\ not\ have\ a\ positive\ impact\ on\ a\ country's\ inward\ foreign\ direct\ investment\ stock$$

Hypothesis 2:

$$H_0:$$
$$There\ are\ ethnicities, excluding\ ethnic\ Chinese, that\ positively\ impact\ a\ country's\ inward\ foreign\ direct\ investment\ stock\ by\ forming\ international\ networks$$

Against:

$$H_1:$$
$$There\ are\ no\ ethnicities, excluding\ ethnic\ Chinese, that\ positively\ impact\ a\ country's\ inward\ foreign\ direct\ investment\ stock\ by\ forming\ international\ networks$$

To be able to answer these hypotheses, a slight alteration is required in how the network variable is constructed. Tong (2005) calculated the variable by taking the product of 2 countries' ethnic Chinese populations. Ethnic Chinese include anyone who has roots in China, counting offspring of former migrants as well. This study calculates the network variable by using the stock of foreign-born individuals in a given country. There is an extensive amount of data available regarding these migrant stocks, making research on varying ethnic backgrounds possible. By making this alteration, ethnic networks might be better defined being migrant networks. Felbermayr et al. (2010) noted however that they found similar estimates when comparing the two with each other.

Summarizing, this study will revisit and extend Tong's (2005) study regarding the impact of Chinese ethnic networks on bilateral FDI. It will deviate in the following three ways:

1) Country-fixed effects are added to the gravity specification to control for multilateral resistance (and other country-specific related variables that influence bilateral FDI).
2) Instead of using a Tobit model, this study estimates the gravity model using a Heckman sample selection model and a Poisson quasi-maximum likelihood (PPML) estimation. The justification for the use of these models will be elaborated in chapter 3.3.2.
3) Instead of using an individual's ethnic background to calculate the proxy for the network variable this study will use its migrant status, which makes it possible to distinguish various ethnic networks apart from the Chinese network.

3) Data and Methodology

3.1) Dependent and independent variables

As Tong's (2005) dataset is unfortunately inaccessible, it is not possible to replicate her estimations and extend it with today's econometric methods. To replicate her study as close as possible this study will make use of mostly the same control variables as Tong added to her empirical model.

Variables mentioned in this chapter are either associated with country i (the country hosting investments), country j (the source of the investments), or both at the same time (ij). Table 3.1 gives an overview of the variables used when estimating the impact of the Chinese network on FDI, together with their data sources. The base year of the data is 2010; Exceptions are noted in chapter 3.2.

The dependent variable in this study is the nominal FDI stock in dollars in country i originating from country j. Stocks are used instead of flows since stocks are not as heavily affected by short-term fluctuations (Tong, 2005). Since this study uses a gravity model, staple gravity variables are added. Standard gravity variables include the nominal gross domestic product (GDP) in dollars of both the host and source countries to measure their economic size, and the bilateral distance between country i and j's capital cities in kilometres.

Apart from these standard gravity variables, the population of both countries has been added to the empirical model. An important reason for adding each country's population is due to the way the proxy for ethnic network strength is calculated. To proxy e.g. Chinese network strength between two countries, a variable is constructed by taking the product of their Chinese migrant populations. Tong (2005) argues that without adding the total population as a variable in the empirical model, an endogeneity bias might occur, distorting the estimates. This happens because the number of migrants in a country is likely to be correlated with its total population, thus causing the proxy variable to capture part the population's effect on the dependent variable.

As mentioned, the Chinese network strength between two countries is measured by taking the product of the ethnic Chinese populations in the two countries. This variable measures all possible connections between the countries' Chinese migrant populations. This will be the main variable of interest in this study. Other explanatory variables added to the empirical model include:

Table 3.1: *Variable names, description and source*

Variable	Description and formula	Data source
F_{ij}	Inward nominal 2010 FDI stock in country i originating from country j	UNCTAD
Y_i	Nominal 2010 GDP in country i	CEPII
Y_j	Nominal 2010 GDP in country j	CEPII
D_{ij}	Distance between i and j's capital cities in kilometers	CEPII
Pop_i	Total population in country i	CEPII
Pop_j	Total population in country j	CEPII
Rem_i	Relative remoteness of country i; following Tong (2005), calculated as: $Rem_i = \sum_i \frac{D_{ij}}{Y_i/Y_w}$ where Y_w stands for the sum of all sample countries GDP	CEPII / Own calculations
Rem_j	Relative remoteness of country j; following Tong (2005), calculated as: $Rem_j = \sum_j \frac{D_{ij}}{Y_j/Y_w}$ where Y_w stands for the sum of all sample countries GDP	CEPII / Own calculations
Tar_i	Average import tariff rate from country i during 2005-2010; calculated as: $Tar_i = \frac{\sum_{t=2005}^{2010} Average\ import\ tariff\ rate_{it}}{6}$	World Bank / Own calculations
$Chin_{ij}$	Proxy variable measuring Chinese ethnic network strength, calculated as: $Chin_{ij} = Chinese\ migrant\ population_i * Chinese\ migrant\ population_j$	UNDESA
Adj_{ij}	A dummy that equals 1 if country i and j are adjacent	CEPII
FTA_{ij}	A dummy that equals 1 if country i and j are in a regional free trade agreement (FTA)	CEPII
$Lang_{ij}$	A dummy that equals 1 if country i and j have a common official first language	CEPII
$Colony_{ij}$	A dummy that equals 1 if country i and j share former colonial ties	CEPII
$GroGDP_i$	Country i's annual GDP growth rate during 2005-2010; calculated as: $GDP\ growth\ rate_{it} = \frac{GDP_{it}}{GDP_{it-1}} \rightarrow GroGDP_i = \frac{\sum_{t=2006}^{2010} GDP\ growth\ rate_{it}}{5}$	World Bank / Own calculations
δ_i	Host country fixed effects	-
δ_j	Source country fixed effects	-
$time_{ij}$	A dummy variable that, as explained in chapter 3.2, has a value of 1 when the FDI stock data is from 2011 instead of 2010	-
ε_{ij}	Error term representing the unspecified variance in F_{ij}.	-

Data uses 2010 as its base year; Exceptions noted in chapter 3.2; Summary statistics provided in appendix table A1

- A remoteness variable that measures whether a country is relatively remotely located from the rest of the world. This is an alternative method to proxy multilateral resistance (Gómez-Herrera, 2012). When a country is relatively remotely located trading is rather expensive due to the distance, thus making (horizontal) FDI more lucrative;
- A variable measuring the host country's average tariff rates as it can give an indication whether tariff jumping is a reason for firms to engage in FDI;
- A variable measuring the host country's average GDP growth rate as it gives the effect of a country's market growth on its incoming FDI;
- A dummy variable that measures the effect of two countries being adjacent to each other. Countries that share a border often have a history of economic relations with each other since distance was much more of an obstacle in the past than it is today;
- A dummy variable that measures the effect of residing in a trading bloc as residing in a trading bloc can affect FDI when interfirm trading is relatively cheaper, which increases the profitability of vertical FDI. It can furthermore decrease the relative usefulness of horizontal FDI as exporting is cheaper, making the fixed costs of setting up a firm relatively higher;
- A dummy variable that measures whether the countries have a common official language, as having a common language makes it easier for people to communicate, which thus makes it easier to engage in economic transactions;
- A dummy variable that measures the effect of former colonial ties as colonial ties indicate a shared history often with cultural similarities like the social and political environment;

3.2) Data and sources

All the data mentioned in this paragraph uses 2010 as its base year. Exceptions will be noted together with the year the data is collected from.

The study regarding the Chinese network uses data from 58 countries and are chosen so to resemble Tong's (2005) sample. The countries are presented in Table 3.2. The countries used in this sample differ from Tong in a few ways. First, Taiwan was removed from Tong's original country set as the United Nations Department of Economic and Social Affairs (UNDESA), the institution providing migrant data, does not provide data for its migrant stock. Second, Namibia was dropped as there was no data regarding its bilateral FDI stock. Third, Czechoslovakia is not included in this study as it is no longer a country. Since the Dissolution of Czechoslovakia (1993) it has parted in two separate countries: the Czech Republic and Slovakia. Data from the Czech Republic is included to replace Czechoslovakia. Finally, the UNDESA does not indicate whether empty values in their data are caused by the fact that there are no migrants from a certain nationality present in a respective country, or whether they are empty because they have no information regarding these nationalities' migrant stock. For this reason,

Table 3.2: *Countries, their Chinese migrant stock and total population circa 2010*

Country	Chinese migrant stock (numbers)	Total population (millions)	Country	Chinese migrant stock (numbers)	Total population (millions)
Americas			***East and Southeast Asia***		
Canada	636,401	34.0	Japan	687,156	127.5
United States of America	1,827,867	309.3	Indonesia	65,307	240.7
Argentina	12,476	40.4	Hong Kong	2,260,045	7.0
Brazil	19,229	195.2	Republic of Korea	490,028	49.4
Chile	5,743	17.2	Malaysia	9,923	28.3
Colombia	1,901	46.4	Philippines	35,398	93.4
Ecuador	2,013	15.0	Thailand	68,811	66.4
Mexico	7,272	117.9	Viet Nam	2,440	87.0
Peru	3,780	29.3	Singapore	365,797	5.1
Venezuela	12,713	29.0			
Bolivia	968	10.2			
Paraguay	800	6.5			
Dominican Republic	1,406	10.0	***European Union***		
El Salvador	248	6.2	Austria	15,145	8.4
			Belgium	8,627	10.9
Other			Bulgaria	824	7.4
Australia	371,590	22.0	Czech Republic	4,601	10.5
Bangladesh	168,119	151.1	Denmark	10,878	5.5
Egypt	722	78.1	Finland	7,362	5.4
Fiji	898	0.9	France	101,734	65.0
Gabon	204	1.6	Germany	98,954	81.8
India	7,372	1205.6	Hungary	11,091	10.0
Libya	2,727	6.0	Ireland	12,416	4.6
Mauritius	2,564	1.3	Italy	200,400	59.3
New Zealand	84,329	4.7	Netherlands	52,904	16.6
Pakistan	338	173.1	Norway	8,852	4.9
Papua New Guinea	87	6.9	Poland	1,221	38.2
South Africa	18,522	50.9	Portugal	9,227	10.6
Sri Lanka	2,312	20.7	Romania	2,109	20.2
Switzerland	15,662	7.8	Spain	154,918	46.6
Togo	459	6.3	Sweden	25,107	9.4
Turkey	1,727	72.1	United Kingdom	162,564	62.8

Cameroon, the Democratic Republic of the Congo (Zaire), Jamaica, Kenya, Liberia, Nigeria, Morocco, Tanzania, Uruguay, and Zimbabwe have been removed from the Tong's original sample. China has also been removed, as it is not possible to be a Chinese migrant when still residing in China.

The last few countries mentioned have only been removed during the estimation of the impact of Chinese migrant networks on FDI. During the estimation of other migrant nationalities that form networks, these countries have been added to the sample again, conditional on the fact that there was data available on the migrant stock regarding the network being estimated. Just as with the estimation of the Chinese migrant network, countries are removed from their sample whenever it lacks migrant data regarding the network being estimated.

Data on the bilateral inward FDI stock was retrieved from the website of the United Nations Conference on Trade And Development (UNCTAD). For the estimation of the Chinese network effect, 45 countries' data provided by the UNCTAD was retrieved from these countries' respective central banks. 3 countries' data has been based on results from the Coordinated Direct Investment Survey of the International Monetary Fund. 10 countries' data are based on the source countries' outward FDI stock. Of the 3,306 possible country pairs, the data by the UNCTAD only provided useable bilateral FDI statistics of 1,697 of the combinations.

Out of the 1,697 combinations, 48 combinations entailed a negative inward FDI stock. According to the Organisation for Economic Cooperation and Development (OECD, 2008), negative FDI stock values can occur when loans from the affiliate to a parent company exceed the loans and equity given by a parent to its affiliate. Apart from that, the OECD states that negative values mainly occur in FDI statistics provided by its source country rather than its host country. Following Wellhausen (2014) the negative values are changed to zeroes as it is arguably the appropriate lower bound for measuring the level of investments in a country.

Out of the 1,697 observations, 1,641 combinations use the inward FDI stock for 2010. For 56 combinations the 2010 stock was not available, thus the inward FDI stock reported for 2011 was used. A dummy variable was added to control for any temporal differences between 2010 and 2011. 253 combinations report an inward FDI stock of 0, which equals roughly 15% of all observations. An overview of each observation's host and source countries' locations, together with an overview regarding the observed zeroes' in the dependent variable, is presented in table

3.3a and 3.3b. The countries that reside in each of the geographical regions mentioned in these tables are distinguished in table 3.2.

Table 3.3a: *Regional overview of countries, observations, and observed zeroes of the dependent variable in numbers*

Region	Countries	Observations		Observed zeroes in F_{ij}	
		Host i	Source j	Host i	Source j
Americas	14	395	328	52	81
East and Southeast Asia	9	214	307	31	43
European Union	19	731	739	110	59
Other	16	357	323	60	70
Total	58	1697		253	

Table 3.3b: *Regional overview of countries, observations, and observed zeroes of the dependent variable in percentages*

Region	Countries	Observations		Observed zeroes in F_{ij}	
		Host i	Source j	Host i	Source j
Americas	24.1	23.3	19.3	20.6	32.0
East and Southeast Asia	15.5	12.6	18.1	12.2	17.0
European Union	32.8	43.1	43.6	43.5	23.3
Other	27.6	21.0	19.0	23.7	27.7
Total			100%		

Table 3.3a and 3.3b show that the observations predominately consist of data regarding countries from the European Union. Zeroes in the dependent variable are however observed more often when the country sourcing the investments is in the Americas or dispersed over the unspecified regions. Chapter 3.3.3 discusses the possible implications for the estimations resulting from the distribution of the observations.

Data regarding bilateral migrant stocks were obtained from the population division of the UNDESA. The data classifies a person as a migrant if he/she was born in a different country as the one he/she is residing. No data is available regarding when the migration took place, so no distinction can be made between new migrants and people that have migrated 30 years ago.

