The pro-export effect of sub-national migration networks: New evidence from Spanish provinces

Anna D'Ambrosio*

Politecnico di Torino.

Sandro Montresor†

Gran Sasso Science Institute

October 26, 2019

*Corresponding author. Email: anna.dambrosio@polito.it. Corso Duca degli Abruzzi 24, Torino, Italy. Declarations of interest: none

†sandro.montresor@gssi.it. Declarations of interest: none

Abstract

We integrate state of art advances in the gravity model literature and investigate the effect that sub-national networks of both immigrants and emigrants have had on the export of Spanish provinces (NUTS3) over the period 2006-2016. We allow for heterogeneity in the provincial export capacity, which significantly reduces pro-trade effects and select the most suitable estimator through diagnostic tests. When both migration flows are considered and instrumented over the period that comprehends the double-deep crisis, the pro-export effect of immigrants found by previous studies vanishes and that of emigrants, instead, prevails, questioning the role of migrants' information effects in the considered context. Over the considered period, the effect of migrants appears to be declining while that of immigrants seems to increase, pointing to a change in the composition of the migration stocks.

JEL codes: F10, F14, F22, C52.

Keywords: Gravity model, migration, subnational units, Poisson PML, Gamma PML, fixed effects.

1 Introduction

The recent upsurge of migration flows across the globe and the change in their geography are putting the economic effects of migration on the top of the academic research agenda. In particular, the trade impact of migrants has been receiving increasing attention since the seminal works by Gould (1994) and

Head and Ries (1998)[1]. Being embedded in transnational networks, migrants can attenuate the information and enforcement costs that, even in the ICT era, still affect the international exchange of commodities and services (Anderson and van Wincoop, 2004; Rauch, 2001). This argument has received much empirical support. However, results appear sensitive to the adopted methodology and perspective (e.g. countries vs. sub-national units, immigrants vs. emigrants, imports vs. exports, standard vs. differentiated commodities, similar vs. dissimilar countries) and keep on revealing the nuances of a still open research issue.

In studying the relationship at stake, the gravity model of international trade (Anderson and van Wincoop, 2003; Chaney, 2008), augmented with migration variables, is generally recognized as a suitable interpretative framework (for the rationale of this augmentation, see, e.g. Gould, 1994; Dunlevy and Hutchinson, 1999; Rauch and Trinidade, 2002; Head and Mayer, 2014).[2] On the other hand, the substantial theoretical and econometric advances recently obtained in the gravity literature (see Head and Mayer, 2014, for a comprehensive review) have only partially percolated into the analysis of the link between migration and trade, and even less into its analysis with panel data.[3] Even more recent econometric and computational advancements (Correia et al., 2019a; Weidner and Zylkin, 2018) can be exploited to disentangle the case for an actual pro-trade effect of migrants, addressing recent claims

[1] For a review of the literature, see Felbermayr et al. (2015); recent contributions include Bratti et al. (2018), Burchardi et al. (2018) and Parsons and Vézina (2018).

[2] The reference to transnational contacts and networks is now established in trade models thanks to the works by Arkolakis et al. (2012) and Chaney (2014).

[3] A notable exception is the recent study by Bratti et al. (2018), though with important methodological differences compared to our own, which we will discuss in the following.

of a possibly overstated nexus (Parsons, 2012; Burchardi et al., 2018). Relying on these methodological developments, we propose to investigate the migration-trade link with panel data in an original way, by jointly considered three important issues.

First of all, unlike the majority of previous studies,[4] we simultaneously address the pro-trade effect of both emigrants and immigrants. In so doing, we claim that to properly account for the social and business network effects that migrants activate in international trade, both directions of migration should be taken into consideration. Otherwise, the effects will likely be overestimated and may be ascribed to the wrong underlying mechanism, especially in countries with large diasporas such as the one we investigate in our empirical application, that is, Spain.

Second, we consider the effect of immigration and emigration on trade for quite disaggregated geographical units of analysis, that is, the provinces of Spain (NUTS3-level regions). Given the spatial heterogeneity in the distributions of both migration and trade, the migration-trade link is arguably a largely localized phenomenon. The extant literature at the regional level is already quite developed[5]; yet, to the best of our knowledge, this is the first study that investigates the trade effect of emigration (along with immigration) at the sub-national scale and that allows for sub-nationally heterogeneous export capacity. More precisely, in line with the most recent contributions on

[4]Among the few exceptions, see Felbermayr et al. (2015); Hiller (2014); Rauch and Trinidade (2002); Murat and Pistoresi (2009).

[5]The migration-trade link has already been investigated with respect to different kinds of subnational units within countries: e.g. US states (Burchardi et al., 2018; Herander and Saavedra, 2005; Dunlevy, 2006; Bandyopadhyay et al., 2008), Canadian provinces (Wagner et al., 2002), Italian provinces (Bratti et al., 2014, 2018), French departments (Briant et al., 2014), and even Spanish provinces (Peri and Requena-Silvente, 2010).

the topic by Burchardi et al. (2018); Bratti et al. (2018) and Briant et al. (2014), we allow for heterogeneity in the "multilateral resistance" factors that inhibit sub-national trade to any partners (i.e., in the so-called "multilateral resistance term" (MRT); Anderson and van Wincoop, 2003).

