Migration and Trade Flows: New evidence from Spanish Regions

Department of Economics and Statistics "Cognetti de Martiis"
Campus Luigi Einaudi, Lungo Dora Siena 100/A, 10153 Torino (Italy)
www.est.unito.it

Dipartimento Economia e Statistica Cognetti de Martiis

Working Paper Series

24/17

MIGRATION AND TRADE OWS: NEW EVIDENCE FROM SPANISH REGIONS

ANNA D'AMBROSIO and SANDRO MONTRESOR

UNIVERSITÀ DEGLI STUDI DI TORINO

The Department of Economics and Statistics "Cognetti de Martiis" publishes research papers authored by members and guests of the Department and of its research centers. ISSN: 2039-4004

Migration and trade flows: new evidence from Spanish regions

Anna D'Ambrosio[*], Sandro Montresor[†]

March 2017

Abstract

We analyze migrants' pro-trade effects through a theory-consistent gravity model augmented with migration variables - both immigration and emigration. We take subnational units, i.e. Spanish NUTS3 regions and allow for subnationally heterogeneous multilateral resistance terms, implying diversified exporting capacity of provinces. We implement an econometric strategy based on Head and Mayer (2014), which leads us to selecting the Gamma PML estimator. Comparing the Gamma with OLS estimator we highlight some shortcomings of previous literature. In particular, language commonality is found to magnify the pro-trade effect of immigrants, differently from previous literature; both emigrants' and immigrants' networks are found to exert a positive and significant pro-trade effect, but in different ways: immigrants affect trade through their local networks, whereas emigrants affect trade through their national networks.

JEL codes: *F10, F14, F22, C52*

Keywords: *Gravity model, migration, subnational units, Gamma PML*

1 Introduction

Local economies are facing unprecedented changes due to their integration in the world economy and to international migration. Among the effects of immigration on the local economies, there is the promotion of trade and internationalisation. A branch of the international economics literature pioneered by Gould (1994) has highlighted that immigrants promote the trade of the countries of destination with their origin countries. This bears developmental implications for the host economies: indeed, beyond increasing the imports of home country goods due to "home biased" consumption, migrants have been proven to increase the host economy's exports to their homeland. This less straightforward finding is interpreted as an indication that migrants decrease bilateral trade costs. Trade barriers can be overcome by migrants because their knowledge of the home language and institutions facilitates transnational information flows about

[*]Polytechnic of Turin

[†]Kore University of Enna

trade opportunities, especially in differentiated goods (the "information effect") and because their networks contribute to the enforcement of contracts in weak institutional settings (the "enforcement effect"; Rauch and Trinidade, 2002; Dunlevy, 2006)[1].

In this paper, we contribute to the empirical analysis of the migration-trade link by emphasizing the implications of a regional perspective, hence by allowing regions to differ in their overall exporting capacity. While other studies have already adopted sub-national units (most importantly, Wagner et al., 2002; Herander and Saavedra, 2005; Bandyopadhyay et al., 2008; Briant et al., 2014; Bratti et al., 2014), they have mainly done so to exploit two convenient features of such units. First, using sub-national units we are able to more precisely measure the issue at stake: the social contacts relevant to reducing information costs associated with trade opportunities are more likely to occur within networks of proximity (Rauch, 1999). Second, sub-national units also conveniently increase the variation in the observations available to the analyst. We focus is on the smallest available regional level, i.e. Spanish NUTS3 regions, i.e. provinces, as Peri and Requena-Silvente (2010). The marked regional concentration of migrants confirms our choice: of the 52 Spanish provinces, seven account for a 60% of the total immigrant population (Madrid, Barcelona, Alicante, Valencia, Malaga, Islas Baleares and Murcia), and eight are home to almost 57% of the Spanish expatriates (Madrid, A Corunña, Pontevedra, Barcelona, Asturias, Ourense, Santa Cruz and Lugo) (INE, *Instituto Nacional de Estadistica*; see also Peri and Requena-Silvente, 2010).

Previous studies, however, with the single exception of (Briant et al., 2014), do not endorse a third implication of the regional perspective: regions may differ in their overall exporting capacity. In a standard gravity equation, this amounts to allowing for heterogeneity at the level of the "multilateral resistance term" of regions; the capacity to export to any country in the world - the exporter's *multilateral* resistance term - affects the extent to which a change in *bilateral* trade costs influences *bilateral* trade (Anderson and van Wincoop, 2003). The literature argues that migrants' stocks effectively reduce the informal barriers to bilateral trade: hence, ignoring sub-national heterogeneity may yield biased estimates of the effect of migration on bilateral trade.

In this study, we also extend previous works by analyzing the immigration and emigration sides jointly. From this perspective, we are essentially adding a regional dimension to the approach by Murat and Pistoresi (2009) and Flisi and Murat (2011), who argued that the analysis of the pro-trade effect of immigrants in countries with large diasporas like Spain must be integrated with the emigrants' contributions.

In line with the other studies in this branch of the literature, we apply a gravity model to predict the effect of migration on trade. The law-like behaviour of the gravity model in the case of Spanish provinces is illustrated in Figure 1, where exports of the province of Madrid in 2008 are plotted against distance-weighted GDP of the partner country. The slope of the fitted line is very close to unity.

Our units of analysis are 3039 trading pairs constituted by Spanish provinces and foreign countries in a panel covering the 2006-2010 period. Because of their greater relevance to local development, we focus on exports and disregard imports. As it is common in the migration-trade link literature adopting sub-national units, our data

[1]See Felbermayr et al. (2012) for a recent review of the literature on the migration-trade link; Genc et al. (2011) provide a meta-analysis of the results.

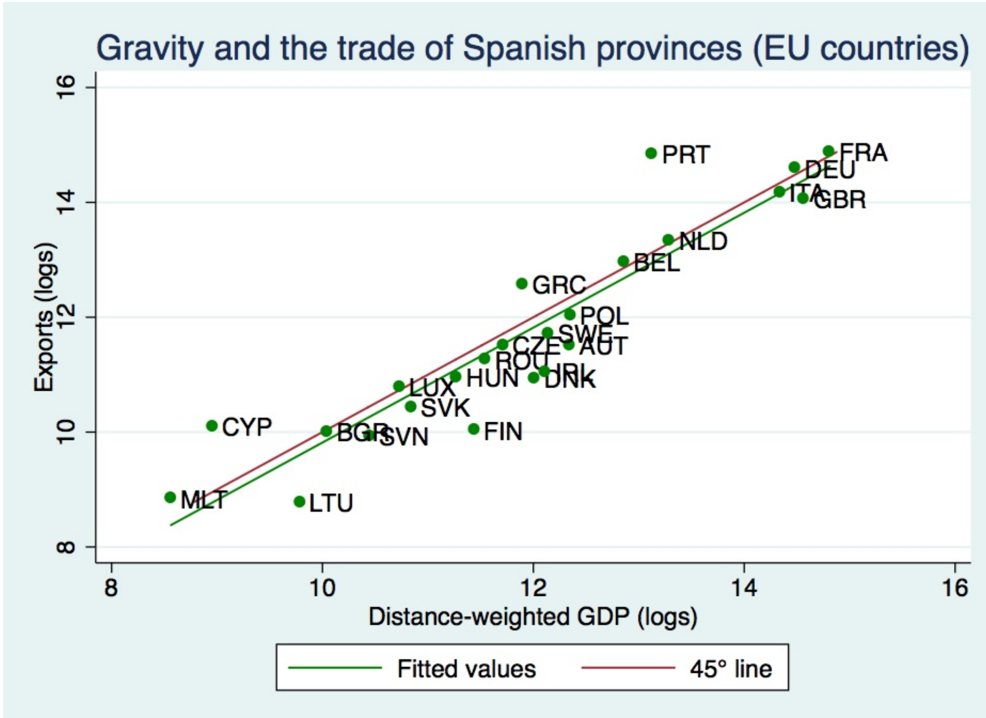

Figure 1: Gravity and the trade of the province of Madrid with EU countries, 2008

is a balanced panel of largely non-zero trade flows. In order to provide an accurate estimate of the pro-trade effect of migrants with these data, the recent developments in the wider gravity literature have to be integrated in the migration-trade link branch. Primarily, as it is by now established, heteroskedasticity in log-linear models leads to violating the assumption of independence of the errors, a problem which can be circumvented if the dependent variable (in levels) is modelled as an exponential function of the covariates (Santos-Silva and Tenreyro, 2006). However, the standard solution of implementing a Poisson PML estimator, suitable to analyze data with many zeros, may not be appropriate for strictly positive data. To select the suitable estimator, we implement an econometric strategy based on Head and Mayer (2014) and on Santos-Silva and Tenreyro (2006): we compare OLS, Poisson PML and Gamma PML estimators and implement diagnostic tests to study the underlying distribution of the errors and potential mis-specification.

The remainder of this paper is organised as follows. In Section 2, we discuss the theoretical framework, and in Section 3 the empirical strategy. In Section 3, we develop the empirical model; in section 4 we describe our data; in section 5 we present our results. Section 6 concludes.