The standard gravity variables have been retrieved from the CEPII database. The data retrieved is compiled by the studies of Head et al. (2010) and Head and Mayer (2014). It includes each countries' GDP, their population and the bilateral distances. The dummy variables regarding adjacency, a common official language, former colonial ties between the countries, and whether the countries take part in a regional free trade agreement are obtained from the dataset provided by the CEPII as well. The bilateral distance and the GDP provided in this dataset are used to calculate the variable that measures the relative remoteness of the sample countries.

Data on the weighted average tariff rates and GDP in constant 2010 dollars from 2005 till 2010 are retrieved from the database provided by the World Bank.

3.3) Methodology

3.3.1) The gravity model

The gravity model is an empirical specification that is often used in the literature to estimate determinants of bilateral trade and financial flows. Its base specification was introduced by Tinbergen (1962). The model implies that economic transactions between two economies (i and j) are proportional to the size of two economies and disproportional to the distance between them. These implications result in the following equation:

$$(3.1) \qquad F_{ij} = \beta_0 \frac{Y_i^{\beta_1} \, Y_j^{\beta_2}}{D_{ij}^{\beta_3}}$$

Equation 3.1 gives a basic gravity specification. F_{ij} denotes trade/financial flows from j to i, Y_i and Y_j respectively denote the size i and j's economy, measured in gross domestic product (GDP) or gross national income (GNI), and D_{ij} measures the geographical distance between i and j, and proxies roughly the transaction costs between i and j. β_0 is a constant. Whereas the size of both economies and the distance are staple in any gravity model, most empirical estimations add an array of covariates that potentially affect the dependent variable. The equation is found to be very versatile in predicting various bilateral flows, and theoretically it has been derived in models predicting bilateral trade (Anderson and van Wincoop, 2003), FDI (Bergstrand and Egger (2007) and migration (Beine et al., 2015).

Combining the explanatory variables discussed in chapter 3.1 leads to the following gravity equation:

$$(3.2) \qquad F_{ij} = \beta_0 \, Y_i^{\beta_1} \, Y_j^{\beta_2} \, D_{ij}^{\beta_3} \, Pop_i^{\beta_4} \, Pop_j^{\beta_5} \, Rem_i^{\beta_6} \, Rem_j^{\beta_7} \, Chin_{ij}^{\beta_8} \exp(\beta_9 \, Tar_i + \beta_{10} Adj_{ij} + \beta_{11} \, FTA_{ij} + \beta_{12} \, Lang_{ij} + \beta_{13} \, Colony_{ij} + \beta_{14} \, GroGDP_i \,)$$

Where every β denotes the coefficient related to its respective variable. Subscripts i and j denote respectively the host and the source country. Traditionally in gravity literature, equation 3.2 is estimated by log-linearizing and using an Ordinary Least Squares (OLS) estimation. Log linearizing equation 3.2, adding country fixed effects and adding a dummy for data taken from 2011 leads to the following equation:

$$(3.3) \qquad \ln F_{ij} = \ln \beta_0 + \beta_1 \ln Y_i + \beta_2 \ln Y_j + \beta_3 \ln D_{ij} + \beta_4 \ln \, Pop_i + \beta_5 \ln Pop_j + \beta_6 \ln Rem_i + \beta_7 \ln Rem_j + \beta_8 \ln Chin_{ij} + \beta_9 \ln Tar_j + \beta_{10} Ajd_{ij} + \beta_{11} \, FTA_{ij} + \beta_{12} \, Lang_{ij} + \beta_{13} \, Colony_{ij} + \beta_{14} \, GroGDP_i + \beta_{15} \, time_{ij} + \delta_i + \delta_i + \varepsilon_{ij}$$

3.3.2) Issues regarding the traditional empirical estimation of a gravity model

There are a few issues regarding the estimation of a traditional gravity model. As mentioned before, multilateral resistance, which entails that the dependent variable is not just subject to bilateral barriers but also to barriers with all other countries, should be dealt with. Anderson and van Wincoop (2003) demonstrated the importance of dealing with multilateral resistance in a gravity model and stressed that omitting a method that deals with it causes a severe bias in the estimates. In a cross-section gravity equation, host and source country fixed effects can be applied to the specification that deal with this issue (Feenstra, 2002; Baier and Bergstrand, 2007; Brouwer et al., 2008; Felbermayr et al., 2010).

A second issue occurs when there are many zeroes observed in the dependent variable. This causes a problem when log-linearizing the equation. There is no natural logarithm of zero, thus every observed value of zero turns out to be unusable. FDI stock has a lower bound of zero, making it prone to this issue. Various studies deal with this problem by leaving the unusable observations completely out of the estimation, or transforming the data (Gómez-Herrera, 2012). Dropping values would not be a problem if the zeroes are randomly distributed over the sample set. This would be the case when zeroes are for example random pieces of missing data. If the zeroes are not randomly distributed however, and the zero stands for zero economic activity, removing them from the dataset causes a selection bias (Westerlund and Wilhelmsson, 2011). Liu (2008) on the other hand stresses that transforming data in an OLS estimation leads to inconsistent estimates of β since it causes the dataset to not reflect the underlying values.

A third issue, brought up by Santos Silva and Tenreyro (2006), can occur in the presence of heteroskedasticity as it causes the property of the error term to change when log-linearizing the gravity equation (Gómez-Herrera, 2012). This is a result of Jensen's inequality. Jensen's inequality states that the expected value of a logarithm of a random variable differs from the logarithm of its expected value. The expected value of a log does not only depend on the mean of its respective variable, but also on the higher-order moments in its distribution. With homoscedastic data the variance and expected value of the error are constant. In case of heteroskedasticity this is not the case, causing log-linearizing to alter the conditional mean of the dependent variable, which violates the zero conditional mean assumption of the error term (Felbermayr et al., 2010). This causes inconsistent results, as the variance of the estimated parameters is biased.

To conquer the three issues mentioned either a nonlinear estimation method or a two-part model is required. Many estimation methods have been tested to tackle the three mentioned issues related to estimating a gravity model. Gómez-Herrera (2012) compared 8 different estimation techniques used to estimate a gravity model for bilateral trade. She found that estimation techniques that don't treat zero trade flows properly perform far worse than methods that do. She concluded that a Heckman sample selection model, as proposed by Helpman et al. (2008), was the estimator with the most desirable properties.

The Heckman sample selection model separates the decision to trade/invest from the value of the bilateral trade/investments by estimating the equation in two steps. The first step of the estimation entails a probit estimation where the dependent variable is a dummy variable that indicates whether a given observation is present in the data sample. The second stage estimates the values of the dependent variable using OLS, conditional on the fact that the dependent variable was non-zero in the first stage. During the second stage of the estimation, the inverse Mills ratio is added as an explanatory variable to correct the selection bias caused by log-linearizing the dependent variable.

The Heckman sample selection model requires a variable which affects a country's decision to engage in FDI. It should be correlated to the fact that two countries are bilaterally investing, but not with the invested amount. Following Gómez-Herrera (2012), this study will use the dummy indicating whether the countries share a common official language as the required selection variable.

Although the Heckman sample selection model has the most desirable properties according to Gómez-Herrera's (2012) study, it still approaches the gravity equation in a log-linear form, which again could lead to questionable results when the assumptions of homoscedasticity and normality of the error term are not met (Tran et al., 2013). Santos Silva and Tenreyro (2006) opt for not taking the log of the dependent variable and allowing heteroskedasticity. Their preferred estimation method is a Poisson quasi-maximum likelihood (PPML) estimation as it is consistent in the presence of heteroskedasticity, and likewise deals naturally with the presence of zeroes of the dependent variable. This estimation technique was also found to provide robust results by Felbermayr et al. (2010) in their study on the effect of migrant networks in international trade.

Both a PPML estimation and a Heckman sample selection model are performed in this study to estimate the effect of Chinese and other migrant networks on bilateral FDI stocks, and the models are tested whether they are correctly specified. Since PPML is not based on the likelihood function, the Akaike information criterion and the Bayesian information criteria cannot be used to calculate which of models provides the best fit. To test the adequacy of each model against the other, an HPC test, as proposed by Santos Silva et al. (2015), is performed. This test is optimized to tell whether single- and double-index models that are used for non-negative data and have many zeroes in the dependent lead to similar fits.

3.3.3) Additional econometric issues

Studies on the effects of migration often encounter endogeneity issues. This is because migration by itself is subject to many determinants. Endogeneity entails that an independent variable is correlated with the error term, which can lead to a bias in the estimated parameters. The three main sources of an endogeneity bias are reverse causality, omitted variable bias and measurement errors.

Reverse causality happens when the dependent variable and an independent variable are a determinant of each other. In trade and FDI literature, migration is often subject to reverse causality. This is especially the case when examining the effect of migration on trade or FDI between the migrants' country of residence and the migrants' country of origin, since migrants might prefer to move towards countries where firms originating from their origin country are active. This study does not include the migrants' origin country in the estimations, thereby mitigating this issue (Felbermayr et al., 2010). Since this study does not contain data over time, it is not possible to test whether the independent variables are subject to reverse causality. Therefore, the assumption is made that there is no reverse causality distorting the estimates. Considering that most of the variables used in this study are found frequently in gravity literature, this assumption seems justified.

Omitted variable bias happens when a variable has explanatory power over another variable that is not added to the empirical model, but that does affect the dependent variable. Due to its absence, the coefficient of the variable it is correlated with will entail some of its explanatory power, causing its estimated coefficient to be biased. A Ramsey Regression Equation Specification Error Test (RESET) is performed to check whether there are omitted variables. The RESET test is performed by adding the squared fitted values as a variable to the regression after its initial estimation, and testing whether this variable provides a coefficient that significantly differs from zero. If the variable turns out insignificant, it indicates that the estimated model is correctly specified.

A measurement error occurs when the data is not correctly measured, causing it to not reflect its actual underlying values. The data used in this study comes from external sources, and therefore, the assumption is made that these external sources measured their provided data correctly. To make sure the variable of interest, $\ln Chin_{ij}$, is not severely biased resulting from an endogeneity issue, an instrumental variable (IV) approach is performed as a robustness check in chapter 4.1.2.2.

Another potential bias might result from the distribution of the observations in the sample. As table 3.3a and 3.3b have shown, observations predominately originated from European countries. Observed zeroes in the dependent variable on the other hand were relatively scarce whenever FDI was originating from a European country. These statistics could indicate that data availability has caused the sample selection to be non-random, which can result in a selection bias. To see whether this is the case, chapter 4.1.2.3 displays results after dividing the sample in two separate groups: one where the set of observations were taken when FDI was originating from developed countries, and one where the set of observations were taken when FDI was originating from developing countries.

4) Results

4.1) Chinese migrant networks

4.1.1) Main estimation results

This section will provide results of this study's findings on the impact of a Chinese network on FDI. Table 4.1, 4.2 and 4.3 provide the OLS, the Heckman sample selection and PPML estimations of equation 3.3. Table 4.1 provides coefficients of the estimated model when only the standard gravity variables and the network variable are added as explanatory variables. Table 4.2 gives the coefficients of the gravity model when all the explanatory variables, except country fixed effects, are added. Table 4.3 gives the estimations with country fixed effects. Table 4.3 deals effectively with multilateral resistance, thus contains the most reliable results.

The results for the Heckman sample selection model are split into 2 parts. The first part is a sample selection equation and provides results of a probit estimation on the full sample. The second part provides results for the outcome equation, which is an OLS regression for the countries that are bilaterally investing with the inverse mills ratio included as an explanatory variable to control for a selection bias caused by log-linearizing the dependent variable. In the discussion below, the results under the second part of the Heckman estimation will be compared to the results obtained from the OLS and PPML estimations. For both the OLS and the Heckman estimation the dependent variable is the log of the inward FDI stock in dollars. In the OLS estimation, zeroes in the dependent variable have been omitted. For the PPML estimation, the dependent variable is the inward FDI stock in millions of dollars. In the OLS and PPML estimations, the observations were clustered by host country to account for intraclass correlation. A Breusch-Pagan/Cook-Weisberg test was furthermore performed after the OLS estimations, which confirmed the presence of heteroskedasticity. This indicates that the linear estimation methods are likely to give inconsistent results.

Apart from the addition of host and source country fixed effects, table 4.3 gives results of the HPC test and the RESET test for both the Heckman sample selection model and the PPML estimation. Looking at the HPC test statistics, both the PPML and the Heckman score above the critical level of 0.05 with p-values of respectively 1.000 and .214. This indicates that both the PPML estimation and the Heckman two-part model lead to similar fits in the region close to zero. Therefore, the results of the HPC test provide no evidence against the use of either the Heckman two-part model or the PPML estimation.

When looking at the RESET test statistics, the PPML estimation is not rejected at 0.05 as its p-value is .283. This indicates that the conditional expectation in areas further away from zero are met by the PPML estimation. The p-value associated with the Heckman model's RESET test is 0.04. The test thus rejects the Heckman two-part model at 0.05 when it comes to explaining values further away from zero. The PPML estimation thus seems to be correctly specified and therefore perform best given this study's empirical model, hence its provided estimates are considered to be the most reliable.