The third issue that we address refers to the delicate methodological choice of how to estimate migration-augmented gravity models of trade with panel data: an issue at the frontier of the econometric "best practices" (page 487) to consistently identify the determinants of international trade (Larch et al., 2019). Drawing on recent econometric advances on Pseudo-Maximum Likelihood (PML) estimators with multi-way fixed effects (Correia et al., 2019b) as well as on the related computational developments (Correia et al., 2019a; Weidner and Zylkin, 2018), we estimate a gravity model that jointly addresses two issues, so far thwarted by computational problems (Larch et al., 2019): i) the heteroskedasticity bias entailed by log-linear models, which is addressed by resorting to PML estimators (Poisson or Gamma) with dependent variables in levels; ii) the simultaneous inclusion of time-varying exporter and importer fixed effects – to account for heterogeneous MRT (Feenstra, 2004; Baldwin and Taglioni, 2007) – and time-invariant region-country fixed effects – to control for bilateral heterogeneity in unobservable trade barriers (Baier and Bergstrand, 2007). In particular, following Head and Mayer (2014); Santos-Silva and Tenreyro (2006); Manning and Mullahy (2001), we select the suitable PML estimator to be used in the estimates based on some diagnostic tests, which address the underlying distribution of the errors and potential misspecification sources.

We integrate the previous three issues in the analysis of 5,450 trading pairs, constituted by 50 Spanish provinces[6] and 109 countries over the 11 years 2006-2017. We carry out our integrated analysis by focusing on exports only, leaving the analysis of the migrants' effect on the imports of Spanish provinces to future research. We first work out the elasticity of exports to both immigration and emigration and compare their size with that of previous studies at the sub-national level. Second, we use our integrated methodology to revisit some previously established stylized facts of the migration-trade link, like the role of institutional and language similarity between trade-partners, and the distinction between local and non-local effects of migration networks. Finally, we provide a first exploration of the extent to which the investigated effects of migrants on exports has changed over the particular temporal window we are considering. Indeed, focussing on the 2006-2017 timeline, we provide a first endurance test of the trade-effect of migration during a relatively negative phase of the business cycle, marked by the entrance in the sub-prime mortgage crisis, by the unfolding of the sovereign-debt default one, and by the subsequent recession.

Our results are only partially confirmative of the available knowledge on the issue and add to it some novel results. Echoing the recent findings by Burchardi et al. (2018) on US counties, when both directions of migration are retained, and their endogeneity is addressed through an instrumental variable (IV) strategy, we do not find robust evidence of an immigrants' effect on exports. Emigrants, instead, increase the export of Spanish provinces to an aligned extent with previous studies. Because the effects of immigrants and emigrants are not entirely symmetric – emigrants directly add up to the de-

[6]Ceuta and Melilla are excluded due to data availability reasons.

mand for home country goods, while immigrants do not– this result actually puts in question the case for a migrants' information effect.

The remainder of this paper is organized as follows. In Section 2 we position our main research issues in the extant literature. In Section 3.1 we present the data for our empirical strategy and in Section 3.2 we illustrate its methodological novelties. Section 4 illustrates our results. Section 5 concludes.

2 Migrants and exports at the sub-national level: heterogeneous export capacities

The mechanisms through which migration can affect exports and, more in general, trade have been extensively studied over the last decades.[7] By moving to a new location, and by keeping a relationship with their origin, migrants bridge the two and create social and business networks that span across countries (Rauch, 2001; Rauch and Casella, 2001). Embeddedness within these transnational networks may allow migrants to circumvent a set of informal barriers that, even in the ICT era, still inhibit trade (Anderson and van Wincoop, 2004). First of all, migrants can help fill the information gap between sellers and buyers on the two sides of their migration route, given the knowledge of customs, laws, language and business practices they have of both (*information effect*). In so doing, they could help to discover new business opportunities. Second, within their transnational networks, migrants could put in place implicit enforcing mechanisms (e.g. punishment, sanctions, and exclusions) of

[7]The first contributions date back to Gould (1994); Head and Ries (1998); Rauch and Trinidade (2002).

international contractual relationships, which could compensate for the lack or weakness of institutional protection mechanisms for trade (*enforcement effect*). Also, emigrants' and immigrants' preferences for their homeland products (their *preference effect*) increase trade unidirectionally, with emigrants' promoting the foreign demand for exports and immigrants increasing the domestic demand for imports (Parsons, 2012; Hatzigeorgiou, 2010).

While the three effects have been mainly disentangled for immigrants, their relevance appears evident also for emigrants, especially when the focus is on exports. In countries with large diasporas, such as Spain, the emigrants' preference effects may be substantial drivers of exports that may confound the results, were the analysis to focus on the side of immigration only. From the perspective of Spanish provinces, indeed, both immigrants and emigrants may exert information and enforcement effects; instead, preference effects accrue to emigrants only. Hence, omitting the emigration side from the analysis may not only overstate the immigrants' effects but more importantly lead to a wrong inference about the underlying mechanism. Indeed, we may attribute to information and enforcement an effect that is, at least partly, a preference effect (Section 2.1).