2 Theoretical framework

Gravity models of international trade, in analogy with the Newtonian law on gravity, predict that exports are a function of the economic mass of the trading partners, and a negative function of a bilateral trade cost term (Tinbergen, 1962). Since Anderson and van Wincoop (2003) established the theoretical microfoundations of the gravity model, such "naive" gravity equation was complemented with importer's and emporter's "multilateral resistance terms" (MRT), capturing the average openness to trade of each trading partner. Changes in bilateral trade barriers must not be evaluated in absolute terms but in terms of their relative effect with respect to the multilateral resistance terms (Anderson and van Wincoop, 2003). The "structural" form of the gravity equation (Head and Mayer, 2014)[2] in a cross-sectional context is then

$$X_{ni} = \frac{Y_i}{\Omega_i} \frac{X_n}{\Phi_n} \phi_{ni} \tag{1}$$

Where X_{ni} is the volume of trade between country n (importing country) and country i (exporting country); Y_i represents the "mass" of production of exporting partner, approximated by province gross product in our case; Y_n represents the "mass" of expenditures of the importing country, approximated by GDP; the term ϕ_{ni} represents the transaction costs of bilateral trade; it includes natural trade barriers such as distance but also other barriers such as tariffs, as well as their respective elasticities. Studies on the migration-trade link see immigration as a factor that decreases bilateral trade costs.

The MRT corresponds to $(\Omega_i \Phi_n)^{-1}$. The factors composing the multilateral resistance terms can be interpreted as, respectively, the average market access available to the exporting country (Ω_i) and the average degree of competition in the importing country (Φ_n). More precisely, Ω_i represents the "expenditure-weighted average of relative access" and Φ_n the "accessibility-weighted sum of exporters' capabilities" (Head and Mayer, 2014, : 9-10).

When sub-national units are adopted, the gravity model becomes asymmetrical: in our case, exporters are NUTS3 regions while the importers are countries. This, however, does not introduce a difference in the interpretation of the terms in the equation. Simply, the exporter-side multilateral resistance term should be seen as the region's (weighted) capacity of exporting to any countries of the word, and the importer-side multilateral resistance term should be seen as the average (weighted) market access of a given country to any regions of Spain, as well as to any other exporter wordwide. Hence, applying a standard structural gravity equation, the exporter-side MRT, allowing for sub-nationally differentiated exporting capability, should be included; omitting this term actually implies incurring the "gold medal mistake" in gravity literature (Baldwin and Taglioni, 2007). Nonetheless, most recent empirical studies focussing on sub-national units have omitted such term. More precisely, Peri and Requena-Silvente (2010); Bandyopadhyay et al. (2008) assume the term is constant across regions in

[2]We will hereinafter often refer to their recent review of the gravity literature and use their notation. In an extensive literature review, they have shown that "structural" gravity equations (and their and "general" form) are compatible with the wide majority of trade models used in the literature, including the one in Chaney (2008), on which Peri and Requena-Silvente (2010) base their work.

the same country; Bratti et al. (2014) assume it is constant within the higher NUTS2 level. To the best of our knowledge, only Briant et al. (2014) include such a term but do not provide a specific discussion supporting their choice. Another reason for allowing for this heterogeneity is that the accessibility-weighted exporting capabilities of regions may be affected by the overall supply of immigrant labour which affects productivity, as well as offshoring decisions of firms (e.g. **??**). The overall supply of immigrant labour is in turn likely to be correlated with the supply of labour from a specific country, hence, again, bear implications on the estimated elasticities.

3 Empirical Strategy

Baldwin and Taglioni (2007) notice that, with panel data, the multilateral resistance term depends on time-varying bilateral trade costs and on time-varying economic masses. Thus, it is time-varying itself. Empirically, they argue, the time variation in the multilateral resistance terms should be captured by time-varying importer and time-varying exporter effects, while the correlation between the unobservable component of the bilateral trade determinants and the included trade determinants should be accounted for by time-invariant pair effects. This is the specification that we will pursue in this paper.

The selection of the suitable estimator for gravity models stands at the "frontiers of gravity research" (Head and Mayer, 2014) and, again, has mostly been discussed with reference to cross-sectional data, with high importance attached to predicting zero trade flows. However, most recent empirical works on the migration trade-link using subnational units use panel data, focus on non-zero trade flows and apply OLS estimation (e.g. Peri and Requena-Silvente, 2010; Bandyopadhyay et al., 2008). Hence, they are affected by the estimation issues associated with heteroskedasticity that Santos-Silva and Tenreyro (2006) popularized: when the error term is heteroskedastic, the procedure of log-linearizing the gravity equation and estimating it by OLS introduces a bias in the estimates, as the conditional mean of the log of errors will depend on both their mean and on the higher-order moments of their distribution. A violation of the homoskedasticity assumption will in general lead to the fact that the expected value of the log-linearized error term depends on the covariates, leading to inconsistent OLS estimates. Pseudo-maximum-likelihood (PML) estimators such as the Poisson and the Gamma, in which the dependent variable is in levels and not in logs, are proposed by them as a solution (Poisson PML being their first choice in a cross-sectional context with zero trade flows). Here, instead, the estimation of zero trade flows is not going to be an issue[3].

Our empirical strategy will focus on minimising the implications from heteroskedasticity on the consistency of the estimates, and will apply to this end the diagnostic tests discussed by Head and Mayer (2014) (which, in turn, are based on Santos-Silva and Tenreyro, 2006; Manning and Mullahy, 2001). Based on these tests, we will compare OLS, Gamma and Poisson PML estimators to select the estimator that is most likely

[3]Neither will we address potential endogeneity or reverse causality, considering that a number of recent works (Briant et al., 2014; Peri and Requena-Silvente, 2010; Bratti et al., 2014) have shown that the effect of migration on trade can actually be safely interpreted as causal.

to be consistent and efficient for the data at stake. Due to the type of data we use, the conclusions from these tests are of relevance for the selection of the most suitable estimator for gravity models by balanced panels of non-zero trade flows.

Hence, as regards our empirical model, we will compare the OLS estimates of the log-linearized gravity model in equation 2:

$$ln(X_{nit}) = b_1 ln(X_{nt}) + b_2 ln(Y_{it}) + \beta_1 \ln(\text{Immi}_{nit}+1) + \beta_2 \ln(\text{Emi}_{nit}+1) + NID_{nit} + \\ + NED_{nit} + \gamma_1 \boldsymbol{\theta_{nt}} + \gamma_2 \boldsymbol{\omega_{it}} + \gamma_3 \boldsymbol{\eta_{ni}} + \varepsilon_{nit} \quad (2)$$

with the equivalent model specification in equation 3 to be estimated by Poisson and Gamma PML (Bosquet and Boulhol, 2010):

$$X_{nit} = exp[b_1 ln(X_{nt}) + b_2 ln(Y_{it}) + \beta_1 \ln(\text{Immi}_{nit}+1) + \beta_2 \ln(\text{Emi}_{nit}+1) + NID_{nit} + \\ + NED_{nit} + \gamma_1 \boldsymbol{\theta_{nt}} + \gamma_2 \boldsymbol{\omega_{it}} + \gamma_3 \boldsymbol{\eta_{ni}} + \varepsilon_{nit}] \quad (3)$$

Where, besides the variables that we already defined,

Immi_{nit} = Stock of immigrants from country n living in province i at time t;

Emi_{nit} = Stock of emigrants from province i living in country n at time t;

NID_{nit}="No immigrants dummy", equal to 1 if no immigrants from country n are residing in province i at time t, and zero otherwise;

NED_{nit}="No emigrants dummy", equal to 1 if no immigrants from country n are residing in province i at time t, and zero otherwise;

$\boldsymbol{\theta_{nt}}$ = vector of the importer-time effects, corresponding to country-time dummies;

$\boldsymbol{\omega_{it}}$= vector of the exporter-time effects, corresponding to province-time dummies;

$\boldsymbol{\eta_{ni}}$= vector of the trading-pair specific fixed effects, corresponding to province-country dummies

ε_{nit}=random error term.

Once the suitable estimation method defined, we will test the hypothesis that immigrants and emigrants have a positive and significant effect on the trade of Spanish provinces. In practice, including $ln(X_{nt})$ jointly with $\boldsymbol{\theta_{nt}}$ and $ln(Y_{it})$ together with $\boldsymbol{\omega_{it}}$ gives rise to perfect collinearity. Hence, the income terms are omitted in the empirical estimation of the basic model. They are, instead, included in the models omitting the corresponding fixed effects.

Along with the hypotheses of a positive and significant effect of immigration and emigration stocks on trade, a few "corollaries" deriving from the literature will be tested. We will in particular investigate sources of non-constancy in the elasticity of exports to immigration and emigration: diminishing returns to migration, cultural and institutional similarity, and geographic proximity (see section 5).

4 Data

The database used for the empirical analysis is a balanced panel based on official export data about 50 Spanish provinces[4] (NUTS 3) and 65 destination countries over 5 years (2006-2010). The selection of the countries is driven by the availability of province-level

[4]The provinces of Ceuta and Melilla are excluded due to data availability reasons.

data on immigrant and emigrant stocks, and by whether their share on total Spanish exports is at least 0,1% every year[5]. Overall, the selected countries account for more than 91% of total Spanish international trade for each year of the panel (see Table A.12 in the Appendix for the complete list of countries)[6].