Table 4.1: *Gravity model estimations without country fixed effects*

	OLS	Heckman Sample Selection Model		PPML
		1st part	2nd part	
$\ln Y_i$	.418***	.075**	.375***	.554***
	(4.44)	(2.42)	(6.98)	(5.27)
$\ln Y_j$	.805***	.249***	.618***	.634***
	(11.55)	(7.09)	(7.62)	(9.36)
$\ln D_{ij}$	-1.219***	-.348***	-.959***	-.759***
	(-13.56)	(-6.84)	(-9.10)	(-11.00)
$\ln Pop_i$				
$\ln Pop_j$				
$\ln Rem_i$				
$\ln Rem_j$				
Tar_i				
$\ln Chin_{ij}$	.307***	.064***	.253***	.166***
	(6.00)	(3.44)	(7.62)	(3.43)
Adj_{ij}				
FTA_{ij}				
$Lang_{ij}$		.609***		
		(4.32)		
$Colony_{ij}$				
$GroGDP_i$				
$Time_{ij}$	-.354	-.237	-.174	-1.582***
	(-1.13)	(-1.17)	(-0.43)	(-5.27)
$Constant$	-9.012***	-5.898***	-3.264	-21.187***
	(-2.94)	(-4.25)	-1.16	-5.66
Country fixed Effects	No	No	No	No
Obs	1444	1697	1444	1697
Mills λ			-2.667***	
			(-3.42)	
R^2	0.532	.106		.414

The dependent variable for the OLS and Heckman estimations is the log of the host country's 2010 inward FDI stock; The dependent variable for the PPML estimation is the host country's 2010 inward FDI stock in millions; ***, ** and * denote a significance level of 1%,5% and 10% respectively; Observations clustered by host countries; Robust t-statistics in parenthesis

Table 4.2: *Gravity model estimations with explanatory variables and without country fixed effects*

	OLS	Heckman Sample Selection Model		PPML
		1st part	2nd part	
$\ln Y_i$	.946***	.039	.920***	.943***
	(6.09)	(0.64)	(12.83)	(4.52)
$\ln Y_j$	1.538***	.481***	1.338***	1.689***
	(19.21)	(9.82)	(14.64)	(9.75)
$\ln D_{ij}$	-.914***	-.269***	-.872***	-.548***
	(-8.00)	(-3.82)	(-10.58)	(-5.28)
$\ln Pop_i$	-.454***	.152**	-.521***	-.312*
	(-2.89)	(2.37)	(-6.72)	(-1.84)
$\ln Pop_j$	-.840**	-.300***	-.734***	-1.042***
	(-12.21)	(-7.59)	(-10.41)	(-7.61)
$\ln Rem_i$	-.874	-2.901***	2.086	1.584
	(-0.36)	(-2.52)	(1.53)	(0.43)
$\ln Rem_j$	-4.609**	-.227	-3.156**	-6.211***
	(-2.68)	(-0.19)	(-2.07)	(-3.13)
Tar_i	.137***	-.034*	.115***	.014
	(2.78)	(-1.82)	(4.53)	(0.17)
$\ln Chin_{ij}$	.199***	.044**	.204***	.105**
	(4.53)	(2.10)	(7.84)	(2.03)
Adj_{ij}	.086	.392	.350	-.106
	(0.47)	(1.28)	(1.56)	(-0.57)
FTA_{ij}	.432**	.086	.405**	.283
	(2.14)	(0.73)	(2.85)	(1.15)
$Lang_{ij}$	1.212***	.672***		.600***
	(8.27)	(4.37)		(4.14)
$Colony_{ij}$	.857***	.131	1.421***	.568***
	(4.30)	(0.47)	(6.39)	(3.28)
$GroGDP_i$	13.478**	-1.024	19.414***	17.861***
	(2.484)	(-0.37)	(6.23)	(3.30)
$Time_{ij}$	-.805**	-.059	-.686**	-1.417***
	(-2.15)	(-0.27)	(-2.20)	(-2.66)
$Constant$	47.141	55.694	-14.923	37.441
	(0.90)	(1.59)	(0.36)	(0.41)
Country fixed Effects	No	No	No	No
Obs	1431	1681	1431	1681
Mills λ		-.826		
		(-1.64)		
R^2	0.651	.296		.590

The dependent variable for the OLS and Heckman estimations is the log of the host country's 2010 inward FDI stock; The dependent variable for the PPML estimation is the host country's 2010 inward FDI stock in millions; ***, ** and * denote a significance level of 1%,5% and 10% respectively; Observations clustered by host countries; Robust t-statistics in parenthesis

Table 4.3: *Gravity model estimations with country fixed effects*

	OLS	Heckman Sample Selection Model		PPML
		1st part	2nd part	
$\ln Y_i$				
$\ln Y_j$				
$\ln D_{ij}$	-1.158***	-.529***	-1.258***	-.575***
	(-11.66)	(-4.27)	(16.05)	(-4.69)
$\ln Pop_i$				
$\ln Pop_j$				
$\ln Rem_i$				
$\ln Rem_j$				
$\ln Tar_i$				
$\ln Chin_{ij}$	.217***	3.680***	.747***	.233***
	(7.87)	(49.50)	(3.09)	(9.96)
Ajd_{ij}	.018	.4312	.108	.105
	(0.09)	(1.54)	(0.56)	(0.73)
FTA_{ij}	.335**	-.097	.283**	.257
	(2.23)	(-0.54)	(2.09)	(1.31)
$Lang_{ij}$	.729***	1.012***		.342**
	(4.36)	(4.13)		(2.19)
$Colony_{ij}$	.985***	.576	1.357***	.572***
	(4.53)	(1.54)	(7.44)	(3.66)
$GroGDP_i$				
$Time_{ij}$	-1.217**	-1.817***	-1.179*	-2.849***
	(--2.91)	(-3.37)	(-1.92)	(-5.58)
Country fixed Effects	Yes	Yes	Yes	Yes
Obs	1444	1697	1444	1694
Mills λ			.212	
			(0.92)	
HPC test		.214		1.000
Reset-test		.040		0.283
R^2	0.778	.374		.823

The dependent variable for the OLS and Heckman estimations is the log of the host country's 2010 inward FDI stock; The dependent variable for the PPML estimation is the host country's 2010 inward FDI stock in millions; ***, ** and * denote a significance level of 1%,5% and 10% respectively; Observations clustered by host countries; Robust t-statistics in parenthesis

Looking at the main variable of interest, $\ln Chin_{ij}$, we see that every column provides positive and significant estimates. The Heckman and PPML estimates in table 4.1, the OLS and Heckman estimates in table 4.2 and the OLS and PPML estimates in table 4.3 provide similar significant results as Tong (2005) found for this variable. The results range from .20 for the OLS estimation in table 4.2 to .25 for the Heckman estimation in table 4.1. Tong's OLS estimation resulted in a value of .21 and her Tobit estimation in a value of 0.19. The coefficient estimated by the PPML estimation in table 4.3, being the most reliable result, has a value of .23. This indicates that an increase of 1% of Chinese migrant population in two countries would cause for an increase in inward FDI in the host country of roughly 0.47%[1], which is an increase of 0.09 percentage points compared to Tong's estimation.

The OLS estimate of the network variable in table 4.1, the PPML estimate in table 4.2 and the Heckman estimate in table 4.3 are divergent from the other estimations. The OLS's estimate in table 4.1 gives a value of .31, which is slightly higher compared to the other estimates. On the other hand, the PPML estimation in table 4.2 gives a value of .11, which is slightly lower compared to the other estimates. The Heckman estimation in table 4.3 suggests a value of approximately .75, which is considerably higher compared to all other estimations.

The estimations for most of the control variables are comparable in sign and magnitude over the estimations. Looking at the standard gravity variables in table 4.1 and 4.2, the results suggest that an increase of both the source and host country's GDP would lead to an increase in FDI from source to host. Distance has a negative coefficient in all 3 tables, indicating that countries are less likely to invest in countries that are distant.

The results in table 4.2 furthermore suggest that the size of the population of both host and source country have a negative influence on the inward FDI stock. The remoteness of the host country is insignificant in all the estimations. Remoteness of the source country gives negative estimates and is significant at a 1 percent level in the OLS estimation and PPML estimation, and at a 5 percent level in the Heckman estimation. These results indicate that less investments are originating from countries which are relatively remotely located. The coefficients related to the average tariff rate of the host country are positive and significant in the OLS and Heckman estimation, but insignificant in the PPML estimation. Having a high average GDP growth appears to have a significant influence on inward FDI, as it is significant in all three estimations.

[1] Calculated by: $((1 + 0.00233)^2 - 1) * 100\% = 0.4665\%$

This indicates that companies tend to invest in countries where economic growth is relatively high.

The estimates FTA dummy in table 4.2 and 4.3 are positive and significant in the OLS and Heckman estimations, but insignificant in the PPML estimations. Adjacency seems to be unrelated to whether countries are investing in one another as its estimated coefficients are insignificant in nearly all estimations. Having a common official language and the existence of a colonial relation in the past are both positive and significant in all estimations and can thus be considered as determinants for bilateral FDI.

4.1.2) Robustness checks

4.1.2.1) Alternative variable

The following section tests whether the results are sensitive to alterations. For the first robustness check, an alternative variable is constructed to estimate the Chinese network effect. Table 4.4 contains results after replacing $\ln Chin_{ij}$ with this alternative variable, which is similar to Felbermayr et al. (2010) and Rauch and Trindade's (2002) method of calculating the network effect. They calculated the network variable by taking the product of the shares of ethnic/migrant Chinese population within the total population of i and j. To be more concise, their approach gives the chance of picking two Chinese migrants when picking two random individuals from country i and j. The variable replacing $\ln Chin_{ij}$ in table 4.4 is calculated as:

$$(4.1) \qquad \ln alt\_Chin_{ij} = \ln\left(\frac{Chinese\ migrant\ population_i}{total\ population_i} * \frac{Chinese\ migrant\ population_j}{total\ population_j}\right)$$

Although this variable is related to the variable used during the previous estimations, its estimated coefficients ought to be interpreted slightly different. The results for an OLS, Heckman sample selection and PPML estimation with country fixed effects are listed in table 4.4.

The estimates for all independent variables hold when compared to table 4.3. The coefficient for the alternative network variable is positive and significant in each estimated coefficient, indicating that whenever the share of Chinese migrants within the total population of two countries increases, more bilateral FDI takes place. As such, the estimates for the alternative variable confirm the results obtained from the previous estimations and suggest the existence of a Chinese ethnic network affecting bilateral FDI.

Table 4.4: *Gravity model estimations with country fixed effects and the alternative network variable*

	OLS	Heckman Sample Selection Model		PPML
		1st part	2nd part	
$\ln Y_i$				
$\ln Y_j$				
$\ln D_{ij}$	-1.158***	-.529***	-1.258***	-.575***
	(-11.66)	(-4.27)	(-16.05)	(-4.69)
$\ln Pop_i$				
$\ln Pop_j$				
$\ln Rem_i$				
$\ln Rem_j$				
$\ln Tar_i$				
$\ln alt\_Chin_{ij}$	.163***	2.567***	.508***	.175***
	(7.87)	(37.20)	(3.09)	(9.96)
Ajd_{ij}	.018	.431	.108	.105
	(0.09)	(1.12)	(0.56)	(0.73)
FTA_{ij}	.335**	-.097	.283**	.257
	(2.23)	(-0.52)	(2.09)	(1.31)
$Lang_{ij}$	.729***	1.012***		.342**
	(4.36)	(4.13)		(2.19)
$Colony_{ij}$	. 985***	.576	1.357***	.572***
	(4.53)	(1.54)	(7.44)	(3.66)
$GroGDP_i$				
$Time_{ij}$	-1.217**	-1.817***	-1.179*	-2.849***
	(-2.91)	(-3.37)	(-1.92)	(-5.58)
Country fixed Effects	Yes	Yes	Yes	Yes
Obs	1444	1697	1444	
Mills λ			.212	
			(0.92)	
HPC test			.214	1.000
Reset-test			.040	0.282
R^2	0.778		.374	.823

The dependent variable for the OLS and Heckman estimations is the log of the host country's 2010 inward FDI stock; The dependent variable for the PPML estimation is the host country's 2010 inward FDI stock in millions; ***, ** and * denote a significance level of 1%,5% and 10% respectively; Observations clustered by host countries; Robust t-statistics in parenthesis

4.1.2.2) Instrumental variable approach

To test whether the estimates of $\ln Chin_{ij}$ are biased resulting from potential endogeneity issues, a two-staged least squares (2SLS) IV estimation is performed. To do so, two variables have been constructed that are correlated with $\ln Chin_{ij}$, but not correlated with the dependent variable F_{ij} through any unspecified mechanism residing in the error term. The first instrumental variable is constructed as:

$$(4.2) \quad IV\ 1_{ij} = \ln\left(Distance\ to\ China_i *\ Distance\ to\ China_j\right)$$

$IV\ 1_{ij}$ is thus the natural logarithm of the product of the host and source countries' distances to China. People tend to migrate predominately to countries which are relatively in close proximity to their country of origin rather than towards countries that are distant (Beine et al., 2015). Therefore, the distance to China should be correlated to the migrant stocks in the respective countries. Following this logic, the product of these distances should be correlated to the product of the respective countries' migrant stocks. Data regarding each countries' distance to China was obtained from the CEPII gravity dataset.

The second instrumental variable is constructed as:

$$(4.3) \quad IV\ 2_{ij} = \ln\left(Chinese\ migrant\ population\ in\ 1990_i * \\ Chinese\ migrant\ population\ in\ 1990_j\right)$$

$IV\ 2_{ij}$ is constructed in the same fashion as $\ln Chin_{ij}$, the difference being that $IV\ 2_{ij}$ uses the Chinese migrant stocks of 1990 rather than the 2010 stocks. $IV\ 2_{ij}$ is logically correlated with $\ln Chin_{ij}$. First, it is likely that some migrants that were residing in a respective country in 1990 were still residing in this country in 2010. Apart from that, people tend to migrate faster towards countries where there are relatively more people with the same ethnicity as they have (Beine et al., 2015). Data regarding the Chinese migrant population in 1990 was taken from the UNDESA. As there was no data available regarding Chinese migrants residing in Romania in 1990, the observations with Romania as either a host or a source country of FDI were not included in the 2SLS estimation.

Whether these variables are completely uncorrelated with the error term remains a point of discussion. Arguably, if two countries are both relatively close to China, there could be a certain cultural proximity between both countries since these countries are possibly also in proximity to each other. This would cause both $IV\ 1_{ij}$ and $IV\ 2_{ij}$ to be correlated to some extent with the error term since both distance and migration are related to the unspecified cultural proximity. This is however untestable, and thus the assumption is made that a possible correlation with the error term is neglectable.