These mechanisms rely strongly on the business and social networks that migrants create, which in turn operate mainly through direct interpersonal contacts and proximity. Because of that, and considering that migrants are typically very unevenly distributed within a country, there are likely subnational differences in their networks, thus calling for a sub-national level of analysis (Section 2.2). Furthermore, as we shall see, the export capacity relative to which their effects should be evaluated is also very sub-nationally heterogeneous, thus requiring to account for heterogeneous multilateral resistance terms (Section 2.3).

2.1 Emigrants and immigrants

In the existing literature, the pro-trade effects of migration have been generally invoked by referring to the networks that immigrants can be expected to create between their hosting and origin countries. The emigration side of the same networking has been instead generally neglected, mainly due to the lack of data.[8] This neglect appears to us unfortunate in three respects and motivates our choice to look at the pro-export effect of immigrants and emigrants simultaneously.

First, emigrants represent moving communities that, similarly to immigrants, are capable of creating social and business ties between their origin and destination loci. Working outwards, rather than inwards, the three effects descending from these ties could operate differently. For example, emigrants could promote the realisation and enforcement of trade opportunities abroad with co-nationals from any other province of their origin country, with whom they share language and social capital, rather than just with those from their own provinces; hence, emigrants may ultimately promote trade of other provinces beyond their specific one.

Second, emigrants and immigrants are likely complementary and may perform their bridging role in different contexts. The emigrants' destination countries may substantially differ from the immigrants' origin countries in terms,

[8] Notable exceptions are the country-level studies by Parsons (2012); Hiller (2014); Felbermayr et al. (2015); Murat and Pistoresi (2009).

for example, of resource endowments, cultural habits, and institutional setups (Girma and Yu, 2002; Dunlevy, 2006). Accordingly, as we will reconsider, their trade contribution could be higher or lower than that of immigrants, depending on them bridging more or less dissimilar communities and the related higher or lower barriers to trade (Rauch, 2001). Also, emigration and immigration could follow distinct historical routes, so that their ability to generate social and business ties could be different in terms of their respective path-dependence (Gould, 1994; Rauch, 2001). Furthermore, differences between emigrants and immigrants in tastes and human capital could translate into different emigrants' and immigrants' effects on the trade of specific commodities and services (Rauch and Trinidade, 2002; Peri and Requena-Silvente, 2010; Briant et al., 2014).

Finally, the complementarity of their routes does not rule out the possibility that immigrants and emigrant stocks are correlated with each other and with trade. In this case, the exclusion of emigrants may bias the estimates of the pro-export effects of immigrants. Emigrants could be among the omitted bilateral variables that, if correlated with both immigrants and exports, make the latter relationship spurious. As we have argued above, this may not only overestimate the immigrants' effect but lead to wrong inference about the relative importance of the information/enforcement and preference effects.

This last consideration connects to the second distinguishing feature of our paper: the choice of addressing the pro-export effect of migrants (both immigrants and emigrants) focusing on sub-national units of analysis.

2.2 The sub-national level

As previously mentioned, the information flows that account for a great part of the migrants' pro-trade effects, are more likely to occur at the local scale (Rauch, 2001).

Investigating such a localized phenomenon at the country level could suffer from the Modifiable Areal Unit Problem (MAUP, Openshaw, 1983), and the choice of the sub-national level of analysis appears preferable. Furthermore, the reference to sub-national observations does increase data variability and mitigates the problems of spurious correlation affecting the relationship between trade and migration (Wagner et al., 2002; Bratti et al., 2014; for a review of the studies at the national vs. sub-national level, see Peri and Requena-Silvente, 2010 and Felbermayr et al., 2015).

For these reasons, the literature has been progressively moving towards a finer geographic disaggregation in the units of analysis, while also exploiting the longitudinal dimension of trade and migration data. The same arguments motivate the choice of our own level of investigation.

Nonetheless, the implications from a sub-national perspective in identifying the pro-export effects of migration have not been fully exploited yet. We address this issue in the next section.

2.3 The gravity model with heterogeneous export capacities

As is well-known, the micro-foundations of the gravity model of international trade by Anderson and van Wincoop (2003) led to an important extension of its standard "naive" formulation, in turn mainly drawn on the analogy with the Newtonian law (for which, see Tinbergen, 1962; Bergstrand, 1985). Not only are country i's exports to country j, X_{ij}, assumed to be a positive function of their economic masses, Y_i and Y_j, and a negative function of their distance and of the relative transaction costs, ϕ_{ij}, respectively;[9] but the "monadic" terms are in turn adjusted by considering the average openness to trade of each trading partner, i.e., their "Multilateral Resistance Terms" (MRT). Denoting with Ω_i the average market size accessible to the exporting country and with Φ_j the average degree of competition of the importing one[10], the "structural" form of the gravity equation (Head and Mayer, 2014), in a cross-sectional context is the following[11]

$$X_{ij} = \frac{Y_i}{\Omega_i} \frac{Y_j}{\Phi_j} \phi_{ij} \tag{1}$$

Thus, any change in bilateral trade barriers, encapsulated in the "dyadic" term, ϕ_{ij}, like their reduction entailed by migration, should be evaluated relative to the MRT, rather than in absolute terms (Anderson and van Wincoop, 2003). Following Baldwin and Taglioni (2007), the application of Equation 1 to a panel context requires recognizing that most variables of interest, including the MRT, are time-varying.