The interpretation of the migration variables is subject to some caveats. Drawing on the literature, we refer to "immigration" in a province as the stock of residents registered in the "Padrón Municipal" (i.e. municipal census) in that province who hold a non-Spanish citizenship (see Table A.13 for a list of data sources). As it is common in migration studies, this is an imperfect measure of immigration as it neglects the portion of foreign-born people that have acquired the nationality of the host country. It also only refers to formally residing people and neglects undocumented immigrants as well as the intra-national mobility that is not registered in changes of residence. Similarly, the emigration variable is imperfect as it refers to the stock of people who have moved their residence outside Spain but are still recorded in the Spanish election registries (as in Flisi and Murat, 2011). These data are uninformative as to the country of birth of these emigrants; thus, one cannot distinguish return migrants from the native Spanish expatriates. Yet, maintaining one's voting rights in Spain implies the persistence of strong ties to Spain. Thus, it seems safe to assume that they reflect the dynamics of the Spanish emigrant population more than the dynamics of return migration. One shortcoming of sub-national level of analysis is that the availability of data on immigrants' characteristics (e.g. skills, employment status and length of stay) at such disaggregation level is severely constrained.

For data availability reasons of the emigrant variable, our panel includes the period of the burst of the global financial crisis, which had an impact on the variables of interest. Over the 1998-2011 period, exports have been growing at an average rate of 6.46%, emigration stocks at an average rate of 4.17%, and immigration stocks have boomed at an average rate of 17,9%. The 2008-2009 crisis period marked a drop in the growth rates of both exports and immigrant stocks. While the exports have rapidly recovered, this period has brought the yearly growth of immigration stocks to stagnation. On the contrary, emigrant stocks have been growing faster since the crisis years on. The extremely high levels of unemployment associated with the crisis in Spain are probably responsible for these changes.

The correlation between the immigration and emigration variables is only 0.10, so the two variables can be assumed to portray quite different phenomena. Indeed the distribution of immigrants and emigrants by provinces follows quite distinct, in some cases opposing, patterns (see Table A.14 in the Appendix, reporting data about immigrants and emigrants distribution across provinces in 2010 and countries of origin); Fig. 2 provides information about the immigrants' and emigrants' countries of origin.

[5] This threshold is motivated by the pragmatic trade-off between the need to maintain tractability in the Poisson and Gamma estimates and the need to account for the highest possible number of countries in empirical estimation. The results of the OLS estimates on the full sample of countries are similar and can be provided upon request.

[6] In order to ensure that the largest possible number of observations was included in the analysis, we included all those dyads for which the panel resulted balanced. This implies that, among the 3039 resulting dyads - leading, over five years, to 15195 observations-, some of the 50 provinces have not been associated with all of the 65 countries.

Figure 2: Top 15 origin countries of immigrants and destination countries of emigrants in Spain, 2010. Source: Own elaboration on INE data.

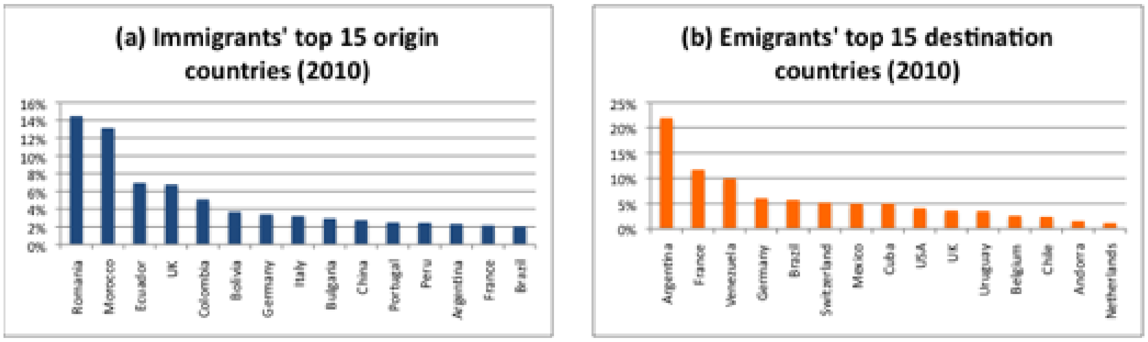

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
X_{nit}	$53,667{\cdot}32$	$243,745$	$0{\cdot}00$	$7,208,594$
$ln(X_{nit})$	$8{\cdot}30$	$2{\cdot}55$	$-6{\cdot}91$	$15{\cdot}79$
$ln(Y_{it})$	$16{\cdot}41$	$0{\cdot}88$	$14{\cdot}472$	$19{\cdot}079$
$ln(X_{nt})$	$19{\cdot}23$	$1{\cdot}65$	$14{\cdot}85$	$23{\cdot}40$
$ln(Immi_{nit}+1)$	$4{\cdot}49$	$2{\cdot}50$	0	$12{\cdot}26$
$ln(Emi_{nit}+1)$	$3{\cdot}03$	$2{\cdot}35$	0	$10{\cdot}75$
NID_{nit}	$0{\cdot}041$	$0{\cdot}20$	0	1
NED_{nit}	$0{\cdot}134$	$0{\cdot}34$	0	1
Observations	$15,195$			

Table 1 reports the summary statistics for main variables of interest of this paper. The correlation is higher between exports and each of the migration variables (respectively, 0.15 with immigration stocks and 0.24 with emigration stocks, than between immigration and emigration stocks. The correlation between province income and immigration is 0.33; between province income and emigration it is 0.17. The correlation between emigration and country income is 0.08; it goes to almost zero (0.001) between immigration and country income.

The distributions of the main variables of interest are characterised by right skew and many small values. This is typical in trade data, and even more so in trade data with sub-national units, with more variation in the data for province-country pairs characterised by larger income and migration stocks. This is a first indication of heteroskedasticity, that will lead to bias if a standard log-linear OLS model is employed; according to Santos-Silva and Tenreyro (2006), the problem may be circumvented by modelling the dependent variable as an exponential function of the covariates specification of the model where the dependent variable is in levels, such as the Poisson PML or the Gamma PML.

The average yearly share of zeros in our explanatory variables is 4.09% for $Immi_{nit}$ and 13.45% for Emi_{nit}. Instead, by construction, the data on exports only have a negligible number of zeros (only 0.24% of all observations). Other works on the pro-trade effects of immigrants using panel data and focussing on non-zero trade flows have employed OLS estimators with fixed effects (for instance, Peri and Requena-Silvente, 2010; Bandyopadhyay et al., 2008); the considerations we made in the above paragraphs, however, suggest there may be additional insights to be gained from enlarging the set of estimators to the PML estimators. The vast majority of strictly positive observations in the dependent variable suggest the Gamma PML may better approximate the distribution of the dependent variable than a Poisson PML (though the latter is still consistent if the conditional mean is correctly specified; see Wooldridge, 2002).

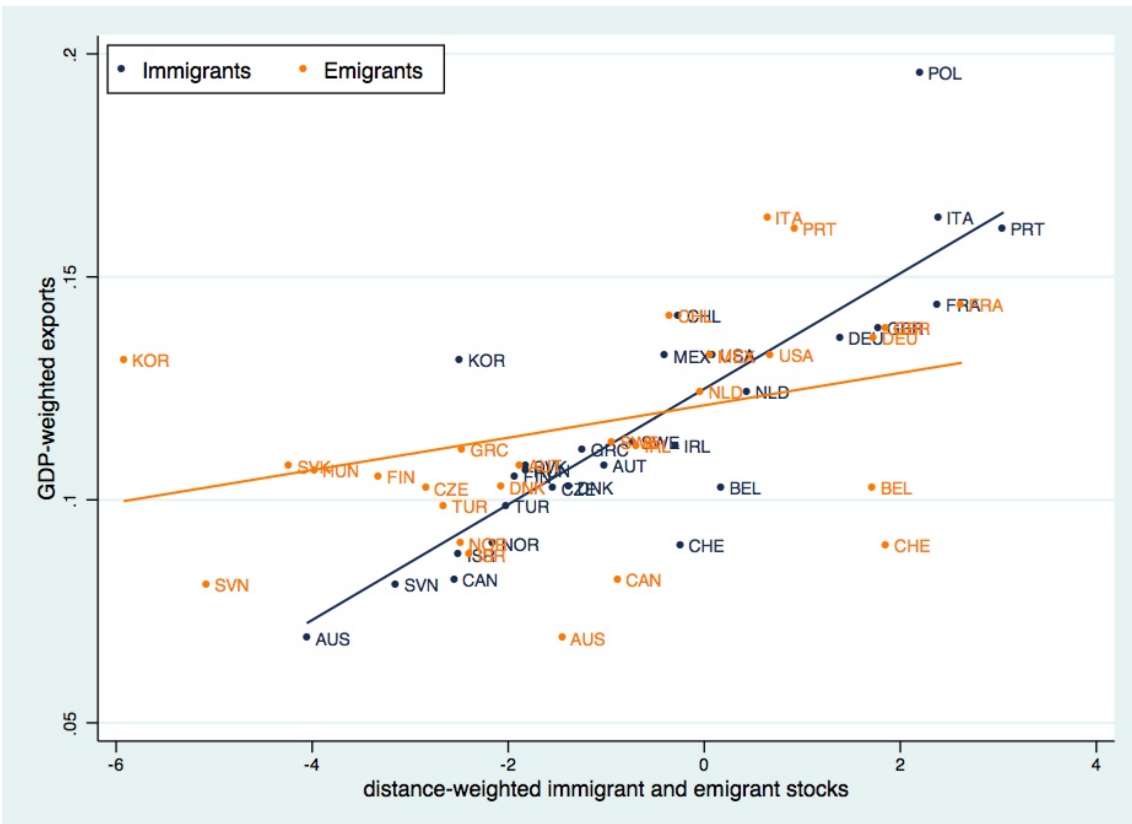

Figure 3: The migration-trade link: province of Madrid and OECD countries, 2010. The graph shows the relationship between GDP-weighted exports and the distance-weighted immigrants and emigrant stocks. Own elaboration on Datacomex, INE, CERA and FMI data.