Table 4.5 contains the outcomes of the 2SLS estimation, together with test statistics regarding the validity of the instruments used. First stage results can be found in appendix table A2. The estimation does not include host and source country fixed effects as the dummies would have perfect explanatory power over $\ln Chin_{ij}$ in the first stage of the estimation, resulting from the way $\ln Chin_{ij}$ is constructed. The log-linearization of the dependent variable might lead to a selection bias in these estimates. Heterogeneity furthermore remains an issue in the results provided.

The first stage F-statistic shows that the instruments used are strongly correlated with $\ln Chin_{ij}$, therefore having explanatory power for this variable. The Sargan-Hansen's test furthermore indicates that the instruments are excludable in the second stage, which entails that they are independent with of the model's residuals (Söderbom, 2011).

Table 4.5: *Gravity model estimations of a two-staged least squares estimation.*

	2SLS
$\ln Y_i$	.915***
	(7.30)
$\ln Y_j$	1.463***
	(12.40)
$\ln D_{ij}$	-.928
	(-12.69)
$\ln Pop_i$	-.457***
	(-6.25)
$\ln Pop_j$	-.800***
	(-15.78)
$\ln Rem_i$	.587
	(0.45)
$\ln Rem_j$	-5.326***
	(-3.46)
$\ln Tar_i$	.108***
	(4.34)
$\ln Chin_{ij}$	.241***
	(7.30)
Ajd_{ij}	.037
	(0.17)
FTA_{ij}	.490***
	(3.59)
$Lang_{ij}$	1.104***
	(7.72)
$Colony_{ij}$	.925***
	(4.13)
$GroGDP_i$	17.108***
	(5.67)
$Time_{ij}$	-.659**
	(-2.12)
$Constant$	58.150
	(1.36)
Country fixed Effects	No
Obs	1389
First-stage F-statistic	831.298***
Sargan–Hansen test	2.158
R^2	0.652

The dependent variable is the log of the host country's 2010 inward FDI stock; ***, ** and * denote a significance level of 1%,5% and 10% respectively; T-statistics in parenthesis

The estimates of $\ln Chin_{ij}$ are marginally higher than the estimates of the regular OLS estimation in table 4.2. Its estimated coefficient of .241 is however more in range of the fixed effects OLS and PPML estimations in table 4.3, insinuating that the results hold, and suggesting that Chinese networks have a significant impact on FDI. The other explanatory variables do not alter significantly, indicating the robustness of the model.

4.1.2.3) Investments originating from developed and undeveloped countries

In this section estimates are provided after dividing the observations in two groups. The first group contains the observations for which a developed country was the source of the FDI. The second group contains the observations for which a developing country was the source of the FDI. A country's development classification was obtained from the World Economic Situation and Prospects report (2010) provided by the United Nations. Appendix table A3 gives an overview of the countries that were considered developed and developing. Appendix table A4a and A4b give an overview of the dispersion of observations over developed and undeveloped countries. The results of the OLS and PPML estimations are provided in table 4.6.

Table 4.6: *Gravity model estimations with country fixed effects where the country sourcing FDI is either developed or developing*

	OLS j = developed	PPML j = developed	OLS j = developing	PPML j = developing
$\ln D_{ij}$	-1.150***	-.444***	-.821***	-.951***
	(-7.78)	(-3.36)	(-3.64)	(-5.22)
$\ln Chin_{ij}$	.200***	.249***	.289***	.088*
	(5.70)	(14.69)	(3.25)	(1.74)
Ajd_{ij}	-.066	.230	.763	-.351
	(-0.34)	(0.124)	(1.59)	(-1.19)
FTA_{ij}	.293	.412*	.322	.056
	(1.31)	(1.92)	(1.15)	(0.19)
$Lang_{ij}$	.355	.342**	1.15***	.907***
	(1.51)	(2.33)	(4.05)	(3.80)
$Colony_{ij}$	1.152***	.548***	.988**	-.519
	(3.87)	(3.42)	(2.30)	(-1.14)
$Time_{ij}$	-1.965***	-3.005***	-1.31	-2.315***
	(-3.47)	(-11.66)	(-0.71)	(-2.65)
Country fixed Effects	Yes	Yes	Yes	Yes
Obs	932	1011	512	682
R^2	.809	.835	.696	.823

The dependent variable for the OLS estimations is the log of the host country's 2010 inward FDI stock; The dependent variable for the PPML estimations is the host country's 2010 inward FDI stock in millions; ***, ** and * denote a significance level of 1%,5% and 10% respectively; Observations clustered by host countries; Robust t-statistics in parenthesis

The estimates regarding investments originating from developed countries remain robust to the results obtained in chapter 4.1.1. The PPML parameter for $\ln Chin_{ij}$ regarding observations where FDI originates from a developing country however differs. This could be caused by zeroes in the dependent variable. In the OLS estimates, the zeroes are removed due to log-linearization of the dependent variable. The PPML estimation however deals efficiently with the zeroes and considering that the largest proportion of zeroes originate from developing countries, they seem to cause a marginally smaller parameter for $\ln Chin_{ij}$.

The results furthermore point towards a selection bias that distort the parameters estimates in the previous tables. As the estimates in table 4.6 show, there is a difference observed between the network effect when investments originate from developing countries and investments originating from developed countries. However, most of the observations in the sample set are investments originating from developed countries, thereby biasing the overall effect given in, among others, table 4.3. The estimates provided in 4.6 all remain statistically significant, therefore still indicating some impact resulting from networking activities by Chinese migrants.

4.2) Other migrant networks

Using the data provided by the UNDESA (2017), bilateral migrant variables have been constructed for multiple countries of origin. The UNDESA provides data for 232 countries of origin. Many countries however lack data on their migrant stock. As mentioned previously in this study, there is no clear indication on the specific reason why these values are missing. Therefore, missing observations were dropped. Since dropping observations decreases the accuracy of the estimated coefficients, a threshold of 700 PPML observations has been set, which left a total of 86 countries. These countries are listed in table A5 in the appendix.

For each country listed in table A5, two variables have been constructed which are similar to the variables used to proxy the Chinese migrant network effect. The two variables are constructed as:

$$\text{(4.4)} \qquad \ln(Mig_{ijk}) = \ln(Migrant\ population_{ik} * \ Migrant\ population_{jk})$$

$$\text{(4.5)} \qquad \ln(alt\_Mig_{ijk}) = \ln\left(\frac{Migrant\ population_{ik}}{total\ population_{i}} * \frac{Migrant\ population_{jk}}{total\ population_{j}}\right)$$

These variables proxy migrant networks for any country of origin k. Due to log-linearization, observations had to be dropped whenever a country was reported to have zero migrants from a specific country k. For each migrant origin k four estimations have been performed: an OLS estimation, a Heckman two-step estimation, a regular PPML estimation and a PPML estimation using the alternative migrant variable as constructed in equation 4.5. All estimations contained host and source country fixed effects to control for multilateral trade resistance, and the observations were clustered by host country to account for intraclass correlation.

Out of the 86 countries listed in table A5 in the appendix, four migrant origins (apart from the Chinese) provided significant and robust results[2], giving indication for the existence of an Estonian, French, Korean and Venezuelan network. The estimates regarding these four countries can be found in appendix tables A6 to A9. There were 25 origin countries that provided semi-robust[3] results. Their estimations can be found in appendix table A9 to A34. Figure 4.1 gives an overview of the estimated PPML coefficients for both the robust and semi-robust network variables. Since the coefficients give the relative effect of each ethnic network, a comparison between network strength is possible. Figure 4.2 gives an overview of the total migrant stock of these countries within their estimated samples.

The results indicate that Korean migrants relatively have the biggest positive impact on bilateral FDI. The estimated coefficient of $\ln(Mig_{ij\,Korea})$ is 1.404, which entails that an increase of 1% of Korean migrant population in both host and source country would cause an increase in inward FDI in the host country from the source country of roughly 2.85%. Following the Korean network, the Portuguese network and Dominican Republic network are the second and third most impactful on bilateral FDI with the variable $\ln(Mig_{ijk})$ respectively being estimated at 0.848 and 0.741.

A remarkable outcome is the estimated coefficient for the Angolan network variable. The PPML estimation of $\ln(Mig_{ij\,Angola})$ being -0.533 indicates that a 1% increase in Angolan migrants in host and source country would lead to a decrease in inward FDI in the host country from the source country of roughly 0.01%. Although this seems like a small decrease, in absolute numbers a decrease of this magnitude could have serious consequences. A negative value of the migrant variable is furthermore not in line with the theoretical implications regarding ethnic business and social networks. Even though the robustness of this coefficient is questionable, potential issues resulting in the negative network variable and suggestions for further research will be discussed in the following concluding section of this study.

[2] Estimates were considered robust if all four estimation techniques provided significant ($p<0.05$) results for $\ln(Mig_{ijk})$ / $\ln(alt\_Mig_{ijk})$ that indicate a similar effect

[3] Estimates were considered semi-robust when both PPML estimations were significant ($p<0.05$) plus either the OLS or Heckman sample selection estimations resulted in significant estimates that were similar to the PPML estimate of the regular migrant variable.

Figure 4.1 *PPML estimates of each migrant network variable for countries that satisfied the set conditions, together with their 1.96 robust standard error plotted around them*

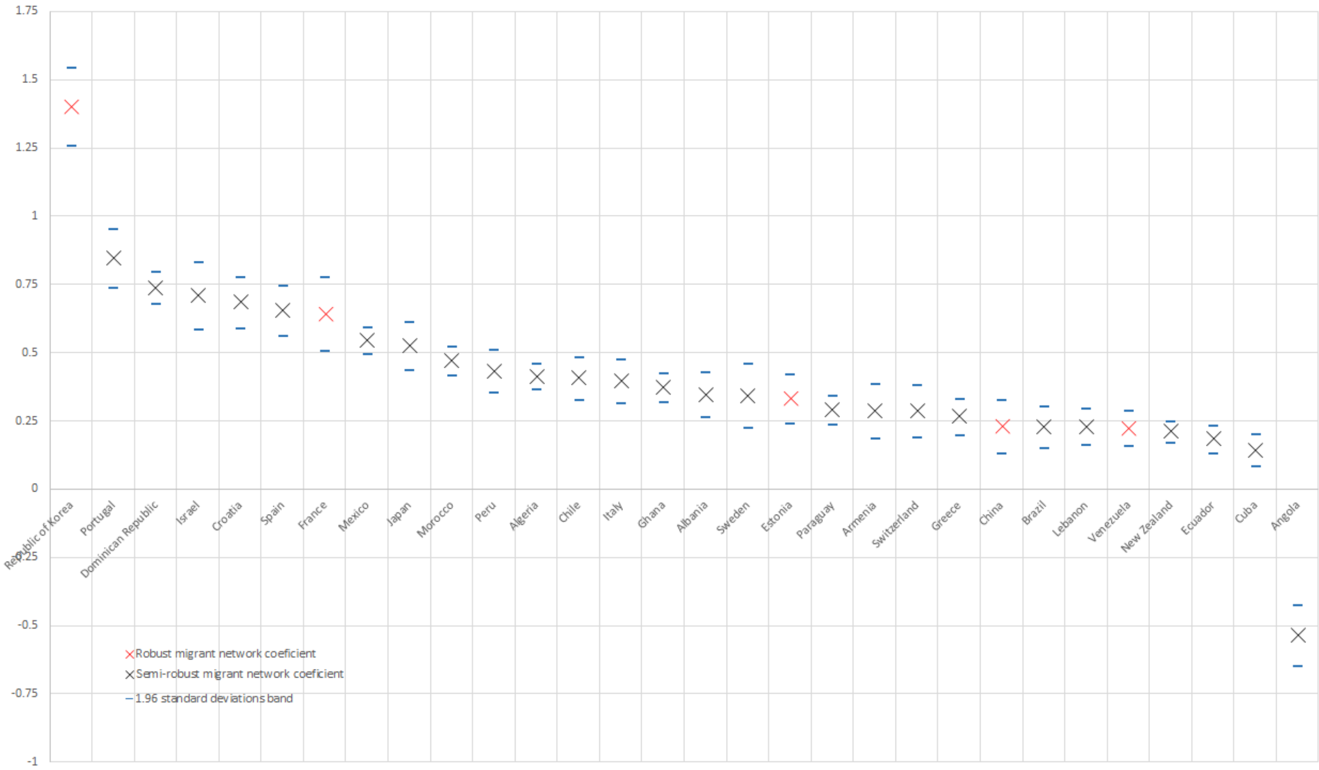
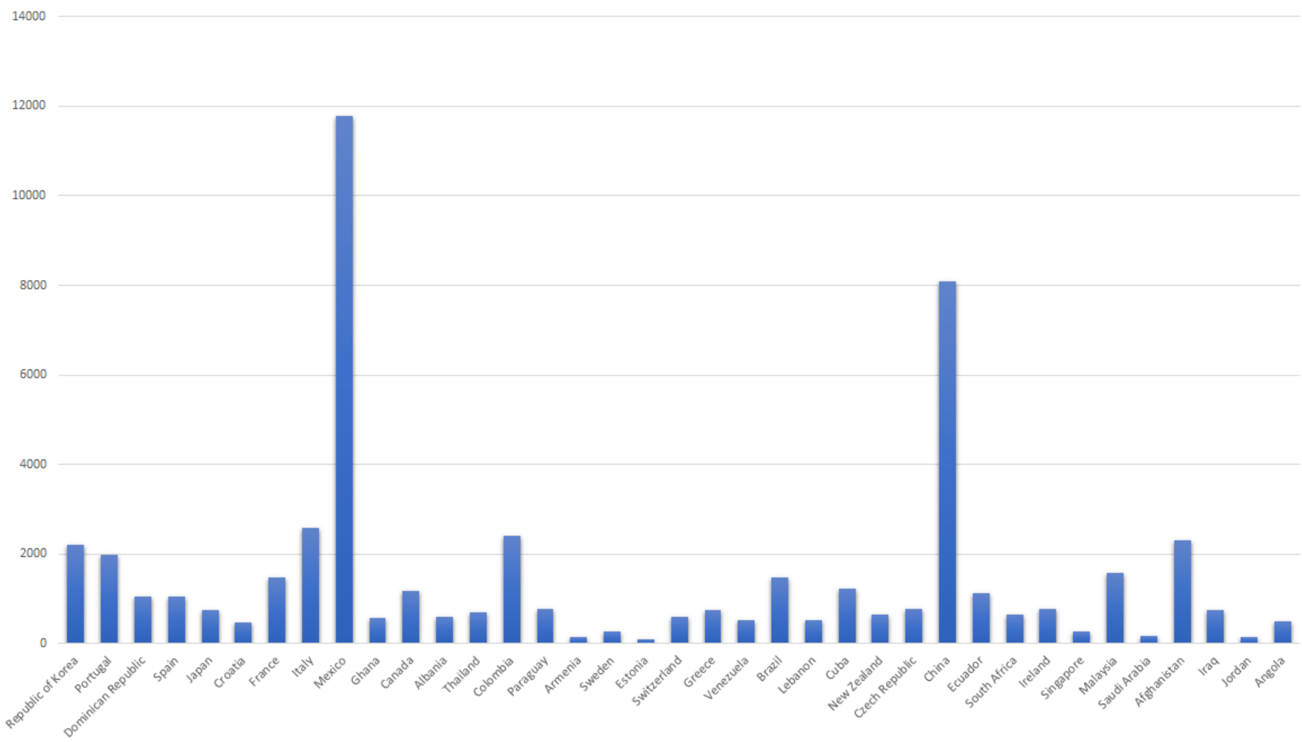