When sub-national units of analysis are adopted, the gravity model be-

[9]In general, Y_i and Y_j represent the "mass" of production and the "mass" of expenditures of the exporting and the importing partner, respectively, and are proxied by their GDP. ϕ_{ij} instead refers to natural trade barriers, such as the geographical distance between countries, but also to other economic barriers, such as tariffs, as well as to their respective elasticities.

[10]More precisely, Ω_i represents the "expenditure-weighted average of relative access" and Φ_j the "accessibility-weighted sum of exporters' capabilities" (Head and Mayer, 2014, : 140-141).

[11]As Head and Mayer (2014) have shown in their review, "structural" gravity equations (and their "general" form) are compatible with the wide majority of trade models used in the literature.

comes asymmetrical: in our case, exporters are the NUTS3 Spanish provinces, while the importers are their destination countries. However, the interpretation of the terms in Equation 1 remains the same (see Bratti et al., 2018). In particular, in order not to incur in what has been called the "gold medal mistake" in the gravity literature (Baldwin and Taglioni, 2007), the need remains to account for the heterogeneity of the exporter-side MRT, meant as the province's (weighted) capacity of exporting to any countries in the world.[12]

Two main arguments support the expectation of sub-national heterogeneity in the MRT. First, as recently formalized by Bratti et al. (2018), sub-nationally heterogeneous exporting capacity can be traced back to the heterogeneous productivity of the firms located in them; this pattern suits countries marked by a geographically fragmented production structure such as Spain.[13] Accounting for this heterogeneity bears implications for the study of the migration-trade link. Indeed, the average productivity of the firms located in a given province is not unrelated to bilateral migration stocks. The overall supply of immigrant labor may affect productivity, wages, as well as the offshoring decisions of firms (e.g. Ottaviano and Peri, 2006; Ottaviano et al., 2013) and can ultimately affect the accessibility-weighted exporting capacities of the exporter. In turn, the overall supply of immigrant labor correlates with the supply of labor from a specific country. Furthermore, provinces with more productive firms may have a more dynamic structure of opportunities and attract more migrants

[12]Similarly, the importer-side MRT should be seen as the average (weighted) market access of a given country to any province of Spain, as well as to any other exporter worldwide.

[13]In their framework, based on the recent model by Arkolakis et al. (2012), the sub-national exporting capacity is modeled as a function of the number of exporters in province i and on the price charged by the exporters on their varieties. In turn, the price charged by province i exporters on products exported to j depends on production costs (i.e. wages), transportation costs and on the productivity of i firms exporting to j relative to the average productivity of firms in i.

from any origin country.

A second argument that calls for the consideration of the sub-national specificity of export capacities (via heterogeneous MRT) concerns the possibility of accounting for the localized and non-localized export effects of migrants, as we will do in our empirical application. As Herander and Saavedra (2005) have shown with respect to US states (1993-1996), immigrants from country j increase the capacity of a certain state i to export towards j, not only directly, that is, through in-state (local) immigrants, but also indirectly: that is, through out-of-state (nation-wide) immigrants, which provide trade-related information about j through their interstate mobility and their interstate networks (e.g., through their national associations and coordination mechanisms).[14]

In spite of this rich theoretical background, the estimate of sub-national gravity models with heterogeneous exporter MRT has been only recently incorporated in panel data analyses (see Bratti et al., 2018). Burchardi et al. (2018) and Briant et al. (2014) include exporter effects, but in a cross-sectional framework. Peri and Requena-Silvente (2010); Bandyopadhyay et al. (2008) use panel data but assume the term constant across regions in the same country, while Bratti et al. (2014) retain it as invariant across the provinces (NUTS3) of the same more aggregated regional (NUTS2) level of analysis.

In trying to fill this gap in the use of gravity models with heterogeneous MRT, our application refers to NUTS3 regions (as "provinces" (*provincias*))

[14]More in general, geographical spillovers in the transmission of export opportunities could also occur through the mobility of emigrants across sub-national units – e.g. emigrants that, once reached their emigration country j return to another province than the original i, to later migrate again to j – or, more commonly, through the network that emigrants from the same country normally establish abroad.

rather than to NUTS2 (*Comunidades Autonomas* as "regions"). Furthermore, we enrich the standard specification of the model by exploiting the panel dimension of our data and, consistently with the arguments developed by Feenstra (2004) and Baldwin and Taglioni (2007), we estimate the panel version of Equation 1 by including exporter-time (province-year) and importer-time (country-year) effects – to account for the time variation in the MRT – as well as exporting region-importing country effects – to account for unobserved heterogeneity in the dyadic trade barriers. As it is theoretically consistent, we believe that this approach is capable to provide more reliable estimates of our focal phenomenon.