Within this framework, Figure 3 focuses on the province of Madrid in 2010 and plots the exports-to-GDP ratio against the immigrant and emigrant stocks (weighted by distance) from each country. To avoid other confounding factors, it only portrays

OECD countries; the picture is similar when looking at EU countries. The relationship appears positive for both immigration and emigration, and stronger for immigrants than for emigrants: the province of Madrid trades more with the countries from which it has a larger immigrant or emigrant community. This purely descriptive result motivates a more rigorous econometric analysis of the relationship.

5 Results

5.1 Immigrants' and emigrants' effects on trade

In Table 2 we compare the three estimation methods to address the hypothesis that immigrants and emigrants have a positive effect on the trade of Spanish provinces[7]. The OLS and Gamma estimates show a positive and significant effect of the immigrants' stocks on trade, with magnitudes that are comparable with each other: the OLS estimates show that, by a 10% increase in the immigrant population, trade is expected to grow by on average 1.6%; according to the Gamma estimates, by the same increase in the immigrant stocks, trade will grow by 1.4%, a higher but comparable estimate than the one found by OLS in Peri and Requena-Silvente (2010) on an earlier time period. The Poisson estimates, instead, do not portray any significant trade creation effect by immigrants[8].

As regards the coefficient of $ln(Emi_{nit})$, none of the estimators we implemented yield an estimate that is statistically significantly different from zero. Thus, the hypothesis of a positive role of local networks of emigrants in promoting the trade of Spanish provinces does not find empirical support when looking at the local networks of emigrants. This does not exclude that the flows of information within the emigrants network be mainly determined at the national level (see below).

The inclusion of the full set of time-varying province effects motivated theoretically above is fully statistically supported in the joint tests performed after each estimation method (the p-values from the F tests and likelihood ratio tests are reported in the last line of the table).

Most recent works in the migration-trade link literature with sub-national units, and in particular Peri and Requena-Silvente (2010), use panel data adopting a fixed effects OLS estimator and include time-varying country effects. However, they assume that there is no sub-national heterogeneity in the exporting capacity of provinces (or equivalently, and more in line with their theoretical model based on Chaney (2008),

[7]All estimations were run in Stata, using the `xtpoisson` command for implementing the Poisson PML with fixed effects (as suggested by Santos Silva and Tenreyro) and the `glm` command with `family(gamma)` and `link(log)` options and including pair, country-time and province-time dummies for implementing the Gamma PML.

[8]Similarly to Peri and Requena-Silvente (2010), the coefficient of the no-immigrant dummy (NID) does not result significantly different from zero in the OLS estimates. Instead, it remains positive and statistically significant in all specifications of the Poisson and the Gamma model. According to this result, the pairs with no immigrants would on average trade more than the pairs with at least one immigrant. NID assumes frequently the value of 1 by two main types of countries: those which enjoyed particularly favourable fiscal conditions in the considered time period (e.g. Andorra, Luxembourg, Cyprus) and the very remote countries, so it refers to a quite peculiar set of countries.

Table 2: Estimation results - The effect of immigrants and emigrants on the trade of Spanish provinces

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	$0{\cdot}162$***	$0{\cdot}045$	$0{\cdot}137$***
	$(0{\cdot}061)$	$(0{\cdot}048)$	$(0{\cdot}032)$
$ln(Emi_{nit})$	$-0{\cdot}012$	$0{\cdot}030$	$0{\cdot}005$
	$(0{\cdot}044)$	$(0{\cdot}031)$	$(0{\cdot}026)$
NID	$0{\cdot}176$	$0{\cdot}217$**	$0{\cdot}216$***
	$(0{\cdot}111)$	$(0{\cdot}108)$	$(0{\cdot}065)$
NED	$0{\cdot}030$	$0{\cdot}083$	$0{\cdot}045$
	$(0{\cdot}067)$	$(0{\cdot}063)$	$(0{\cdot}040)$
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	$0{\cdot}119$		
log-likelihood	$-14537{\cdot}1$	$-1{\cdot}54e{+}07$	$-143798{\cdot}1$
AIC	$29994{\cdot}2$	$3{\cdot}08e{+}07$	$288516{\cdot}1$
Joint test on province-time effects			
p-value	$0{\cdot}000$	$0{\cdot}000$	$0{\cdot}000$

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

that there is no subnational variation in wages). Hence, in Table 3 we report the results from a specification which, similar to theirs, restricts all sub-national resistance terms to be equal across provinces, and which includes province income as it is no longer collinear with the (excluded) time-varying province effects lnY_{it}. Comparing this specification with the results of our baseline model in Table 2, we can study whether the restriction that all subnational resistance terms being equal to each other is supported in our data.

While, as already mentioned, the inclusion of the time-varying province effects is supported statistically (see the p-values of the joint tests in Table 2 and compare the AIC in that table with those in Table 3) for all models, excluding it from the specification makes virtually no difference in the OLS estimates of the variables of interest (Table 3). Relying on OLS would suggest that the effect of immigration on trade goes entirely through the bilateral trade costs term and is unrelated with the exporting capacity of the province; the more parsimonious model adopted in Peri and Requena-Silvente (2010) would then be a reasonable simplification.

The Gamma estimate of the immigration effect on trade, instead, is about 8% smaller when the province effects are excluded. This implies some correlation between the bilateral stocks of immigrants and the overall province-level exporting capacity, which may operate through different channels, including - most plausibly in the Chaney

(2008) framework - through the wage channel[9]; also, it implies that it is not necessarily a harmless simplification to assume homogeneous exporting capacity across provinces.

Table 3: Estimation results - Sub-national heterogeneity in the MRT

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	$0{\cdot}161^{***}$	$0{\cdot}049$	$0{\cdot}126^{***}$
	$(0{\cdot}061)$	$(0{\cdot}054)$	$(0{\cdot}032)$
$ln(Emi_{nit})$	$-0{\cdot}001$	$-0{\cdot}005$	$0{\cdot}004$
	$(0{\cdot}044)$	$(0{\cdot}033)$	$(0{\cdot}027)$
$ln(Y_{it})$	$0{\cdot}384$	$0{\cdot}660^{***}$	$0{\cdot}306$
	$(0{\cdot}373)$	$(0{\cdot}202)$	$(0{\cdot}205)$
NID	$0{\cdot}169$	$0{\cdot}173$	$0{\cdot}182^{***}$
	$(0{\cdot}110)$	$(0{\cdot}115)$	$(0{\cdot}065)$
NED	$0{\cdot}028$	$0{\cdot}060$	$0{\cdot}036$
	$(0{\cdot}069)$	$(0{\cdot}067)$	$(0{\cdot}041)$
N	15195	15195	15195
r2	$0{\cdot}077$		
log-likelihood	$-14886{\cdot}3$	$-1{\cdot}71e{+}07$	$-143904{\cdot}7$
AIC	$30302{\cdot}6$	$3{\cdot}42e{+}07$	$294417{\cdot}3$
Trading pair effects	Yes	Yes	Yes
Province-time effects	No	No	No
Country-time effects	Yes	Yes	Yes

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

5.2 Selection of the estimator

In this section we will discuss the estimation issues affecting the estimators and select the estimator that best suits the data according to the procedure described in Head and Mayer (2014), including the analysis of the distribution of the residuals, and a set of Ramsey RESET tests.

A priori, a scenario in which the two PML estimates, both consistent if the conditional mean is correctly specified (Wooldridge, 2002), display different results while the OLS and the Gamma PML estimates are similar, points to mis-specification in the elasticities (Head and Mayer, 2014). According to the simulations in Head and Mayer (2014), wrongly assuming constant elasticities when the elasticity is actually non-constant leads the Poisson PML to be biased because it gives more weight to larger observations in the trade variable. The Gamma PML and the OLS, giving more

[9]The negative relation between the bilateral immigration stocks and the overall exporting capacity implied by the direction of the omitted variable bias is simply another way to look at the MRT issue as defined in Anderson and van Wincoop (2003): bilateral resistance factors, such as immigration stocks, affect trade *relative* to the MRT.

weight to smaller observations, yield estimates of the marginal effects that are closer to the true average effects even by mis-specification of the elasticity.

Hence, because the gravity model *per se* is unlikely to be wrongly specified, we should expect to find non-constant effects of the migration variables, an issue that we will explore in the next section. It is important to note that, with the exceptions of the earlier works by Gould (1994) and Wagner et al. (2002) having modelled diminishing returns from migration, the majority of empirical works on the pro-trade effects of immigration assume a linear relationship between the log of the immigration stocks and the log of trade. However, many studies have identified interactions of the immigrants' stocks with relevant determinants of trade, which are also a cause of non-linearity and may ultimately represent an argument against the use of Poisson PML estimators for the research question at stake. Finally, by construction, our data exclude zero trade flows and do not meet the distributional assumptions of the Poisson distribution. Thus, the Gamma PML seems a better candidate to consistently estimate the effect of immigration when there is heteroskedasticity, non-constant immigration effects and no zero trade flows.