Figure 4.2 *Total migrant stock (thousands) in the sample originating from the robust and semi-robust countries.*



5) Conclusion

This paper has studied the impact of internationally dispersed ethnic networks on FDI by re-examining and extending Tong's (2005) study on ethnic Chinese networks. Estimates have been obtained by using a standard gravity model. Econometric issues related to estimating a gravity model have been accounted for. Multilateral resistance has been controlled for by adding country fixed effects to the empirical model. The issue of having many zeroes in the dependent variable has been tackled by estimating the gravity model with a Heckman sample selection model and a PPML estimation. As the PPML estimation deals with inconsistency caused by log-linearizing heteroskedastic data, its estimates are the most reliable.

Positive and significant estimations have been obtained for both the variable proxying Chinese network strength as for several other proxied ethnic networks. These results indicate that for both hypothesis 1 and hypothesis 2, H_1 can be rejected. The results suggest the presence of several ethnic networks, including ethnic Chinese, affecting bilateral investments. A total of 30 ethnic networks have been distinguished, with the South Korean network being the most impactful. The Chinese ethnic network ranks 23d in terms of its impact on bilateral investments. Unfortunately, the results appear to be affected by a selection bias, causing the main estimates in this study to predominately measure the network effect when FDI originates from developed economies.

An interesting estimate has been obtained for the coefficient for the Angolan ethnic network. Its estimated coefficient is negative, implying that an increase in Angolan migrants in two countries results a decrease of investments between these two countries. There are a few options which could cause this outcome. It could mean that Angolan networks are actively discouraging companies to invest overseas, although this seems highly unlikely. Another option is that the variables constructed to estimate network strength between two countries are not properly specified to only proxy ethnic networks. A bias resulting from endogeneity issues or due to the geographical distribution of the observations seems likely.

Due to the mentioned negative results for the Angolan ethnic network variable, further research on this study's topic is recommended. More in-depth research on determinants of Angolan migration could lead to a better insight regarding the causality related to the results found in this study. An alternative approach regarding the construction of a proxy measuring network strength between countries would furthermore be of added value to overcome endogeneity issues.

6) Limitations

As mentioned in chapter 3.1 and 3.2, the estimations in this study are done using macro data on bilateral FDI and macro data on migrant stocks. The data does not include detailed information regarding the companies or the people involved in international investments. There is furthermore no data publicly available regarding actual networking activities of migrants. This restricts this study to work with a proxy variable measuring ethnic network strength, and the outcomes of this proxy variable should therefore be interpreted as an indication of networking activity, rather than being hard evidence of networking activity.

The scarcity of bilateral FDI data has furthermore restricted this study to work with cross sectional data rather than panel data. Panel data would have been preferred over cross-sectional data since the extra time dimension leads to more sample variability and therefore improves efficiency of the estimated coefficients (Hsiao, 2007). Due to the lack of data over time, certain assumptions must be considered. For instance, the assumption is made that the variables used in the estimation are not subject to reverse causality. This assumption is required since a Granger causality test cannot be performed as it requires lagged variables.

Missing pieces of data furthermore led towards a dataset where observations were not completely randomly distributed. There were relatively more observations for European countries than for countries in either the Americas, South-East Asia, or dispersed over the other regions. This caused a selection bias that affected the main estimates, and with that the validity of this study.

Concluding, a re-evaluation of study should be considered when more data over time, and data regarding developing countries is available. Apart from that, access to data regarding actual networking activities (for example, data from LinkedIn, a social network focussed on users' professional relationships) would result in a far more precise measure of ethnic networking activity rather than the using migrant data. If this data should become accessible to scholars, a re-estimation of the impact of ethnic networking activities on international investments would be advised.

7) References

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

Ammon, C., and Baiardi., A. (2016). *The Benefits of the Bamboo Network in International Trade*. Retrieved from Anna Baiardi's personal website: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CSAE2016&paper_id=917

Beine, M., Bertoli, S., and Fernández-Huertas Moraga, J. (2015). A Practitioners' Guide to Gravity Models of International Migration. *The World Economy, 39*(4), 496-512.

Bergstrand, J H., and Egger, P. "A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises." Journal of International Economics 73.2 (2007): 278-308.

Baier, S., and Bergstrand, J. (2007). Do free trade agreements actually increase members' international trade?. Journal of International Economics, 71(1), 72-95.

Brouwer, J., Paap, R., and Viaene, J. (2008). The trade and FDI effects of EMU enlargement. *Journal of International Money And Finance, 27*(2), 188-208.

Centre d'Etudes Prospectives et d'Informations Internationales (2018, February). Gravity. Retrieved from: http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8

Donald Trump gets serious about building a border wall. (2017, January 25). The Economist. Retrieved from https://www.economist.com/democracy-in-america/2017/01/25/donald-trump-gets-serious-about-building-a-border-wall

Emigrant and immigrant networks in FDI. (2009). *Applied Economics Letters, 16*(12), 1261-1264.

Feenstra, R. (2002). Border Effects and the Gravity Equation: Consistent Methods for Estimation. *Scottish Journal of Political Economy, 49*(5), 491-506.

Felbermayr, G., Jung, B., and Toubal, F. (2010). Ethnic Networks, Information, and International Trade: Revisiting the Evidence. *Annals of Economics And Statistics*, (97/98), 41.

Flisi, S., and Murat, M. (2011). The hub continent. Immigrant networks, emigrant diasporas and FDI. *The Journal of Socio-Economics, 40*(6), 796-805.

Gao, T. (2003). Ethnic Chinese networks and international investment: evidence from inward FDI in China. *Journal of Asian Economics, 14*(4), 611-629.

Genc, M., Gheasi, M., Nijkamp, P., and Poot, J. (2012). The impact of immigration on international trade: a meta-analysis. In P. Nijkamp, J. Poot, and M. Sahin (Eds.), *Migration impact assessmemt. New horizons in regional science*, ch. 9 (pp. 301–337). Northhampton (MA): Edward Elgar Publishing Limited

Girma, S., and Yu, Z. (2002). The link between immigration and trade: Evidence from the United Kingdom. *Weltwirtschaftliches Archiv, 138*(1), 115-130.

Gómez-Herrera, E. (2012). Comparing alternative methods to estimate gravity models of bilateral trade. *Empirical Economics, 44*(3), 1087-1111.

Gould, D. (1994). Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows. *The Review of Economics And Statistics*, *76*(2), 302.

Head, K., and Mayer, T. (2014). Gravity Equations: Workhorse,Toolkit, and Cookbook. *Handbook of International Economics*, 131-195.

Head, K., and Ries, J. (1998). Immigration and Trade Creation: Econometric Evidence from Canada. *The Canadian Journal of Economics*, *31*(1), 47.

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

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes*. *Quarterly Journal of Economics*, *123*(2), 441-487.

Hsiao, C. (2007). Panel data analysis—advantages and challenges. Test, 16(1), 1-22.

Javorcik, B., Özden, Ç., Spatareanu, M., and Neagu, C. (2011). Migrant networks and foreign direct investment. *Journal of Development Economics*, *94*(2), 231-241.

Liu, X. (2009). GATT/WTO Promotes Trade Strongly: Sample Selection and Model Specification. *Review of International Economics*, *17*(3), 428-446.

Murat, M., and Pistoresi, B. (2009). Migrant Networks: Empirical Implications for the Italian Bilateral Trade. *International Economic Journal*, *23*(3), 371-390.

OECD. (2008). *Foreign direct investment statistics Explanatory notes* [Ebook] (pp. 1-6). OECD. Retrieved from https://www.oecd.org/daf/inv/FDI-statistics-explanatory-notes.pdf

Raising the drawbridge. (2016, August 27). The Economist. Retrieved from https://www.economist.com/britain/2016/08/27/raising-the-drawbridge

Rauch, J., and Casella, A. (2003). Overcoming Informational Barriers to International Resource Allocation: Prices and Ties*. *The Economic Journal*, *113*(484), 21-42.

Rauch, J., and Trindade, V. (2002). Ethnic Chinese Networks in International Trade. *Review of Economics And Statistics*, *84*(1), 116-130.

Silva, J., and Tenreyro, S. (2006). The Log of Gravity. *Review of Economics And Statistics, 88*(4), 641-658.

Silva, J., Tenreyro, S., and Windmeijer, F. (2015). Testing Competing Models for Non-negative Data with Many Zeros. *Journal of Econometric Methods*, *4*(1).

Söderbom, Måns. "Econometrics II, Lecture 4: Instrumental Variables Part I." (2011).

Tinbergen, J., and Hekscher, A. (1962). *Shaping the World Economy. Suggestions for an International Economic Policy.* New-York

Tong, S. (2005). Ethnic Networks in FDI and the Impact of Institutional Development. *Review of Development Economics*, *9*(4), 563-580.

Tran, N., Wilson, N., and Hite, D. (2013). 5 Choosing the Best Model in the Presence of Zero Trade: A Fish Product Analysis. *Frontiers of Economics And Globalization*, 127-148.

United Nations Conference on Trade And Development. (2018, February). Bilateral FDI Statistics. Retrieved from: http://unctad.org/en/Pages/DIAE/FDI%20Statistics/FDI-Statistics-Bilateral.aspx

United Nations, Department of Economic and Social Affairs (2012), World Economic Situation and Prospects 2010, United Nations, New York.

United Nations, Department of Economic and Social Affairs, Population Division (2017). International Migration Report 2017: Highlights(ST/ESA/SER.A/404).

United Nations, Department of Economic and Social Affairs, Population Division. (2018, February). International migrant stock: The 2017 revision. Retrieved from: http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates17.shtml

Westerlund, J., and Wilhelmsson, F. (2011). Estimating the gravity model without gravity using panel data. *Applied Economics*, *43*(6), 641-649.

World Bank. (2018, March). GDP (constant 2010 US$). Retrieved from World Bank Indicators: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD

World Bank. (2018, March). Tariff rate, applied, weighted mean, all products (%). Retrieved from World Bank Indicators: https://data.worldbank.org/indicator/TM.TAX.MRCH.WM.AR.ZS

World Bank. (2018, March). Trade (% of GDP). Retrieved from World Bank Indicators: https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS

World Trade Organization. (2012). *A practical guide to trade policy analysis* (pp. 101-136). Geneva: World Trade Organization.

Appendix

Table A1: *Summary statistics of the dependent and independent variables*

Variable	Observations	Mean	Std. Dev.	Min	Max
F_{ij}	1,697	6,800,000,000	24,600,000,000	0	387,000,000,000
Y_i	1,697	1,190,000,000,000	2,600,000,000,000	3,140,000,000	15,000,000,000,000
Y_j	1,697	1,410,000,000,000	2,730,000,000,000	3,140,000,000	15,000,000,000,000
D_{ij}	1,697	6999.75	4,628.69	160,93	19,539.48
Pop_i	1,697	81,200,000	186,000,000	860,559	1,210,000,000
Pop_j	1,697	80,800,000	190,000,000	860,559	1,210,000,000
Rem_i	1,697	654,000,000	25,700,000	588,000,000	710,000,000
Rem_j	1,697	658,000,000	24,500,000	588,000,000	710,000,000
Tar_i	1,681	3.41	2.88	0	14.99
$Chin_{ij}$	1,697	219,000,000,000	141,000,000,000	244,036	4,130,000,000,000
Adj_{ij}	1,697	0.05	0.23	0	1
FTA_{ij}	1,697	0.40	0.49	0	1
$Lang_{ij}$	1,697	0.14	0.35	0	1
$Colony_{ij}$	1,697	0.05	0.21	0	1
$GroGDP_i$	1,697	1.03	0.02	1	1.08

Table A2: *First-stage estimates of the two-stage least squares estimation.*

Variable	Coefficient
$\ln Y_i$	.741***
	(14.79)
$\ln Y_j$	.798***
	(19.08)
$\ln D_{ij}$	.322***
	(5.76)
$\ln Pop_i$	-.453***
	(-8.37)
$\ln Pop_j$	-.530***
	(-15.04)
$\ln Rem_i$	2.277**
	(2.36)
$\ln Rem_j$	4.931***
	(4.41)
$\ln Tar_i$	.031*
	(1.67)
$IV\ 1_{ij}$	-.755
	(-16.94)
$IV\ 2_{ij}$	.477
	(30.73)
Ajd_{ij}	.227
	(1.36)
FTA_{ij}	.046
	(0.45)
$Lang_{ij}$	.308***
	(2.77)
$Colony_{ij}$	.306*
	(1.83)
$GroGDP_i$	-10.213***
	(-4.52)
$Time_{ij}$	.047
	(0.20)
$Constant$	-138.107***
	(-4.48)
Country fixed Effects	No
Obs	1389
R^2	0.652