3 Empirical application

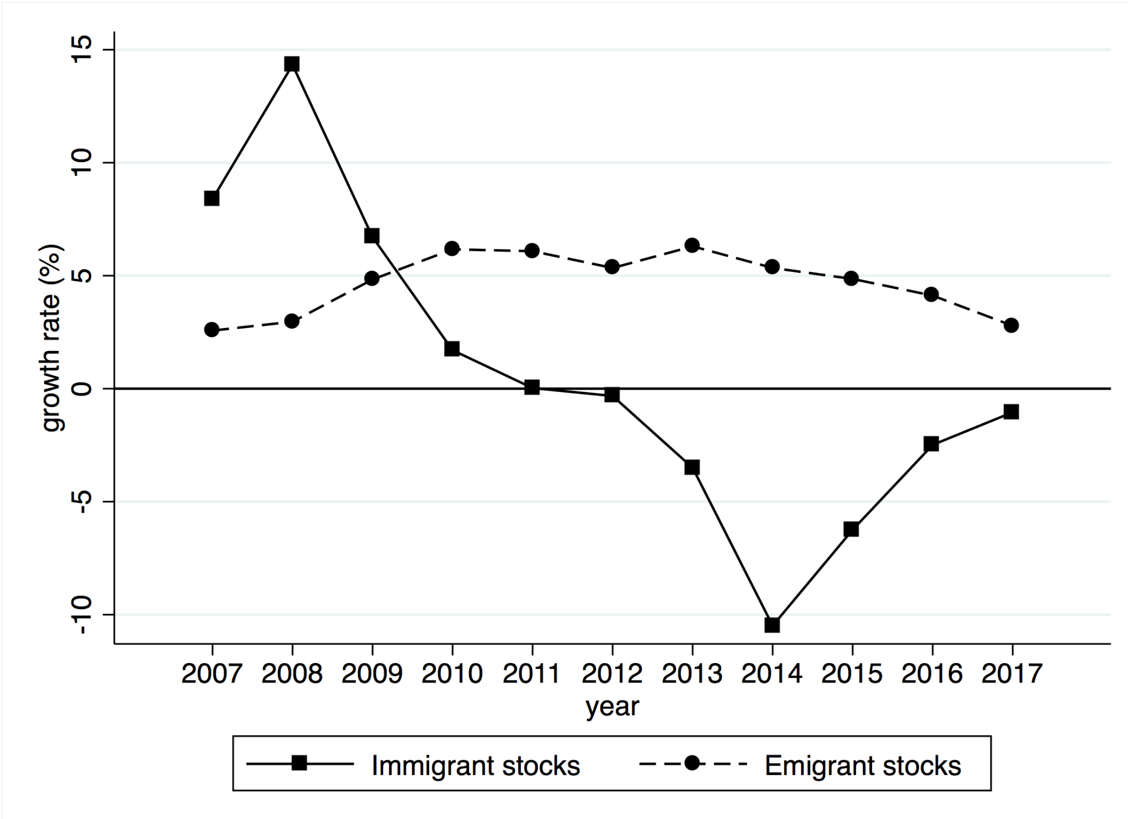
We investigate the role of migration in driving the export performance of 50 Spanish provinces (NUTS 3) towards 109 destination countries over the 11 years 2006-2017. Compared to the previous study of the migration-trade link in Spanish provinces by Peri and Requena-Silvente (2010), we include a wider their set of countries by drawing on the publicly available province-level dataset supplied by the Ministry of Economics and Competitiveness, not eliminating the dyads for which there are zero trade flows. Furthermore, unlike Peri and Requena-Silvente (2010), who focused on the pre-crisis period (1998-2007), when immigration was booming, our analysis concentrates on a relatively negative phase of the business cycle, marked by the burst of the subprime mortgage crisis, the unfolding of the sovereign-debt default one, and the subsequent recession. Over such a period, Spanish exports have been growing at an average rate of 4.1%, while emigration and immigration stocks (measured as described in the following) at an average rate of 4.7% and of 0.7%, respectively. The underlying patterns in the growth rates of immigrants and emigrants stocks have been, however, very different, as illustrated in Figure 1 for the countries in our sample. While exports' growth rates (not shown) faced a single substantial drop in 2009 and rapidly recovered, the growth in immigrants' stocks has been slowing down over the entire period, taking negative values from 2011 onwards. On the contrary, emigrants' stocks have been increasing relatively stably over the considered period, but more strongly during the crisis years.

Given the particular trend that migration and export revealed over this crisis period, looking at their relationship represents an interesting exercise to evaluate the endurance of migration effects along the business cycle. To the best of our knowledge, this is the first study to perform such an analysis.

3.1 Data

The dataset used for the empirical analysis is a balanced panel. Exports data are retrieved from the official statistics of the Ministerio de Economia, Industria y Competitividad (MEIC) in Spain. For the sake of illustration, Figure 2 represents the relationship between exports of the province of Madrid and the distance-weighted GDP of EU partner countries in 2008. The resulting picture is reassuring about the choice of the gravity model as an interpretative framework for Spanish provinces' exports. The slope of the fitted line is 0.94, very close to one, in line with the stylized facts highlighted in the gravity literature