The large number of fixed effects implied by our theory consistent specification imposes a heavy computational burden to the estimation. A way to partially circumvent it is the "tetrad" approach to gravity modelling proposed in Head et al. (2010), that allows algebraic simplification of 575 importer-time and exporter-time effects by computing ratios of ratios of both the dependent variable and the regressors with respect to a reference province and a reference country. We applied the "tetrad" approach to the Gamma (and Poisson) estimation in Table 4, along with Head et al. (2010) original application of the tetrad method to OLS[10]. The Gamma PML estimates for the coefficients of the variables of interest are virtually unaffected (yet, a large number of time-invariant dummies is still included in the model).

Turning to the diagnostic test, we perform a Park test for heteroskedasticity. The results are reported in Table 5. As it turns out, the variance of the residuals significantly increases in the size of the immigration stocks. Hence, heteroskedasticity is at play and the OLS estimates will be biased upwards.

In order to retrieve the underlying distribution of the errors and to select the most efficient estimator, we are interested in analyzing the relation between the variance and the conditional mean (Head and Mayer, 2014):

$$\text{var}[X_{ni}|\mathbf{z}_{ni}] = h\text{E}[X_{ni}|\mathbf{z}_{ni}]^{\lambda} \tag{4}$$

Where $\mathbf{z}_{ni}$ is the vector of covariates. This relation can be estimated empirically as

[10]Because of the repeated use of reference country and province, the errors are correlated across observations. Hence, standard errors must be clustered multi-way at the level of the pair, of the importer-year and of the exporter-year, as argued by Head et al. (2010); their web appendix provides the code for the implementation in Stata, which employs the method developed by Cameron et al. (2011). To the best of our knowledge, there are no statistical packages allowing to estimate Gamma and Poisson regressions with multi-way clustering, so we used clustering at the pair level for the Gamma PML and robust standard errors for the Poisson PML estimator. As regards the "tetraded" OLS, the estimates with multi-way clustering and with clustering at the pair level turn out to have very similar point estimates and standard errors (OLS results with pair-level clustering are available upon request).

Table 4: Estimation results - "Tetrads" method

Model	OLS (Head et al., 2010)	PPML	GammaPML
$ln(Immi_{nit})$	$0{\cdot}187{*}{*}{*}$	$-0{\cdot}122$	$0{\cdot}140{*}{*}{*}$
	$(0{\cdot}053)$	$(0{\cdot}103)$	$(0{\cdot}031)$
$ln(Emi_{nit})$	$-0{\cdot}0229$	$-0{\cdot}127$	$0{\cdot}038$
	$(0{\cdot}052)$	$(0{\cdot}110)$	$(0{\cdot}027)$
NID	$0{\cdot}244{*}$	$0{\cdot}179$	$0{\cdot}194{*}{*}{*}$
	$({\cdot}114)$	$(0{\cdot}175)$	$(0{\cdot}066)$
NED	$0{\cdot}016$	$-0{\cdot}102$	$0{\cdot}051$
	$({\cdot}069)$	$(0{\cdot}138)$	$(0{\cdot}044)$
Trading pair effects	Yes°	Yes	Yes
Province-time effects	Yes°	No	No
Country-time effects	Yes°	No	No
Year dummies	Yes	Yes	Yes
Constant	$-1{\cdot}02e{-}09$		$13{\cdot}843{*}{*}{*}$
	$({\cdot}018)$		$(4{\cdot}077)$
N	14625	15195	15195

Note: Reference importer is France, reference exporter is the province of Madrid. All dependent and independent variables are "tetraded" and demeaned with respect to the reference importer and exporter (see Head and Mayer, 2014; Head et al., 2010). Multi-way clustered standard errors in parentheses in column "OLS (Head et al., 2010)"; standard errors are clustered at the pair level in column "GammaPML". Robust standard errors in column "PPML".
° All variables tetraded and de-meaned by pair, which is equivalent to including the three sets of effects.
* p<0.1, ** p<0.05, *** p<0.01

follows:

$$\ln\hat{\epsilon}^2_{ni} = \text{constant} + \lambda\widehat{\ln X}_{ni} \tag{5}$$

Which is the test implemented by Manning and Mullahy (2001) and Santos-Silva and Tenreyro (2006) to select the most efficient estimator (we will call it "MaMu test" as Head and Mayer, 2014). The coefficient of interest in the MaMu test is the λ in equation 5. If the λ is close to 2, this reflects a constant coefficient of variation, which is compatible with the Gamma distributional assumptions and with a log-normal distribution. The most efficient estimators in this case are the homoskedastic OLS on logs, which is the MLE if the homoskedasticity assumption is reasonable, and the Gamma PML. This explains why the Gamma and OLS estimates are quite similar. If λ is significantly below 2, the Poisson PML should be preferred; the Poisson distributional assumptions can be generalised to correspond to a λ of 1 (Manning and Mullahy, 2001).

The results of the test are reported in Table 6. Regressing the log of the squared residuals on the log of the fitted values of the OLS regression, the estimate for λ is 1.56. Because, however, the OLS estimates for the MaMu test may be affected

Table 5: Park test: linear regression of the log of the squared residuals on the covariates

Model	OLS residuals	Poisson PML residuals	Gamma PML residuals
$ln(Immi_{nit})$	$0{\cdot}485^{***}$	$0{\cdot}312**$	$0{\cdot}354^{***}$
	$(0{\cdot}115)$	$(0{\cdot}126)$	$(0{\cdot}095)$
$ln(Emi_{nit})$	$-0{\cdot}190$	$-0{\cdot}033$	$-0{\cdot}038$
	$(0{\cdot}119)$	$(0{\cdot}091)$	$(0{\cdot}076)$
NID	$0{\cdot}636^{***}$	$0{\cdot}367$	$0{\cdot}465**$
	$(0{\cdot}222)$	$(0{\cdot}233)$	$(0{\cdot}186)$
NED	$-0{\cdot}0322$	$0{\cdot}081$	$0{\cdot}065$
	$(0{\cdot}160)$	$(0{\cdot}142)$	$(0{\cdot}119)$
Constant	$10{\cdot}93 * **$	$14{\cdot}476^{***}$	$25{\cdot}821^{***}$
	$(0{\cdot}728)$	$(0{\cdot}731)$	$(0{\cdot}636)$
N	15195	15195	15195
r2	$0{\cdot}0970$	$0{\cdot}114$	$0{\cdot}936$

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

by heteroskedasticity just like those of the gravity regressions, Manning and Mullahy (2001) advise to rely on the PML estimates for λ. The coefficients for λ estimated by Poisson and Gamma PML are in both cases very close to 2. Because we saw that heteroskedasticity is at play, the MaMu test leads to selecting the Gamma PML estimator as the most efficient estimator.

Table 6: Manning and Mullahy test on the underlying distribution of the errors

Model	OLS residuals	Poisson PML residuals	Gamma PML residuals
$ln(\hat{\mu})$	$1{\cdot}562^{***}$	$1{\cdot}981^{***}$	$2{\cdot}123^{***}$
	$({\cdot}008)$	$(0{\cdot}130)$	$(0{\cdot}006)$
Constant	$-0{\cdot}486^{***}$	$15{\cdot}728^{***}$	$-1{\cdot}405^{***}$
	$({\cdot}0725)$	$(0{\cdot}066)$	$(0{\cdot}059)$
N	15195	15195	15195
r2	$0{\cdot}702$	$0{\cdot}014$	$0{\cdot}922$

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

To detect functional form mis-specification and non-constancy in the covariates, as discussed above, in Table 7 we report the coefficients and p-values resulting from Ramsey (1969) RESET-tests on each estimation method, where, differently from Santos-Silva and Tenreyro (2006), we also include the cube of the fitted values. The joint tests indicate functional form mis-specification for both the Poisson and OLS models. The p-values of the tests for the Gamma PML are extremely small. However, they are associated with coefficients for the square and cube of the fitted values that are so

Table 7: RESET tests on the estimation methods

Model	OLS model	PPML model	GammaPML
Square of the fitted values	$-\cdot 241$	$-0\cdot 1103$	$-1\cdot 33e-13$
P-value	0·0004	0·1455	0·000
Cube of the fitted values	·0067	0·0107	$1\cdot 42e-20$
P-value	0·0137	0·4352	0·000
Joint p-value	0·0000	0·0092	·

close to zero that they actually support the interpretation that the Gamma regression is free from specification errors.

The results of the tests, thus, unambiguously lead to selecting the Gamma PML as our preferred estimator.

Based on the above considerations, we will rely on the Gamma estimates to test hypotheses in the next sections. For the sake of comparison, however, we will also report the results of the estimates for OLS and Poisson PML. As implied by our diagnostic test, the standard errors associated with the Gamma PML estimator turn out to be always smaller than those associated with the Poisson PML and of with the OLS.