The dependent variable is $\ln Chin_{ij}$; ***, ** and * denote a significance level of 1%,5% and 10% respectively; T-statistics in parenthesis

Table A3: Developed vs developing countries, as classified in the World Economic Situation and Prospects report (2010) provided by the United Nations

Developed	Developing
Australia	Argentina
Austria	Bangladesh
Belgium	Bolivia (Plurinational State of)
Bulgaria	Brazil
Canada	Chile
Czech Republic	Hong Kong
Denmark	Colombia
Finland	Dominican Republic
France	Ecuador
Germany	Egypt
Hungary	El Salvador
Ireland	Fiji
Italy	Gabon
Japan	India
Netherlands	Indonesia
New Zealand	Libya
Norway	Malaysia
Poland	Mauritius
Portugal	Mexico
Romania	Pakistan
Spain	Papua New Guinea
Sweden	Paraguay
Switzerland	Peru
United Kingdom	Philippines
United States of America	Republic of Korea
	Singapore
	South Africa
	Sri Lanka
	Thailand
	Togo
	Turkey
	Venezuela
	Viet Nam

Table A4a: *Overview of development classification, observations, and observed zeroes of the dependent variable in numbers*

Classification	Countries	Observations		Observed zeroes in F_{ij}	
		Host i	Source j	Host i	Source j
Developed	25	876	1011	133	79
Undeveloped	33	821	686	120	174
Total	58	1697		253	

Table A4b: *Overview of development classification, observations, and observed zeroes of the dependent variable in percentages*

Classification	Countries	Observations		Observed zeroes in F_{ij}	
		Host i	Source j	Host i	Source j
Developed	43.1	51.6	59.6	52.6	31.2
Undeveloped	56.9	48.4	40.4	47.4	68.8
Total			100		

Table A5: *Countries with >700 observations in their PPML estimations*

Country of origin	Observations	Country of origin	Observations
Afghanistan	907	Japan	1523
Albania	788	Jordan	988
Algeria	977	Kenya	874
Angola	817	Lebanon	1040
Argentina	1014	Libya	798
Armenia	808	Lithuania	745
Australia	1461	Malaysia	1495
Austria	1024	Mexico	1012
Bangladesh	1321	Morocco	941
Belgium	1042	Myanmar	823
Bolivia	837	Nepal	878
Brazil	1159	Netherlands	1149
Bulgaria	859	New Zealand	1163
Cambodia	776	Nigeria	1161
Canada	1664	Norway	1019
Chile	968	Pakistan	1633
China	1694	Panama	818
Colombia	878	Paraguay	796
Costa Rica	788	Peru	977
Croatia	878	Philippines	1380
Cuba	905	Poland	1059
Czech Republic	892	Portugal	1128
Republic of Korea	734	Republic of Korea	1284
D.R. Congo	772	Romania	994
Denmark	1018	Russian Federation	1380
Dominican Republic	740	Saudi Arabia	1044
Ecuador	795	Senegal	804
Egypt	1032	Singapore	1124
Estonia	763	South Africa	1095
Ethiopia	811	Spain	1178
Finland	918	Sri Lanka	1262
France	1477	Sweden	1105
Germany	1484	Switzerland	1182
Ghana	834	Syrian Arab Republic	1012
Greece	984	Thailand	1295
Guatemala	825	Tunisia	829
Hungary	890	Turkey	992
India	1660	Ukraine	970
Indonesia	1479	United Kingdom	1758
Iran	1220	United States	1750
Iraq	1207	Uruguay	877
Ireland	1092	Venezuela	923
Israel	982	Vietnam	1220

Disclaimer regarding table A6 to A34:
Dependent variable in the OLS and Heckman sample selection model is the log of the 2010 inward FDI stock in i from j; Dependent variable in the PPML estimations is the 2010 inward FDI stock in millions in i from j; Country fixed effects are added to control for multilateral resistance; ***,** and * denote a significance level of 0.1%, 1% and 5% respectively; Observations clustered by host countries; Robust t-statistics in parenthesis.

Table A6: *Results for migrants from Estonia*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.142***	-0.541*	-1.179***	-0.254	-0.254
	(-8.51)	(-2.29)	(-10.47)	(-1.27)	(-1.27)
$\ln Mig_{ij}$	0.191*	-0.805	0.592***	0.336***	
	(2.27)	(-0.01)	(9.13)	(3.94)	
$\ln alt\_Mig_{ij}$					0.254***
					(3.94)
Ajd_{ij}	-0.00023	0.832	0.0884	0.306	0.306
	(0)	(1.04)	(0.41)	(1.7)	(1.7)
$Lang_{ij}$	0.608*	1.979**		0.432**	0.432**
	(2.58)	(2.71)		(2.67)	(2.67)
FTA_{ij}	0.335	0.786	0.256	0.478	0.478
	(1.28)	(1.68)	(1.14)	(1.73)	(1.73)
$Colony_{ij}$	1.150**	-0.165	1.456***	0.445**	0.445**
	(3.37)	(-0.26)	(6.27)	(2.71)	(2.71)
$Time_{ij}$	2.057***	1.384	2.016**	3.427***	3.427***
	(-4.27)	(-1.18)	(-2.65)	(-5.75)	(-5.75)
Obs	681	763	681	763	763
Mills λ			-0.546		
			(-1.89)		

Table A7: *Results for migrants from France*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.106***	-0.471***	-1.225***	-0.560***	-0.560***
	(-11.49)	(-3.67)	(-14.13)	(-4.45)	(-4.45)
$\ln Mig_{ij}$	0.652***	0.455*	0.293*	0.647***	
	(31.97)	(2.44)	(2.54)	(23.84)	
$\ln alt\_Mig_{ij}$					0.582***
					(3.54)
Ajd_{ij}	0.149	1.210*	0.21	-0.0674	-0.0674
	(0.58)	(2.30)	(1.00)	(-0.4)	(-0.4)
$Lang_{ij}$	1.048***	1.509***		0.430*	0.430*
	(6.39)	(5.84)		(2.24)	(2.24)
FTA_{ij}	0.203	-0.223	0.162	0.237	0.237
	(1.21)	(-1.08)	(1.05)	(0.98)	(0.98)
$Colony_{ij}$	0.818**	0.394	1.357***	0.689**	0.689**
	(3.16)	(0.96)	(6.32)	(3.26)	(3.26)
$Time_{ij}$	1.268**	1.975***	1.089	3.085***	3.085***
	(-3.19)	(-3.65)	(-1.73)	(-4.5)	(-4.5)
Obs	1235	1477	1235	1477	1477
Mills λ			-0.086		
			(-0.35)		

Table A8: *Results for migrants from the Republic of Korea*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.223***	-0.517***	-1.326***	-0.414**	-0.414**
	(-11.2)	(-3.47)	(-15.73)	(-2.95)	(-2.95)
$\ln Mig_{ij}$	0.974***	3.981***	0.909***	1.404***	
	(6.06)	(42.29)	(3.58)	(12.74)	
$\ln alt\_Mig_{ij}$					1.166***
					(14.03)
Ajd_{ij}	0.0232	1.011	0.0976	0.304	0.304
	(0.1)	(1.75)	(0.49)	(1.73)	(1.73)
$Lang_{ij}$	0.659**	1.118***		0.428*	0.428*
	(3.49)	(3.51)		(2.1)	(2.1)
FTA_{ij}	0.241	-0.115	0.168	0.950**	0.950**
	(1.14)	(-0.5)	(1.14)	(3.1)	(3.1)
$Colony_{ij}$	1.164***	0.442	1.534***	0.660***	0.660***
	(5.18)	(1.06)	(7.79)	(4.01)	(4.01)
$Time_{ij}$	1.799**	2.190***	1.817*	3.158***	3.158***
	(-3.25)	(-3.31)	(-2.29)	(-5.74)	(-5.74)
Obs	1115	1284	1115	1284	1284
Mills λ			0.135		
			(0.52)		

Table A9: *Results for migrants from Venezuela*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.192***	-0.605***	-1.288***	-0.392*	-0.392*
	(-8.99)	(-3.30)	(-13.51)	(-2.04)	(-2.04)
$\ln Mig_{ij}$	0.489***	7.351***	2.712***	0.227***	
	(4.18)	(51.88)	(6.14)	(7.91)	
$\ln alt\_Mig_{ij}$					0.274***
					(13.06)
Ajd_{ij}	-0.00894	1.007	0.0484	0.238	0.238
	(-0.04)	(1.56)	(0.24)	(1.39)	(1.39)
$Lang_{ij}$	0.871**	1.668***		0.444**	0.444**
	(3.37)	(3.44)		(2.93)	(2.93)
FTA_{ij}	1.026**	0.123	1.479***	0.470**	0.470**
	(3.24)	(0.24)	(6.73)	(2.66)	(2.66)
$Colony_{ij}$	0.282	-0.04	0.182	0.325	0.325
	(1.03)	(-0.12)	(0.95)	(1.19)	(1.19)
$Time_{ij}$	-1.526***	-0.828	-1.443*	-3.343***	-3.343***
	(-7.16)	(-1.02)	(-2.33)	(-6.84)	(-6.84)
Obs	814	923	814	923	923
Mills λ			-0.512*		
			(-1.97)		

Table A10: *Results for migrants from Albania*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.189***	-0.487*	-1.244***	-0.317	-0.317
	(-9.35)	(-2.36)	(-11.13)	(-1.62)	(-1.62)
$\ln Mig_{ij}$	0.300***	30600.6***	-0.328	0.349***	
	(4.28)	(364085.12)	(-0.54)	(4.36)	
$\ln alt\_Mig_{ij}$					0.365***
					(4.36)
Ajd_{ij}	-0.168	0.567	-0.0188	0.253	0.253
	(-0.82)	(0.83)	(-0.09)	(1.39)	(1.39)
$Lang_{ij}$	0.750**	1.836**		0.410*	0.410*
	(2.86)	(3.19)		(2.44)	(2.44)
FTA_{ij}	0.114	-0.0527	0.159	0.485	0.485
	(0.41)	(-0.15)	(0.65)	(1.79)	(1.79)
$Colony_{ij}$	0.910**	0.512	1.185***	0.456**	0.456**
	(2.81)	(1.03)	(5.51)	(2.71)	(2.71)
$Time_{ij}$	1.366*	1.402	1.266	2.997***	2.997***
	(-2.54)	(-1.65)	(-1.87)	(-9.62)	(-9.62)
Obs	694	788	694	788	788
Mills λ			-0.472		
			(-1.61)		

Table A11: *Results for migrants from Algeria*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.142***	-0.707***	-1.224***	-0.395*	-0.395*
	(-9.26)	(-3.96)	(-12.46)	(-2.08)	(-2.08)
$\ln Mig_{ij}$	0.414***	2.278	-1.845	0.416***	
	(8.5)	(0.34)	(-0.39)	(8.58)	
$\ln alt\_Mig_{ij}$					0.168***
					(8.96)
Ajd_{ij}	-0.0282	0.968	0.065	0.222	0.222
	(-0.13)	(1.51)	(0.32)	(1.3)	(1.3)
$Lang_{ij}$	1.008***	1.889***		0.455**	0.455**
	(4.67)	(4.15)		(3)	(3)
FTA_{ij}	0.282	-0.00621	0.225	0.326	0.326
	(1.04)	(-0.02)	(1.17)	(1.19)	(1.19)
$Colony_{ij}$	0.870**	0.29	1.322***	0.485**	0.485**
	(3.15)	(0.62)	(6.51)	(2.67)	(2.67)
$Time_{ij}$	1.562***	1.978**	1.350*	3.378***	3.378***
	(-7.07)	(-2.89)	(-2.15)	(-7.09)	(-7.09)
Obs	849	977	849	977	977
Mills λ			-0.615*		
			(-2.51)		

Table A12: *Results for migrants from Angola*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.179***	-0.769***	-1.319***	-0.296	-0.296
	(-8.67)	(-3.6)	(-12.92)	(-1.42)	(-1.42)
$\ln Mig_{ij}$	-0.287***	-1.376	0.132	-0.533***	
	(-18.3)	(-0.01)	(0.46)	(-5.31)	
$\ln alt\_Mig_{ij}$					-0.556***
					(-5.32)
Ajd_{ij}	0.0375	10.29	0.0614	0.438*	0.438*
	(0.14)	((.))	(0.27)	(2.13)	(2.13)
$Lang_{ij}$	0.900***	1.175*		0.430*	0.430*
	(4.09)	(2.42)		(2.54)	(2.54)
FTA_{ij}	0.232	0.0104	0.205	0.409	0.409
	(0.85)	(0.03)	(1.05)	(1.38)	(1.38)
$Colony_{ij}$	1.051**	0.634	1.631***	0.542**	0.542**
	(3.24)	(0.83)	(7.35)	(2.96)	(2.96)
$Time_{ij}$	1.740**	2.468**	1.555*	3.376***	3.376***
	(-3.09)	(-3.02)	(-2.03)	(-5.18)	(-5.18)
Obs	710	817	710	817	817
Mills λ			-0.159		
			(-0.59)		

Table A13: *Results for migrants from Armenia*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.343***	*** -0.883***	-1.384***	-0.430*	-0.430*
	(-8.93)	(-4.00)	(-11.97)	(-2.11)	(-2.11)
$\ln Mig_{ij}$	0.603***	2.128***	-0.18	0.289**	
	(4.31)	(9.13)	(-0.71)	(2.76)	
$\ln alt\_Mig_{ij}$					0.329**
					(2.76)
Ajd_{ij}	-0.266	0.697	-0.173	0.215	0.215
	(-1.37)	(1.00)	(-0.82)	(1.17)	(1.17)
$Lang_{ij}$	0.694*	1.227*		0.437*	0.437*
	(2.57)	(2.37)		(2.51)	(2.51)
FTA_{ij}	-0.158	-0.122	-0.227	0.221	0.221
	(-0.46)	(-0.3)	(-0.91)	(0.75)	(0.75)
$Colony_{ij}$	0.880*	0.359	1.170***	0.393*	0.393*
	(2.62)	(0.7)	(5.3)	(2.35)	(2.35)
$Time_{ij}$	1.613***	1.348	1.535*	3.404***	3.404***
	(-6.58)	(-1.63)	(-2.52)	(-7.27)	(-7.27)
Obs	713	808	713	808	808
Mills λ			-0.521		
			(-1.88)		