Figure 1: Growth rates of immigrant and emigrants' stocks in Spain, 2007-2017

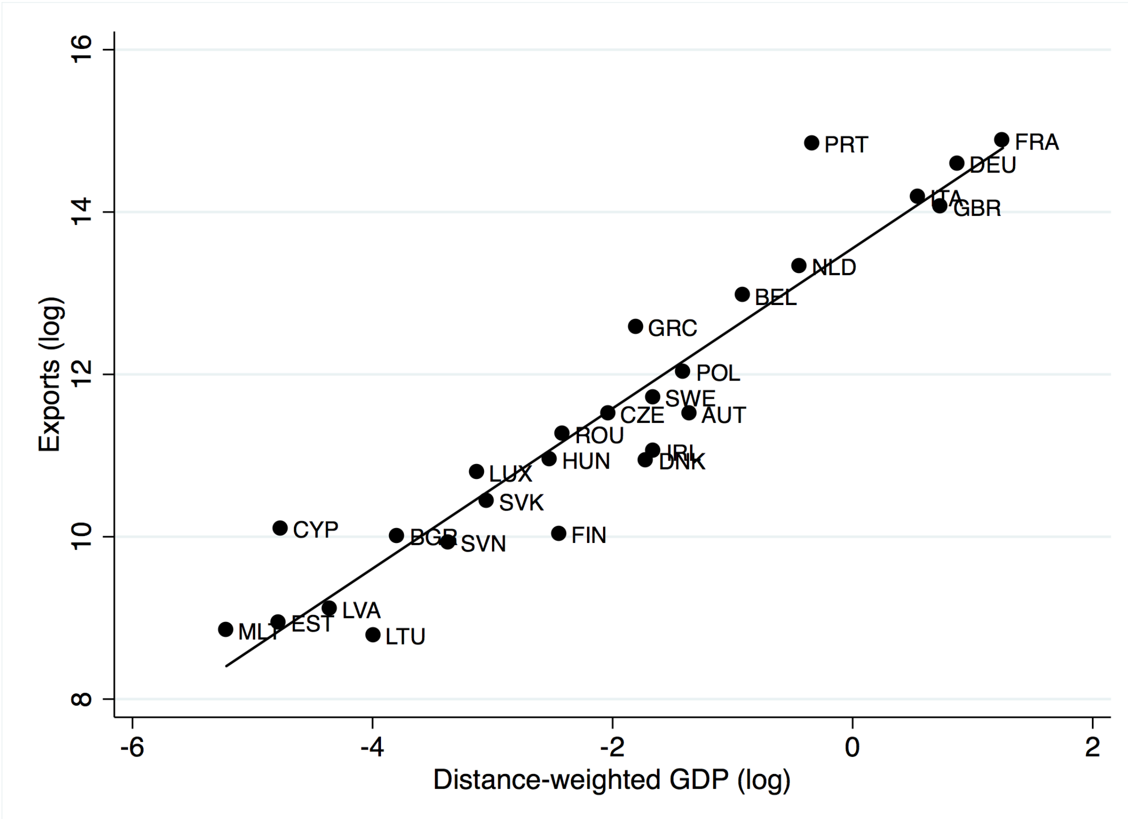


The total immigrant and emigrant stocks on which the growth rates are computed are obtained by aggregating our bilateral immigration and emigration stocks data by year. Source: Own elaborations on INE data

(Head and Mayer, 2014).

We matched trade data with demographic data sourced from the Spanish National Statistical Institute (INE); our main variables and the relative sources are reported in Table A.10 in Appendix. Some words of caveat are due concerning migration variables, whose characteristics (e.g. skills, employment status and length of stay) are affected by severe data constraints at the sub-national

Figure 2: Gravity model and the exports of Madrid to EU countries, 2008



The figure plots the log exports of the province of Madrid to EU countries vs. the log of the ratio between each country's GDP and its distance to Madrid. Source: own elaborations on MEIC and INE data

level. Drawing on the extant literature, we measure immigrants $Immi_{ijt}$ by the stock of residents registered in province i according to the municipal registries ("Padrón Municipal") who hold the citizenship of non-Spanish country j at time t.

As is well-known 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. Furthermore, the same stock refers only to formally residing people; it also neglects undocumented immigrants and intra-national movements of foreigners that are not registered in official changes of residence. Imperfect is also our measurement of emigrants, Emi_{ijt}, captured (as in Flisi and Murat, 2011) by the stock of people recorded in the Spanish election registries of province i who have moved their residence to foreign country j at time t. Our variable likely underestimates the actual emigrant stocks. Indeed, migrants typically inscribe in the electoral registries when they are fairly established in a foreign country and are intending to stay in a long-term perspective. Hence, we may underestimate the emigrants of more recent expatriation. Furthermore, while the data report the emigrants' last province of residence in Spain, they are uninformative as to the emigrants' country of birth, so that we 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 to us safe to assume that Emi_{ijt} reflects the dynamics of the Spanish emigrant population more than the dynamics of return migration.