5.3 Sources of non-constancy in the elasticities of trade to immigration and emigration

In what follows, we will explore the non-linearity that may drive the divergence in the Poisson and Gamma PML estimates, and we will compare the results of the Gamma PML with those of the OLS, in order to show the implications of applying a Gamma PML estimator with respect to a more standard OLS estimator and to learn about the effects of institutional and language similarity and geography.

The most obvious form of non-linearity which may affect estimates is a non-linear function of the variables of interest. This would be a way to test whether there are increasing or diminishing returns from immigration on trade (Gould, 1994; Wagner et al., 2002). However, including the squared terms of $ln(Immi_{nit})$ and $ln(Emi_{nit})$ in the model has the only effect of reducing the efficiency of the estimates through multicollinearity, while the squared terms result both economically and statistically insignificant. The (unreported, but available) results are similar when including a cubic term.

5.3.1 Institutional similarity and language commonality

The hypotheses that the immigrants' and emigrants' effect on the trade of Spanish provinces is stronger with more institutionally and culturally distant countries are tested in Tables 9a and 9b. As in Girma and Yu (2002) and Blanes-Cristóbal (2008), this implies interacting $ln(Immi_{nit})$ and $ln(Emi_{nit})$ with, respectively, a D_{EU} dummy (equal to 1 for EU Member States and zero otherwise) and a D_{Spa} dummy (equal to 1 when the country has Spanish as an official language and zero otherwise).

Table 8: Regression results - Non-linearity in migration

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	0·18*	0·11	0·10*
	(0·11)	(0·08)	(0·06)
$ln(Immi_{nit})^2$	−0·00	−0·01	0·01
	(0·02)	(0·01)	(0·01)
$ln(Emi_{nit})$	−0·03	−0·04	0·01
	(0·08)	(0·06)	(0·05)
$ln(Emi_{nit})^2$	0·00	0·01	−0·00
	(0·01)	(0·01)	(0·01)
NID	0·19	0·27 **	0·19 **
	(0·13)	(0·12)	(0·07)
NED	0·02	0·03	0·05
	(0·08)	(0·07)	(0·05)
N	15195	15195	15195
r2	0·12		

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

In Table 9a, only the immigrants from extra-EU countries result significant in increasing the trade of Spanish provinces in the Gamma and OLS estimates. This result is consistent with the literature (e.g. Girma and Yu, 2002) and with the interpretation that immigrants are brokers of the flow of communication and facilitators of the enforcement of contracts in international transactions mainly with countries that do not share the same institutional setting. The reason is that sharing the institutional setting or a regional trade agreement (RTA), as well as a common language, are factors that increase bilateral trade *per se*, independently from the immigrant population; the direct effect of these determinants is absorbed in the fixed effects. This is shown more explicitly in Table 10, where the estimated pair-specific fixed effects from the OLS regression are regressed on a series of traditional gravity determinants (this procedure is suggested in Cheng and Wall, 2005). From this regression, Spain results to trade on average 23% more ($[exp(0.207) - 1] * 100$) with EU countries and countries of the European Economic Area (EEA), and 88% more ($[exp(0.631) - 1] * 100$) with OECD countries.

As regards language commonality, D_{Spa} has a coefficient of 0.761: Spanish provinces trade on average about 114% more ($[exp(0.761) - 1] * 100$) with Spanish-speaking countries, irrespective of the immigrants that they host from these countries. The literature would predict the effect of immigration to be redundant. Instead, the Gamma estimates in Table 9b show that immigrants from Spanish-speaking countries further increase trade with their origin countries to a much greater extent - 82% more - than do the immigrants from non-Spanish speaking countries.

Hence, language commonality and institutional similarity cannot be viewed as two

Table 9: Regression results: cultural and institutional similarity

(a) Regression results - Institutional similarity: EU countries

Model	OLS	PPML	GammaPML
$ln(Immi_{nit}^{EU})$	0·123	−0·013	0·080
	(0·094)	(0·071)	(0·051)
$ln(Immi_{nit}^{NEU})$	0·173**	0·088	0·155***
	(0·071)	(0·064)	(0·038)
$ln(Emi_{nit}^{EU})$	0·001	−0·036	0·036
	(0·074)	(0·052)	(0·046)
$ln(Emi_{nit}^{NEU})$	−0·015	0·064*	−0·004
	(0·049)	(0·034)	(0·029)
NID	0·173	0·228**	0·211***
	(0·111)	(0·109)	(0·065)
NED	0·033	0·077	0·054
	(0·069)	(0·065)	(0·041)
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	0·119		

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

(b) Regression results - Language commonality

Model	OLS	PPML	GammaPML
$ln(Immi_{nit}^{Spa})$	0·234	−0·038	0·235**
	(0·216)	(0·235)	(0·109)
$ln(Immi_{nit}^{NSpa})$	0·156**	0·049	0·129***
	(0·064)	(0·049)	(0·033)
$ln(Emi_{nit}^{Spa})$	−0·002	0·074	−0·026
	(0·056)	(0·121)	(0·036)
$ln(Emi_{nit}^{NSpa})$	−0·015	0·029	0·017
	(0·055)	(0·032)	(0·033)
NID	0·173	0·221**	0·211***
	(0·113)	(0·108)	(0·065)
NED	0·027	0·082	0·054
	(0·071)	(0·063)	(0·043)
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	0·119		

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

sides of a somewhat redundant effect of immigration on trade. Integration within the EU results to effectively decrease the barriers to trade, making the effect of immigration redundant. Instead, language commonality activates the capacity of immigrants to promote trade, contributing to bridging remaining hindrances to information flows and trade. Notice that relying on the OLS estimates would have led to opposite conclusions, even if the magnitude of $ln(Immi_{nit}^{Spa})$ estimated by OLS is very similar to the Gamma estimate. This result can be explained within the random encounter model proposed by Wagner et al. (2002), modelling the probability that, given a set of realisable trade opportunities, the immigrant actually realises them: commonality of language gives the immigrant easier access to information on trading opportunities not just in the origin country but also in the host country, and increases her capacity to successfully realise the trading opportunity. In short, language commonality increases the probability that an immigrant has the capacity to facilitate the exchange.

As regards emigration, coherently with the findings in Table 2, no emigrant variable results statistically to determine the level of trade.

Table 10: Determinants of the fixed effects (OLS estimates)

$ln(Y_i)$	1·292***
	(0·014)
$ln(Y_j)$	0·642***
	(0·011)
$ln(DIST)$	−1·051***
	(0·030)
D_{colo_tie}	−0·137
	(0·150)
D_{Spa}	0·761***
	(0·148)
D_{common_border}	0·890***
	(0·065)
D_{EUEEA}	0·207***
	(0·046)
D_{OECD}	0·631***
	(0·043)
Constant	−25·598***
	(0·340)
N	3039
r2	0·516

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

5.3.2 Geographic proximity

In this section we will test Herander and Saavedra (2005)'s hypothesis that local networks of immigrants from the same province have a stronger effect on the trade of provinces than more distant networks, and will extend it to the analysis of emigrants' networks. In particular, we will distinguish the effect on trade of immigrants residing in the province from those residing outside the province by adding two additional variables $ln(ImmiOut_{nit})$ and $ln(EmiOut_{nit})$. $ln(ImmiOut_{nit})$ represents the total stock of immigrants from country n living in provinces other than i at time t, and $ln(EmiOut_{nit})$ represents the total stock of emigrants registered in provinces other than i who had migrated to country n at time t. They are meant to represent national networks of immigrants and emigrants that extend beyond the province.

Table 11 reports the results of the regression. They confirm the hypothesis that it is local networks of immigrants, rather than more far-reaching networks, that determine their trade-facilitation effect. The flows of information relevant to trade creation by emigrants, instead, are to be found at the level of nation-wide networks of expatriates; as in the specification about language commonality, the OLS yields a large point estimate but not a statistically significant effect and would have led to wrong inference. From the point of view of the emigrants, their effect on trade is an import-facilitating effect. Hence, taste effects could be at play as well as information effects and a distinction between the two is unfortunately impossible with the data at stake. Whatever the mechanism, this is not specific of the province but of their host country. Just to make an example, to promote trade between the province of Alicante and China, a larger network of emigrants from any Spanish province is more effective than a large network of natives from Alicante residing in China. If this can be attributed to a taste effect, demand for Alicante goods in China is not specific to natives from Alicante but to the whole expatriated Spanish community in China; if this can be attributed to an information effect, trade promotion depends more on the knowledge of the Chinese market than on the knowledge of the Alicante products.

More generally, these results suggest that immigration and emigration networks operate through different dynamics[11].

6 Conclusions

This study has analyzed the effects of immigration and emigration on the trade of Spanish provinces applying a theory-consistent gravity model. It integrates contributions to the literature on the migration-trade link at the methodological level with theoretical arguments in support of the inclusion of controls for sub-national heterogeneity in the

[11] One could approach the issue differently and argue that any migrant linkages, in one direction or another, promote trade, or, alternatively, that the net change in the stock of individuals able to create a bridge between the two countries matters more to trade than the stocks of immigrants and emigrants separately. We tested these hypotheses in a series of unreported regressions, where the migration variables were linearly combined. In all cases, the explanatory power of the joint variable (either adding the two stocks or computing the net effect of the two) is lower than that of the two variables separately. Comparison of the related AIC strongly support the interpretation that the two variables operate through different dynamics and no specific insight is drawn from their combination.