Table A14: *Results for migrants from Brazil*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.215***	-0.538***	-1.340***	-0.605***	-0.605***
	(-10.71)	(-3.49)	(-15.11)	(-3.72)	(-3.72)
$\ln Mig_{ij}$	0.298***	3.038***	-0.0106	0.232**	
	(4.73)	(35.64)	(-0.03)	(2.82)	
$\ln alt\_Mig_{ij}$					0.252**
					(2.82)
Ajd_{ij}	-0.079	0.573	0.0256	0.0367	0.0367
	(-0.37)	(0.97)	(0.13)	(0.23)	(0.23)
$Lang_{ij}$	0.839***	1.417***		0.422**	0.422**
	(4.05)	(4)		(2.77)	(2.77)
FTA_{ij}	0.251	-0.206	0.204	0.264	0.264
	(1.05)	(-0.83)	(1.2)	(1)	(1)
$Colony_{ij}$	0.934**	0.47	1.334***	0.511**	0.511**
	(3.42)	(1.09)	(6.53)	(2.71)	(2.71)
$Time_{ij}$	1.208**	1.453*	1.089	3.042***	3.042***
	(-2.95)	(-2.14)	(-1.55)	(-10.27)	(-10.27)
Obs	1006	1159	1006	1159	1159
Mills λ			-0.0731		
			(-0.28)		

Table A15: Results for migrants from Chile

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.182***	-0.457*	-1.286***	-0.393*	-0.393*
	(-9.02)	(-2.27)	(-13.08)	(-2.02)	(-2.02)
$\ln Mig_{ij}$	0.382***	1.663***	-0.293	0.409***	
	(4.35)	(16.96)	(-0.87)	(4.88)	
$\ln alt\_Mig_{ij}$					0.125***
					(11.41)
Ajd_{ij}	-0.0154	1.325*	-0.00258	0.214	0.214
	(-0.07)	(2.03)	(-0.01)	(1.21)	(1.21)
$Lang_{ij}$	0.815***	1.793***		0.438**	0.438**
	(3.91)	(4.02)		(2.86)	(2.86)
FTA_{ij}	0.188	-0.116	0.105	0.402	0.402
	(0.67)	(-0.34)	(0.55)	(1.44)	(1.44)
$Colony_{ij}$	0.948**	0.2	1.377***	0.468**	0.468**
	(2.98)	(0.35)	(6.42)	(2.69)	(2.69)
$Time_{ij}$	1.964***	2.220*	1.807*	3.341***	3.341***
	(-3.87)	(-2.55)	(-2.35)	(-6.4)	(-6.4)
Obs	849	968	849	968	968
Mills λ			-0.617*		
			(-2.41)		

Table A16: Results for migrants from Croatia

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.174***	-0.597**	-1.262***	-0.418*	-0.418*
	(-9.14)	(-2.94)	(-13.14)	(-2.19)	(-2.19)
$\ln Mig_{ij}$	0.642***	5.176	-0.745	0.688***	
	(5.42)	(0.03)	(-0.82)	(4.5)	
$\ln alt\_Mig_{ij}$					0.602***
					(4.5)
Ajd_{ij}	-0.109	0.937	-0.0401	0.19	0.19
	(-0.49)	(1.23)	(-0.2)	(1.08)	(1.08)
$Lang_{ij}$	0.812**	2.061***		0.451**	0.451**
	(3.53)	(3.7)		(2.93)	(2.93)
FTA_{ij}	0.318	0.556	0.213	0.379	0.379
	(1.16)	(1.46)	(1.1)	(1.35)	(1.35)
$Colony_{ij}$	1.087**	-0.231	1.524***	0.481**	0.481**
	(3.44)	(-0.4)	(6.88)	(2.66)	(2.66)
$Time_{ij}$	1.923***	1.609	1.830*	3.330***	3.330***
	(-3.84)	(-1.47)	(-2.41)	(-6.46)	(-6.46)
Obs	775	878	775	878	878
Mills λ			-0.556*		
			(-2.23)		

Table A17: *Results for migrants from Cuba*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.167***	-0.539**	-1.294***	-0.376*	-0.376*
	(-7.81)	(-2.99)	(-13.56)	(-1.97)	(-1.97)
$\ln Mig_{ij}$	0.379***	-2.461***	0.114	0.145***	
	(4.31)	(-20.21)	(0.36)	(7.02)	
$\ln alt\_Mig_{ij}$					0.294**
					(2.83)
Ajd_{ij}	-0.0352	1.091	-0.00025	0.249	0.249
	(-0.16)	(1.7)	(0)	(1.45)	(1.45)
$Lang_{ij}$	0.871***	1.886***		0.441**	0.441**
	(3.65)	(3.84)		(2.91)	(2.91)
FTA_{ij}	0.283	0.0782	0.181	0.345	0.345
	(0.94)	(0.24)	(0.95)	(1.25)	(1.25)
$Colony_{ij}$	1.009**	0.0337	1.482***	0.471**	0.471**
	(3.17)	(0.06)	(6.75)	(2.67)	(2.67)
$Time_{ij}$	1.552***	1.345	1.445*	3.350***	3.350***
	(-7.28)	(-1.65)	(-2.35)	(-6.8)	(-6.8)
Obs	793	905	905	905	905
Mills λ			-0.336		
			(-1.31)		

Table A18: *Results for migrants from the Dominican Republic*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.320***	-0.651*	-1.438***	-0.420*	-0.420*
	(-8.46)	(-2.31)	(-13.35)	(-2.08)	(-2.08)
$\ln Mig_{ij}$	0.430***	-2.060***	0.173	0.741***	
	(32.23)	(-10.76)	(0.64)	(12.27)	
$\ln alt\_Mig_{ij}$					0.656***
					(11.15)
Ajd_{ij}	-0.14	11.87	-0.0987	0.233	0.233
	(-0.57)	((.))	(-0.49)	(1.29)	(1.29)
$Lang_{ij}$	0.804**	1.750*		0.451**	0.451**
	(2.91)	(2.49)		(2.63)	(2.63)
FTA_{ij}	-0.0656	0.363	-0.185	0.208	0.208
	(-0.2)	(0.65)	(-0.83)	(0.7)	(0.7)
$Colony_{ij}$	0.924*	0.337	1.418***	0.376*	0.376*
	(2.55)	(0.34)	(5.95)	(2.25)	(2.25)
$Time_{ij}$	1.880***	1.846	1.756*	3.429***	3.429***
	(-3.7)	(-1.36)	(-2.39)	(-6.33)	(-6.33)
Obs	659	740	659	740	740
Mills λ			-0.381		
			(-1.37)		

Table A19: Results for migrants from Ecuador

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.287***	-0.622*	-1.405***	-0.436*	-0.436*
	(-8.48)	(-2.54)	(-12.59)	(-2.16)	(-2.16)
$\ln Mig_{ij}$	0.140***	-1.481***	0.108	0.186***	
	(3.61)	(-10.14)	(0.46)	(8.58)	
$\ln alt\_Mig_{ij}$					0.219***
					(11.81)
Ajd_{ij}	-0.119	19.15	-0.081	0.22	0.22
	(-0.5)	((.))	(-0.39)	(1.22)	(1.22)
$Lang_{ij}$	0.872**	1.460*		0.459**	0.459**
	(3.5)	(2.54)		(2.69)	(2.69)
FTA_{ij}	-0.0718	-0.303	-0.158	0.175	0.175
	(-0.22)	(-0.69)	(-0.71)	(0.59)	(0.59)
$Colony_{ij}$	0.768*	0.511	1.284***	0.366*	0.366*
	(2.18)	(0.55)	(5.43)	(2.2)	(2.2)
$Time_{ij}$	1.895**	2.092*	1.721*	3.431***	3.431***
	(-3.34)	(-2.32)	(-2.28)	(-6.35)	(-6.35)
Obs	702	795	702	795	795
Mills λ			-0.449		
			(-1.58)		

Table A20: Results for migrants from Ghana

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.216***	-0.614**	-1.260***	-0.327	-0.327
	(-8.47)	(-3.07)	(-10.54)	(-1.64)	(-1.64)
$\ln Mig_{ij}$	0.0183	-0.0119	0.229*	0.375***	
	(0.06)	(-0.11)	(2.4)	(6.23)	
$\ln alt\_Mig_{ij}$					0.388***
					(3.77)
Ajd_{ij}	-0.241	0.591	-0.188	0.281	0.281
	(-1.06)	(0.82)	(-0.77)	(1.62)	(1.62)
$Lang_{ij}$	0.295	1.298**		0.358*	0.358*
	(1.11)	(2.78)		(2.24)	(2.24)
FTA_{ij}	0.172	0.321	0.183	0.383	0.383
	(0.63)	(0.95)	(0.78)	(1.4)	(1.4)
$Colony_{ij}$	1.118**	0.397	1.285***	0.442**	0.442**
	(3.51)	(0.8)	(5.52)	(2.73)	(2.73)
$Time_{ij}$	2.476***	2.471**	2.527***	3.765***	3.765***
	(-7.13)	(-3.03)	(-3.44)	(-8.13)	(-8.13)
Obs	731	834	731	834	834
Mills λ			0.392		
			(1.37)		

Table A21: Results for migrants from Greece

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.185***	-0.626***	-1.294***	-0.384*	-0.384*
	(-8.99)	(-3.71)	(-13.19)	(-2.02)	(-2.02)
$\ln Mig_{ij}$	0.260***	11.52***	-0.731	0.268***	
	(6.53)	(87.51)	(-0.34)	(6.76)	
$\ln alt\_Mig_{ij}$					0.238***
					(6.76)
Ajd_{ij}	-0.0251	1.023	0.0314	0.243	0.243
	(-0.11)	(1.7)	(0.16)	(1.43)	(1.43)
$Lang_{ij}$	0.878***	1.310**		0.442**	0.442**
	(3.94)	(3.21)		(2.94)	(2.94)
FTA_{ij}	0.174	-0.148	0.128	0.327	0.327
	(0.6)	(-0.52)	(0.68)	(1.2)	(1.2)
$Colony_{ij}$	0.951**	0.458	1.363***	0.468**	0.468**
	(3.25)	(0.99)	(6.62)	(2.65)	(2.65)
$Time_{ij}$	1.524***	1.587*	1.401*	3.352***	3.352***
	(-6.96)	(-2.29)	(-2.25)	(-6.79)	(-6.79)
Obs	851	984	851	984	984
Mills λ			-0.295		
			(-1.16)		

Table A22: Results for migrants from Israel

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.193***	-0.578***	-1.310***	-0.385*	-0.385*
	(-8.78)	(-3.33)	(-13.75)	(-2.02)	(-2.02)
$\ln Mig_{ij}$	0.534***	26.83***	-1.433	0.713***	
	(35.49)	(185.99)	(-0.28)	(29.98)	
$\ln alt\_Mig_{ij}$					0.273***
					(9.22)
Ajd_{ij}	-0.0478	1.157	-0.00441	0.24	0.24
	(-0.21)	(1.85)	(-0.02)	(1.41)	(1.41)
$Lang_{ij}$	0.922***	1.672***		0.445**	0.445**
	(4.19)	(3.95)		(2.96)	(2.96)
FTA_{ij}	0.156	-0.0851	0.0847	0.328	0.328
	(0.54)	(-0.29)	(0.46)	(1.19)	(1.19)
$Colony_{ij}$	0.923**	0.186	1.401***	0.469**	0.469**
	(2.97)	(0.37)	(6.7)	(2.65)	(2.65)
$Time_{ij}$	1.541***	1.765*	1.369*	3.358***	3.358***
	(-7.08)	(-2.57)	(-2.21)	(-6.72)	(-6.72)
Obs	852	982	852	982	982
Mills λ			-0.424		
			(-1.74)		

Table A23: Results for migrants from Italy

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
ln D_{ij}	-1.195***	-0.581***	-1.345***	-0.476**	-0.476**
	(-11.53)	(-4.01)	(-15.07)	(-2.74)	(-2.74)
ln Mig_{ij}	0.303***	0.739	0.0923	0.399***	
	(22.92)	(0.77)	(0.13)	(24.43)	
ln alt_Mig_{ij}					0.704***
					(3.42)
Ajd_{ij}	0.0187	1.112*	0.056	0.173	0.173
	(0.09)	(2.01)	(0.28)	(1.09)	(1.09)
$Lang_{ij}$	1.012***	1.418***		0.463**	0.463**
	(5.74)	(5.16)		(3.05)	(3.05)
FTA_{ij}	0.253	-0.162	0.211	0.283	0.283
	(1.22)	(-0.69)	(1.29)	(1.06)	(1.06)
$Colony_{ij}$	0.944***	0.384	1.529***	0.483**	0.483**
	(3.71)	(0.93)	(7.68)	(2.74)	(2.74)
$Time_{ij}$	1.210**	1.553*	1.001	1.973***	1.973***
	(-3.04)	(-2.54)	(-1.65)	(-4.89)	(-4.89)
Obs	1084	1289	1084	1289	1289
Mills λ			-0.00626		
			(-0.03)		