Table 1 reports the summary statistics for the main variables of our application, and Table 2 shows their mutual correlation. First of all, let us notice that the correlations between exports (X_{ijt}) and each of the two migration variables (0.182 with $Immi_{ijt}$ and 0.281 with Emi_{ijt}) are higher than that between the two. Indeed, the correlation between $Immi_{ijt}$ and Emi_{ijt} is quite low (0.115). Furthermore, the main origin and destination countries of immigrants and emigrants and their respective distributions differ substantially (see Figure 3 for the year 2010). Quite distinct are also the distribution patterns of immigrants and emigrants by Spanish province (still for 2010, see Table A.11 in the Appendix). As we argued in Section 2.1, $Immi_{ijt}$ and Emi_{ijt} can be assumed to portray different phenomena and, supporting the first of our methodological choices, are both in need of consideration.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
Exports X_{ijt}	36 557.36	202 660.20	0.00	7 718 364.00
Prov. p.c. gross product Y_{it}	21.56	4.51	14.64	38.10
Country p.c. GDP Y_{jt}	16.12	21.10	0.25	120.79
Immigrants $Immi_{ijt}$	954.83	5 459.89	0.00	219 567.00
Emigrants Emi_{ijt}	266.07	1 692.90	0.00	55 022.00
No Immigrants NI_{ijt}	0.08		0.00	1.00
No Emigrants NE_{ijt}	0.28		0.00	1.00
Km distance $Dist_{ij}$	4 985.11	3 487.49	164.69	19 959.60

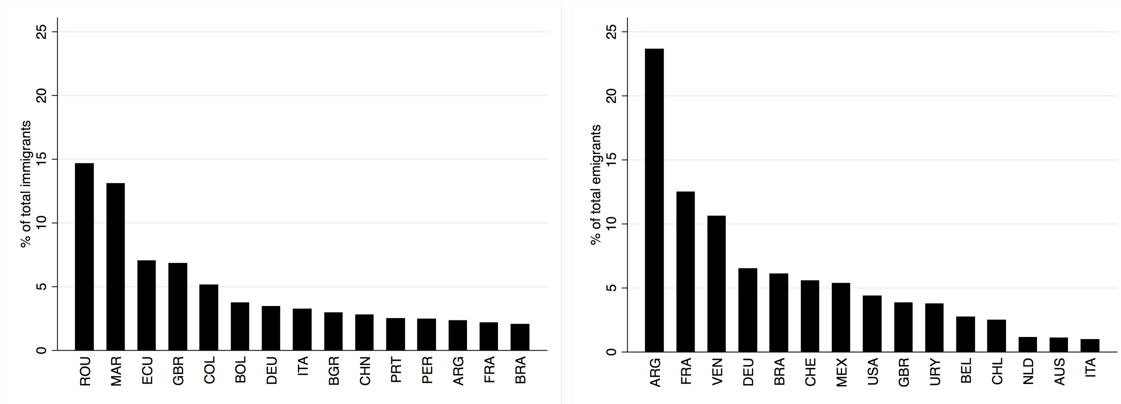
Observations: 54,200.

Table 2: Correlations

Variable		1	2	3	4	5	6	7	8
1	X_{ijt}	1.000							
2	Y_{it}	0.119	1.000						
3	Y_{jt}	0.154	−0.004	1.000					
4	$Immi_{ijt}$	0.174	0.080	−0.015	1.000				
5	Emi_{ijt}	0.274	0.027	0.081	0.107	1.000			
6	NI_{ijt}	−0.049	−0.045	0.035	−0.051	−0.045	1.000		
7	NE_{ijt}	−0.109	−0.056	−0.301	−0.097	−0.100	0.195	1.000	
8	$Dist_{ij}$	−0.119	−0.013	−0.168	−0.035	0.057	−0.003	−0.162	1.000

Indeed, a descriptive overview of the data suggests that both immigration and emigration can have an impact on exports at the province level. For illustration, Figure 4 focuses on the province of Madrid in 2010 and plots the exports-to-GDP ratio against the immigrant and emigrant stocks, weighted by