Table 11: Regression results - Geographic proximity

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	0·154 ∗ ∗	0·045	0·144∗∗∗
	(0·064)	(0·048)	(0·032)
$ln(ImmiOut_{nit})$	−1·117	−0·544∗∗∗	−0·082
	(1·307)	(0·190)	(0·065)
$ln(Emi_{nit})$	−0·002	0·013	0·025
	(0·051)	(0·035)	(0·026)
$ln(EmiOut_{nit})$	0·524	−0·429	0·279∗∗∗
	(0·936)	(0·379)	(0·057)
NID	0·170	0·218 ∗ ∗	0·189∗∗∗
	(0·112)	(0·109)	(0·065)
NED	0·034	0·067	0·051
	(0·068)	(0·062)	(0·039)
Constant	14·490		10·635∗∗∗
	(14·577)		(0·956)
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	0·119		

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

multilateral resistance terms; the resulting empirical estimates provide insights that are partially confirming the existing literature and partially extending it.

From the theoretical point of view, the model used to analyze the effect of the migration-trade link allowed for province-level heterogeneity in the multilateral resistance terms besides trading-pair time-invariant fixed effects and time-varying country-level effects. While the inclusion of time-varying exporter effects is an obvious implication of the gravity model, it has been often neglected in empirical studies on the migration-trade link that adopt sub-national units. The rationale for including such controls has been discussed and statistical tests have showed that they contribute to a better fit of the model.

From the methodological point of view, the application of our empirical strategy led to identifying the Gamma estimator as the most suited estimator for the data at stake on both efficiency and consistency grounds. The OLS estimator was discarded on grounds of heteroskedasticity, which leads to bias in log-linear models; throughout the different specifications, however, the magnitudes of the OLS and Gamma PML estimates have resulted comparable with each other, with the Gamma PML generally outperforming the OLS in terms of efficiency of the estimates and the OLS estimates exceeding those of the Gamma by between 10% and 35%, a difference attributable to the bias. By construction, and similarly to other papers in the literature on the immigrants' pro-trade effects using panel data, our database only includes a negligible number of zeros. Hence, it is not very surprising that the Poisson distributional assumptions appeared restrictive and the Poisson PML led to very inefficient estimates that were hardly of any usefulness in addressing the research question. The Poisson estimates often diverged from those yielded by the more flexible Gamma regression. The reason for this divergence was primarily found in functional form mis-specification due to non-constancy in the immigration and emigration effects for different levels of institutional and language similarity. This non-constancy could account for the worse performance of the Poisson estimator because the Poisson estimator gives more weight to larger observations, i.e. larger trade flows (as found in Head and Mayer, 2014): because Spanish provinces trade mostly with EU countries, this implies high institutional similarity and no language commonality, hence a small effect of immigrants. Hence, the application of the Poisson PML is not advisable to the analysis of the research question at stake.

The methodology adopted in this study represents an application to panel data of the empirical strategy proposed in Head and Mayer (2014) and is *per se* novel: it applies a quite recent methodology standing at the "frontier of gravity research" (Head and Mayer, 2014) to the analysis panel data. Comparing the results obtained by applying Gamma regression with those obtained by OLS and Poisson, it becomes clear that the application of a sound methodology has important implications on the findings for the research question at stake, i.e. whether immigrants and emigrants have an effect on trade.

As regards the main empirical findings, overall, the Gamma (and OLS) estimators robustly confirm a positive effect of immigrants on trade. The identified magnitude of the effect ranges between 0.126 and 0.144, implying that a 10% increase in the immigrant population from a given country in a given province would increase its exports to the origin country by between 1.26% and 1.4%. If we only looked at the OLS estimates, we would conclude that including time-varying exporter effect has only a minor impact on the magnitude of the immigrants' and emigrants' elasticities on the aggregate sample estimated by OLS; instead, inclusion of these effects decreases the Gamma estimates by 8%, suggesting that the immigration effect and the exporting capacity of the province are correlated.

The effect of immigration also results stronger in the trade with more institutionally distant countries, i.e. with non-EU countries. This confirms the findings of the literature on the migration-trade link in this regard: the integration of Member States within the European Union *per se* increases trade and immigrants don't play a significant role in this regard; immigrants, instead, play a relatively large and statistically significant role in promoting trade towards these countries with which there are no institutionalized trade agreements. Hence, immigrant stocks contribute to realizing potential trade opportunities by decreasing the impact of informal trade barriers that do not seem to apply to the trade with EU countries.

The results in terms of language commonality, instead, suggest that, while the value of trade with Spanish-speaking countries is much higher *per se* than the trade with other countries, immigrants from Spanish-speaking countries have a magnifying effect and further increase trade. This result suggests that language commonality is a factor that increases the immigrants' capacity to promote trade. Among the potential trade opportunities that an immigrant could facilitate, some could be lost due to language differences, which would have the effect of reducing the immigrant capacity to promote trade (cfr. the random encounter model in Wagner et al., 2002). An implication of this finding would be that promoting the knowledge of the host country language among the immigrant communities could, among other, contribute to realizing potential opportunities for exports, facilitating in particular the realization of trade opportunities with strategic emerging economies with which there are currently no institutional arrangements.

The analysis also confirmed that it is mainly localized, rather than more far-reaching networks of immigrants, that result relevant in the promotion of the trade of provinces and, thus, that immigration is an issue of relevance for local production systems. In addition to this, in a series of regressions that are not reported here for space reasons, the analysis showed that there is marked sub-national variation in the effects of immigration and emigration, which implies, *coeteris paribus*, differentiated capacity of the local systems to enable the immigrants' potential to promote trade.

The networks of expatriates, instead, appear to affect trade through different mechanisms, with a strong role of nation-wide networks and negligible effects of the local networks: this result means that what matters to trade is the existence of a network of Spanish expatriates in the same country, irrespective of their provinces of origin, that enables emigrants to promote imports from their origin provinces. The magnitude of this effect, implying that a 10% increase in the emigrant population to a country would increase its exports to the destination country by 2.7%, is aligned with the elasticities identified for imports in the meta-analysis carried out by Genc et al. (2011) on the migration-trade link literature and is comparatively larger than the effect of immigrants in promoting exports. This results bears slightly optimistic implications about the fact that the massive expatriation from Spain in the years of the crisis has not only been a way to escape unemployment but has also contributed to create trade opportunities.

Bibliography

Anderson, J. and E. van Wincoop (2003). Gravity with *gravitas*: a solution to the border puzzle. *American Economic Review 93*, 170—-92.

Baldwin, R. and D. Taglioni (2007). Gravity for dummies and dummies for gravity equations. NBER working paper series, Working Paper 12516, `http://www.nber.org/papers/w12516`.

Bandyopadhyay, S., C. Coughlin, and H. Wall (2008). Ethnic networks and US exports. *Review of International Economics 16*(1), 199–213.

Blanes-Cristóbal, J. (2008). Characteristics of immigrants and bilateral trade. *Revista de Economía Aplicada XVI*(48), 133–159.

Bosquet, C. and H. Boulhol (2010, November). Scale-dependence of the Negative Binomial Pseudo-Maximum Likelihood Estimator. Documents de travail du Centre d'Economie de la Sorbonne 10092, Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne.

Bratti, M., L. De Benedictis, and G. Santoni (2014). On the pro–trade effects of immigrants. *Review of World Economics 150*, 557–594.

Briant, A., P.-P. Combes, and M. Lafourcade (2014, 01). Product Complexity, Quality of Institutions and the Protrade Effect of Immigrants. *The World Economy 37*(1), 63–85.

Cameron, A., J. Gelbach, and D. Miller (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Chaney, T. (2008). Distorted gravity: Heterogeneous firms, market structure and the geography of international trade. *American Economic Review 98*, 1707—21.

Cheng, I. and H. Wall (2005). Controlling for heterogeneity in gravity models of trade and integration. *Federal Reserve Bank of St. Louis Review 87*, 49–63.

Dunlevy, J. (2006). The impact of corruption and language on the pro-trade effect of immigrants: Evidence from the American States. *Review of Economics and Statistics 88*(1), 182—-186.

Felbermayr, G., V. Grossmann, and W. Kohler (2012, October). Migration, International Trade and Capital Formation: Cause or Effect? IZA Discussion Papers 6975, Institute for the Study of Labor (IZA).

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

Genc, M., M. Gheasi, P. Nijkamp, and J. Poot (2011, November). The Impact of Immigration on International Trade: A Meta-Analysis. IZA Discussion Papers 6145, Institute for the Study of Labor (IZA).

Girma, S. and Z. Yu (2002). The link between immigration and trade: Evidence from the United Kingdom. *Weltwirtschaftliches Archiv 138*, 115–30.

Gould, D. M. (1994). Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. *The Review of Economics and Statistics 76*(2), 302–316.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*. Elsevier.

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

Herander, M. and L. Saavedra (2005, May). Exports and the structure of immigrant-based networks: The role of geographic proximity. *The Review of Economics and Statistics 87*(2), 323–335.

Manning, W. and J. Mullahy (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics 20*(4), 461–494.

Murat, M. and B. Pistoresi (2009). Migrant networks: Empirical implications for the Italian bilateral trade. *International Economic Journal 23*(3), 371–390.