Table A24: Results for migrants from Japan

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
ln D_{ij}	-1.215***	-0.442***	-1.335***	-0.532***	-0.532***
	(-11.6)	(-3.39)	(-15.7)	(-3.62)	(-3.62)
ln Mig_{ij}	0.243	6.343***	1.391**	0.529***	
	(1.38)	(73.62)	(3.14)	(28.49)	
ln alt_Mig_{ij}					0.762***
					(8.67)
Ajd_{ij}	-0.016	0.788	0.0592	0.249	0.249
	(-0.07)	(1.66)	(0.3)	(1.54)	(1.54)
$Lang_{ij}$	0.770***	1.114***		0.448*	0.448*
	(4.69)	(4.08)		(2.32)	(2.32)
FTA_{ij}	0.264	-0.107	0.197	0.780**	0.780**
	(1.45)	(-0.53)	(1.34)	(2.68)	(2.68)
$Colony_{ij}$	1.078***	0.481	1.505***	0.674***	0.674***
	(4.7)	(1.17)	(7.65)	(4.05)	(4.05)
$Time_{ij}$	1.182	2.035***	1.059	2.962***	2.962***
	(-1.86)	(-3.3)	(-1.44)	(-5.11)	(-5.11)
Obs	1304	1523	1523	1523	1523
Mills λ			0.223		
			(0.87)		

Table A25: *Results for migrants from Lebanon*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.102***	-0.513**	-1.247***	-0.401*	-0.401*
	(-9.13)	(-3.18)	(-12.65)	(-2.11)	(-2.11)
$\ln Mig_{ij}$	0.234***	-0.0821	0.123	0.232***	
	(6.86)	(-0.18)	(0.36)	(9.04)	
$\ln alt\_Mig_{ij}$					0.242***
					(6.86)
Ajd_{ij}	0.0448	1.111	0.111	0.217	0.217
	(0.2)	(1.87)	(0.55)	(1.27)	(1.27)
$Lang_{ij}$	1.025***	1.897***		0.460**	0.460**
	(4.95)	(4.66)		(3.05)	(3.05)
FTA_{ij}	0.215	-0.119	0.192	0.31	0.31
	(0.79)	(-0.43)	(1)	(1.13)	(1.13)
$Colony_{ij}$	0.860**	0.209	1.378***	0.481**	0.481**
	(3.15)	(0.47)	(6.68)	(2.65)	(2.65)
$Time_{ij}$	1.566***	1.965**	1.452*	3.377***	3.377***
	(-6.86)	(-3.01)	(-2.25)	(-7.1)	(-7.1)
Obs	891	1040	891	1040	1040
Mills λ			-0.0828		
			(-0.34)		

Table A26: *Results for migrants from Mexico*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.179***	-0.415*	-1.284***	-0.367	-0.367
	(-10.23)	(-2.37)	(-13.16)	(-1.86)	(-1.86)
$\ln Mig_{ij}$	0.631***	-9.412***	0.553	0.548***	
	(11.85)	(-79.76)	(0.7)	(10.77)	
$\ln alt\_Mig_{ij}$					0.459**
					(3.06)
Ajd_{ij}	0.0619	1.299*	0.108	0.231	0.231
	(0.27)	(2.16)	(0.56)	(1.37)	(1.37)
$Lang_{ij}$	0.782***	1.582***		0.417**	0.417**
	(3.6)	(3.63)		(2.64)	(2.64)
FTA_{ij}	0.187	0.181	0.13	0.36	0.36
	(0.94)	(0.6)	(0.69)	(1.33)	(1.33)
$Colony_{ij}$	0.950**	0.173	1.372***	0.462**	0.462**
	(3.06)	(0.34)	(6.49)	(2.8)	(2.8)
$Time_{ij}$	1.648***	1.535*	1.536*	3.408***	3.408***
	(-7.63)	(-2.16)	(-2.47)	(-6.88)	(-6.88)
Obs	885	1012	885	1012	1012
Mills λ			-0.213		
			(-0.82)		

Table A27: Results for migrants from Morocco

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.191***	-0.624***	-1.293***	-0.383*	-0.383*
	(-8.99)	(-3.6)	(-13.12)	(-2.01)	(-2.01)
$\ln Mig_{ij}$	0.106***	-3.067	0.193	0.472***	
	(4.24)	(-0.02)	(0.41)	(8.47)	
$\ln alt\_Mig_{ij}$					0.188***
					(9.24)
Ajd_{ij}	-0.0555	0.968	0.0101	0.244	0.244
	(-0.25)	(1.59)	(0.05)	(1.43)	(1.43)
$Lang_{ij}$	0.875***	1.539***		0.442**	0.442**
	(3.67)	(3.48)		(2.93)	(2.93)
FTA_{ij}	0.233	-0.0337	0.157	0.325	0.325
	(0.81)	(-0.11)	(0.81)	(1.19)	(1.19)
$Colony_{ij}$	1.043***	0.358	1.451***	0.467**	0.467**
	(3.66)	(0.76)	(6.88)	(2.64)	(2.64)
$Time_{ij}$	1.572***	1.391	1.450*	3.356***	3.356***
	(-7.24)	(-1.91)	(-2.32)	(-6.71)	(-6.71)
Obs	818	941	818	941	941
Mills λ			-0.360		
			(-1.42)		

Table A28: Results for migrants from New Zealand

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.117***	-0.537**	-1.158***	-0.482***	-0.482***
	(-10.39)	(-2.91)	(-14.18)	(-3.93)	(-3.93)
$\ln Mig_{ij}$	0.246***	-8.831	1.685	0.213***	
	(5.51)	(-0.02)	(0.87)	(5.71)	
$\ln alt\_Mig_{ij}$					0.202***
					(5.71)
Ajd_{ij}	-0.0515	1.142	0.0472	0.224	0.224
	(-0.23)	(1.6)	(0.24)	(1.56)	(1.56)
$Lang_{ij}$	0.736***	1.999***		0.318*	0.318*
	(3.65)	(3.85)		(2.09)	(2.09)
FTA_{ij}	0.267	0.282	0.185	0.25	0.25
	(1.5)	(0.98)	(1.24)	(1.26)	(1.26)
$Colony_{ij}$	0.974***	-0.141	1.314***	0.532***	0.532***
	(3.58)	(-0.25)	(6.5)	(3.4)	(3.4)
$Time_{ij}$	1.796**	2.345**	1.677*	3.350***	3.350***
	(-3.06)	(-2.66)	(-2.18)	(-6.45)	(-6.45)
Obs	1033	1165	1033	1163	1163
Mills λ			-0.655**		
			(-2.75)		

Table A29: Results for migrants from Paraguay

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.099***	-0.612**	-1.269***	-0.361	-0.361
	(-7.64)	(-2.59)	(-13.01)	(-1.89)	(-1.89)
$\ln Mig_{ij}$	0.253***	0.713	-0.0561	0.294***	
	(3.73)	(0.01)	(-0.53)	(5.16)	
$\ln alt\_Mig_{ij}$					0.339***
					(5.16)
Ajd_{ij}	0.0207	10.29	0.0586	0.294	0.294
	(0.09)	((.))	(0.29)	(1.71)	(1.71)
$Lang_{ij}$	0.935***	1.459*		0.427**	0.427**
	(3.95)	(2.57)		(2.83)	(2.83)
FTA_{ij}	0.294	-0.113	0.198	0.337	0.337
	(0.94)	(-0.29)	(1.05)	(1.22)	(1.22)
$Colony_{ij}$	1.049**	0.426	1.691***	0.467*	0.467*
	(3.19)	(0.48)	(7.35)	(2.5)	(2.5)
$Time_{ij}$	1.852*	1.576	1.677*	3.365***	3.365***
	(-2.69)	(-1.34)	(-2.24)	(-5.62)	(-5.62)
Obs	709	796	796	796	796
Mills λ			-0.125		
			(-0.42)		

Table A30: Results for migrants from Peru

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.169***	-0.430*	-1.276***	-0.518**	-0.518**
	(-9.17)	(-2.34)	(-13.7)	(-2.96)	(-2.96)
$\ln Mig_{ij}$	0.395***	2.922***	-0.612	0.435***	
	(4.43)	(31.98)	(-0.99)	(5.17)	
$\ln alt\_Mig_{ij}$					0.235***
					(23.4)
Ajd_{ij}	-0.253	1.094	-0.188	0.144	0.144
	(-1.33)	(1.57)	(-0.93)	(0.89)	(0.89)
$Lang_{ij}$	0.899***	1.692***		0.413**	0.413**
	(4.32)	(3.59)		(2.76)	(2.76)
FTA_{ij}	0.351	0.199	0.242	0.203	0.203
	(1.37)	(0.65)	(1.32)	(0.79)	(0.79)
$Colony_{ij}$	0.974**	-0.169	1.483***	0.488**	0.488**
	(3.16)	(-0.26)	(6.84)	(2.7)	(2.7)
$Time_{ij}$	1.803**	2.439*	1.675*	3.360***	3.360***
	(-3.45)	(-2.2)	(-2.2)	(-6.56)	(-6.56)
Obs	864	977	864	977	977
Mills λ			-0.404		
			(-1.47)		

Table A31: Results for migrants from Portugal

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.183***	-0.602***	-1.286***	-0.432*	-0.432*
	(-10.13)	(-3.57)	(-14.55)	(-2.54)	(-2.54)
$\ln Mig_{ij}$	0.579***	-1.956***	0.237	0.848***	
	(3.6)	(-20.61)	(0.76)	(10.97)	
$\ln alt\_Mig_{ij}$					0.959***
					(10.98)
Ajd_{ij}	-0.0174	1.13	0.0411	0.19	0.19
	(-0.08)	(1.83)	(0.21)	(1.18)	(1.18)
$Lang_{ij}$	0.832***	1.648***		0.397**	0.397**
	(4.56)	(3.9)		(2.79)	(2.79)
FTA_{ij}	0.165	-0.00074	0.0907	0.329	0.329
	(0.78)	(0)	(0.56)	(1.25)	(1.25)
$Colony_{ij}$	1.015**	0.136	1.445***	0.521**	0.521**
	(3.28)	(0.27)	(6.81)	(3.03)	(3.03)
$Time_{ij}$	1.615***	1.586*	1.498*	3.915***	3.915***
	(-5.18)	(-2)	(-2.13)	(-4.6)	(-4.6)
Obs	992	1128	992	1128	1128
Mills λ			-0.298		
			(-1.24)		

Table A32: Results for migrants from the Spain

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
$\ln D_{ij}$	-1.134***	-0.457**	-1.283***	-0.438*	-0.438*
	(-10.18)	(-2.86)	(-14.04)	(-2.5)	(-2.5)
$\ln Mig_{ij}$	0.380***	3.798	0.38	0.657***	
	(4.06)	(0.57)	(0.08)	(9.95)	
$\ln alt\_Mig_{ij}$					0.719***
					(9.96)
Ajd_{ij}	-0.0105	1.300*	0.0382	0.283	0.283
	(-0.04)	(2.27)	(0.19)	(1.45)	(1.45)
$Lang_{ij}$	1.024***	1.900***		0.244	0.244
	(5.52)	(4.94)		(1.66)	(1.66)
FTA_{ij}	0.338	-0.0742	0.288	0.265	0.265
	(1.5)	(-0.29)	(1.63)	(1.07)	(1.07)
$Colony_{ij}$	0.819*	0.23	1.281***	0.591***	0.591***
	(2.63)	(0.4)	(5.35)	(3.34)	(3.34)
$Time_{ij}$	1.114**	1.707**	0.922	1.590***	1.590***
	(-2.93)	(-2.64)	(-1.51)	(-3.56)	(-3.56)
Obs	1008	1178	1008	1178	1178
Mills λ			-0.225		
			(-0.93)		

Table A33: *Results for migrants from the Sweden*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
ln D_{ij}	-1.172***	-0.617***	-1.305***	-0.410*	-0.410*
	(-9.97)	(-3.84)	(-13.87)	(-2.32)	(-2.32)
ln Mig_{ij}	0.227*	10918.7***	-0.661	0.346***	
	(2.27)	(64665.82)	(-0.5)	(3.31)	
ln alt_Mig_{ij}					0.247***
					(3.31)
Ajd_{ij}	-0.035	0.981	0.0286	0.114	0.114
	(-0.15)	(1.68)	(0.14)	(0.69)	(0.69)
$Lang_{ij}$	0.865***	1.329***		0.458**	0.458**
	(4.57)	(3.7)		(3.13)	(3.13)
FTA_{ij}	0.151	-0.156	0.0712	0.414	0.414
	(0.65)	(-0.6)	(0.41)	(1.45)	(1.45)
$Colony_{ij}$	0.867**	0.434	1.297***	0.410*	0.410*
	(3.18)	(1)	(6.12)	(2.39)	(2.39)
$Time_{ij}$	1.567***	1.594*	1.510*	3.360***	3.360***
	(-6.8)	(-2.33)	(-2.36)	(-7.31)	(-7.31)
Obs	952	1105	952	1105	1105
Mills λ			0.0214		
			(0.09)		

Table A34: *Results for migrants from the Switzerland*

	OLS	Heckman Sample Selection Model		PPML	Alt PPML
		1st part	2nd part		
ln D_{ij}	-1.217***	-0.506***	-1.385***	-0.470*	-0.470*
	(-10.2)	(-3.3)	(-14.57)	(-2.43)	(-2.43)
ln Mig_{ij}	0.358***	20.15***	1.305	0.289***	
	(8.64)	(163.64)	(0.38)	(6.15)	
ln alt_Mig_{ij}					0.274***
					(6.15)
Ajd_{ij}	-0.023	1.091	-0.0155	0.142	0.142
	(-0.1)	(1.88)	(-0.07)	(0.92)	(0.92)
$Lang_{ij}$	1.153***	1.402***		0.622***	0.622***
	(6.01)	(4.26)		(3.55)	(3.55)
FTA_{ij}	0.182	-0.162	0.105	0.402	0.402
	(0.81)	(-0.65)	(0.59)	(1.29)	(1.29)
$Colony_{ij}$	0.831**	0.37	1.446***	0.472**	0.472**
	(3.11)	(0.86)	(7)	(2.74)	(2.74)
$Time_{ij}$	1.105**	1.648*	0.915	1.835***	1.835***
	(-2.89)	(-2.49)	(-1.51)	(-4.25)	(-4.25)
Obs	1014	1182	1014	1182	1182
Mills λ			-0.0328		
			(-0.13)		