Peri, G. and F. Requena-Silvente (2010). The trade creation effect of immigrants: Evidence from the remarkable case of Spain. *Canadian Journal of Economics 43*(4), 1433–1459.

Ramsey, J. (1969). Tests for specification errors in classical linear least-squares analysis. *Journal of the Royal Statistical Association 71*, 350–371.

Rauch, J. (1999). Networks versus markets in international trade. *Journal of International Economics 48*(1), 7–35.

Rauch, J. and V. Trinidade (2002). Ethnic Chinese networks in international trade. *Review of Economics and Statistics 84*(1), 116–30.

Santos-Silva, J. and S. Tenreyro (2006). The Log of Gravity. *Review of Economics and Statistics 88*, 641–58.

Tinbergen, J. (1962). *Shaping the World Economy; Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

Wagner, D., K. Head, and J. Ries (2002). Immigration and the trade of provinces. *Scottish Journal of Political Economy 49*(5), 507—525.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge (MA): MIT Press.

A Appendix

Table A.12: **List of Countries**

Europe	Americas	Asia	Africa	Transition economies	Oceania
Andorra	Argentina	Bangladesh	Algeria	Bulgaria	Australia
Austria	Brazil	China	Angola	Croatia	
Belgium	Chile	India	Dem. Rep. of Congo	Czech Republic	
Cyprus	Dominican Rep.	Iran	Egypt	Hungary	
Denmark	Canada	Israel	Guinea	Lithuania	
Finland	Colombia	Lebanon	Morocco	Poland	
France	Cuba	Japan	Nigeria	Romania	
Germany	Mexico	Pakistan	Syria	Russia	
Greece	Panama	Philippines	South Africa	Slovakia	
Ireland	Peru	South Korea	Tunisia	Slovenia	
Italy	USA			Ukraine	
Luxembourg	Venezuela				
Malta					
The Netherlands					
Norway					
Portugal					
Sweden					
Switzerland					
Turkey					
UK					

Table A.14: **Total population, immigrants and emigrants by province (2010)**

Province	Population	Nationality		Province pop. share		Immigrant pop. share	Emigrants	Province pop. share		Emigrant pop. share
	(Persons)	*Spanish*	*Foreigners*	%	*Level*	%		%	*Level*	%
SPAIN	47,021,031	41,273,297	5,747,734	12.2	-	100	1,408,825	3.0	-	100
Alicante	1,926,285	1,459,186	467,099	24.2	high	8.1	21,371	1.1	low	1.5
Balears	1,106,049	863,793	242,256	21.9	high	4.2	14,328	1.3	low	1.0
Almería	695,560	544,401	151,159	21.7	high	2.6	27,772	4.0	high	2.0
Girona	753,046	590,799	162,247	21.5	high	2.8	9,884	1.3	low	0.7
Tarragona	808,420	658,106	150,314	18.6	high	2.6	10,087	1.2	low	0.7
Castellón	604,274	492,009	112,265	18.6	high	2.0	5,267	0.9	low	0.4
Lleida	439,768	359,278	80,490	18.3	high	1.4	11,471	2.6	mid	0.8
Málaga	1,609,557	1,334,530	275,027	17.1	high	4.8	33,211	2.1	mid	2.4
Madrid	6,458,684	5,378,740	1,079,944	16.7	high	18.8	174,819	2.7	mid	12.4
Murcia	1,461,979	1,220,114	241,865	16.5	high	4.2	19,607	1.3	mid	1.4
Guadalajara	251,563	212,359	39,204	15.6	high	0.7	2,247	0.9	low	0.2
S.C.Tenerife	1,027,914	874,587	153,327	14.9	high	2.7	72,454	7.0	high	5.1
Barcelona	5,511,147	4,705,660	805,487	14.6	high	14.0	104,302	1.9	mid	7.4
LaRioja	322,415	275,735	46,680	14.5	high	0.8	10,237	3.2	mid	0.7
LasPalmas	1,090,605	936,553	154,052	14.1	mid	2.7	25,548	2.3	mid	1.8
Zaragoza	973,252	845,610	127,642	13.1	mid	2.2	15,388	1.6	mid	1.1
Cuenca	217,716	189,747	27,969	12.8	mid	0.5	2,269	1.0	low	0.2
Segovia	164,268	143,194	21,074	12.8	mid	0.4	2,304	1.4	mid	0.2
Valencia	2,581,147	2,266,752	314,395	12.2	mid	5.5	36,944	1.4	mid	2.6
Huesca	228,566	200,756	27,810	12.2	mid	0.5	5,063	2.2	mid	0.4

Teruel	145,277	127,643	17,634	12.1	mid	0.3	3,656	2.5	mid	0.3
Toledo	697,959	613,984	83,975	12.0	mid	1.5	6,627	0.9	low	0.5
Melilla	76,034	67,161	8,873	11.7	mid	0.2	3,527	4.6	high	0.3
Navarra	636,924	565,555	71,369	11.2	mid	1.2	16,766	2.6	mid	1.2
Soria	95,258	85,388	9,870	10.4	mid	0.2	4,421	4.6	high	0.3
Burgos	374,826	340,260	34,566	9.2	mid	0.6	12,122	3.2	mid	0.9
Araba/Álava	317,352	289,142	28,210	8.9	mid	0.5	4,139	1.3	low	0.3
C. Real	529,453	483,452	46,001	8.7	mid	0.8	4,175	0.8	low	0.3
Huelva	518,081	475,328	42,753	8.3	mid	0.7	5,200	1.0	low	0.4
Albacete	401,682	369,277	32,405	8.1	mid	0.6	5,129	1.3	low	0.4
Ávila	171,896	159,283	12,613	7.3	mid	0.2	6,005	3.5	mid	0.4
Granada	918,072	853,738	64,334	7.0	mid	1.1	34,317	3.7	high	2.4
Cantabria	592,250	553,049	39,201	6.6	mid	0.7	25,170	4.2	high	1.8
Valladolid	533,640	500,984	32,656	6.1	mid	0.6	9,005	1.7	mid	0.6
Gipuzkoa	707,263	664,814	42,449	6.0	mid	0.7	19,313	2.7	mid	1.4
Bizkaia	1,153,724	1,085,014	68,710	6.0	mid	1.2	27,011	2.3	mid	1.9
León	499,284	473,321	25,963	5.2	mid	0.5	35,339	7.1	high	2.5
Ourense	335,219	318,508	16,711	5.0	mid	0.3	82,134	24.5	high	5.8
Ceuta	80,579	76584	3,995	5.0	low	0.1	2,132	2.6	mid	0.2
Salamanca	353,619	336,113	17,506	5.0	low	0.3	23,265	6.6	high	1.7
Asturias	1,084,341	1,035,055	49,286	4.5	low	0.9	83,041	7.7	high	5.9
Palencia	172,510	165,301	7,209	4.2	low	0.1	5,510	3.2	mid	0.4
Zamora	194,214	186,173	8,041	4.1	low	0.1	14,820	7.6	high	1.1
Pontevedra	962,472	922,678	39,794	4.1	low	0.7	106,279	11.0	high	7.5
Sevilla	1,917,097	1,840,007	77,090	4.0	low	1.3	22,326	1.2	low	1.6
Lugo	353,504	339,328	14,176	4.0	low	0.2	50,352	14.2	high	3.6
Cádiz	1,236,739	1,188,972	47,767	3.9	low	0.8	19,825	1.6	mid	1.4
Cáceres	415,083	399,767	15,316	3.7	low	0.3	12,705	3.1	mid	0.9
Badajoz	692,137	668,097	24,040	3.5	low	0.4	8,803	1.3	low	0.6
ACoruña	1,146,458	1,107,469	38,989	3.4	low	0.7	128,090	11.2	high	9.1
Córdoba	805,108	779,849	25,259	3.1	low	0.4	13,920	1.7	mid	1.0
Jaén	670,761	650,094	20,667	3.1	low	0.4	9,128	1.4	mid	0.6

Table A.13: **Main Data Sources**

Variable	Description	Source
X_{nit}	value of the exports from province i to country n in year t (thousands of €)	1995–2011: Datacomex, `http://datacomex.comercio.es/principal_comex_es.aspx`. Full database at 3-digit disaggregation requested and received by email
X_{nt}	Country n GDP in year t (billions of US$)	1995–2011: IMF World Economic Outlook Database, `http://www.imf.org/external/pubs/ft/weo/2012/01/weodata/index.aspx`
Y_{it}	Province i gross product in year t (thousands of €)	1995-2010: INE - "PIB a precios de Mercado precios Corrientes", `http://www.ine.es/jaxi/menu.do?L=0&type=pcaxis`
$Immi_{nit}$	Foreign residents with country n nationality residing in province i at year t	1998–2011: INE – "Población extranjera por sexo, comunidades y provincias y nacionalidad" `http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe245`
NID_{nit}	"No Immigrants Dummy", equal to 1 if at time t there are no immigrants from country n in province i, and zero otherwise	
Emi_{nit}	Spanish expatriates registered in province i and residing in country n at year t	2006-2011: Censo Electoral de españoles residentes en el extranjero (CERA) por provincia de inscripción y país de residencia, `http://www.ine.es/ss/Satellite?c=Page&cid=1254735793323`
NED_{nit}	"No Emigrants dummy", equal to 1 if at time t there are no emigrants from province i residing in country n, and zero otherwise	