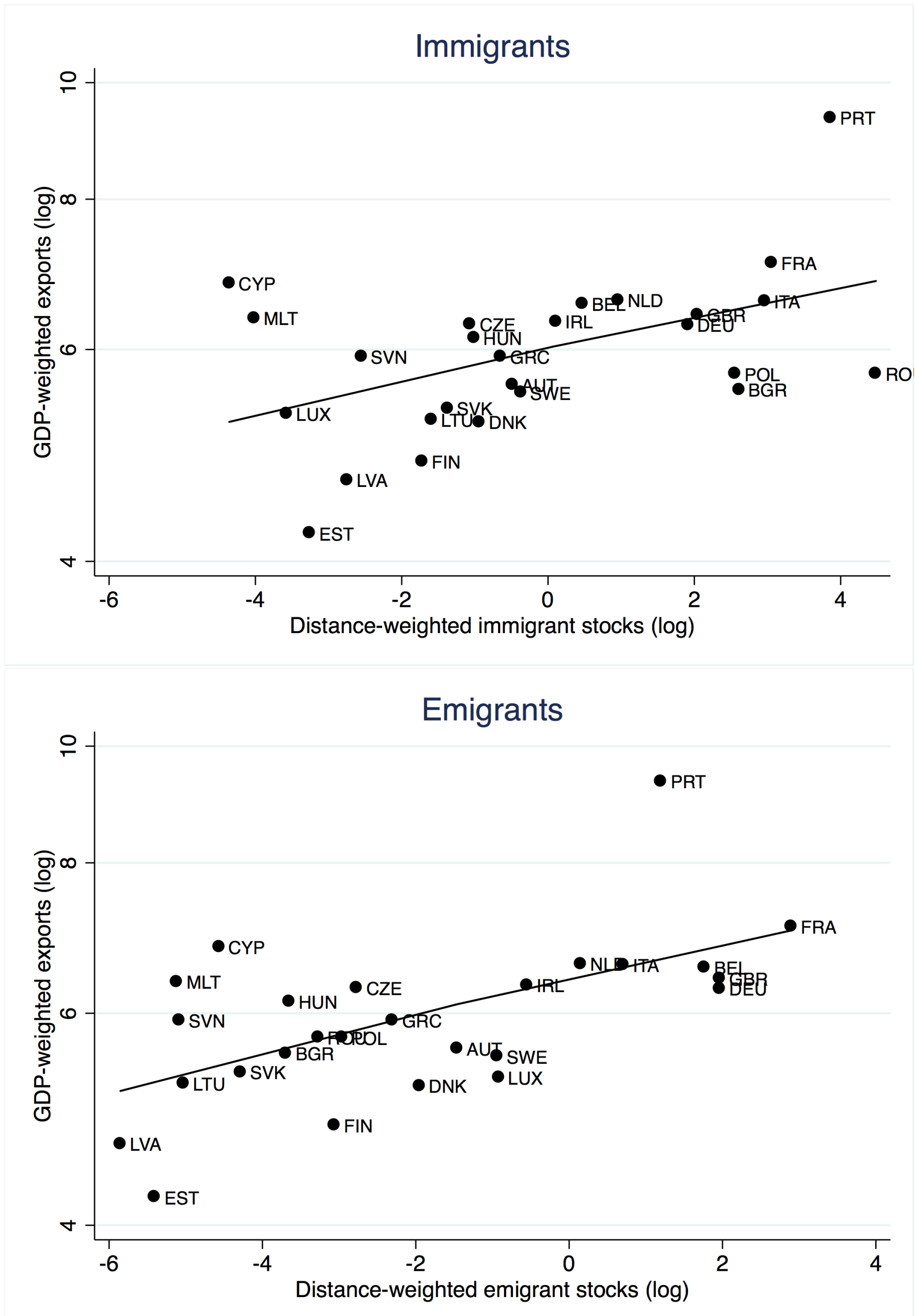
Figure 3: Top 15 origin and destination countries of immigrants and emigrants (2010)



Source: Own elaboration of INE data.

distance, for each OECD partner country. The relationship appears positive for both immigration and emigration, and slightly stronger for the former than for the latter. The province of Madrid exports more to the countries with which immigrants and emigrants have helped to build up larger transnational communities.

Figure 4: Migration and Trade



GDP-weighted exports towards EU countries vs. the respective distance-weighted immigrants and emigrant stocks: Madrid and EU countries, 2008. To mitigate the role of confounding factors, the Figure only portrays EU countries. Similar results are obtained when looking at OECD countries. Our own elaboration on Datacomex, INE, CERA and FMI data.

Before turning to a more rigorous econometric analysis of this evidence, we should notice two further features of the data. The first derives from an inspection of the summary statistics. The distribution of our dependent variable is characterized by a right skew and many small values. This is typical of trade data, and even more of trade data on sub-national units, which typically display more variation for province-country pairs characterized by larger trade volumes. This is a first indication of heteroskedasticity, which we will have to take into consideration through a suitable estimator. As to the issue of zero trade flows, 7.45% of our trade observations have a nil value. The shares of zeroes in our variables of interest ($Immi_{ijt}$ and Emi_{ijt}) are 8% and 28%, respectively.

3.2 Econometric strategy

The specification we propose to estimate our gravity model in Eq.1 follows Baldwin and Taglioni (2007), and includes a vector of country-year effects, $\theta_{\mathbf{jt}}$, and a vector of province-year effects, $\omega_{\mathbf{it}}$.[15] Even including these fixed effects, the presence of preferential ties liking specific dyads could confound the estimation of the migrants' effects. Historical reasons, including colonial history, past migration, but also geography and transport infrastructure may be responsible for tighter trade relationships between specific pairs, but also larger bilateral migration stocks. In this case, migration variables would also capture the unobserved role of history and geography (Briant et al., 2014; Burchardi et al., 2018). In order to account for the correlation between the

[15]According to them, the regional MRT would depend on time-varying bilateral trade costs and on time-varying economic masses, and should thus be time-varying itself.

unobservable and observable components of the bilateral trade determinants (Baier and Bergstrand, 2007; Baldwin and Taglioni, 2007), we thus further include the bilateral distance $Dist_{ij}$ between dyads and, most importantly, time-invariant region (NUTS2) $\times$ country effects, $\eta_{\mathbf{rj}}$ (not present in the recent specifications by Bratti et al., 2018). This is an important advantage of the panel approach we are implementing. Such a choice overcomes the difficulty to include dyadic province (NUTS3) $\times$ country effects, as recommended by Baier and Bergstrand (2007), given the low residual variation in exports when province-year and country-year effects are included jointly with province-country effects. Indeed, these effects together explain between 90% and 98% of the variation in our dependent variable, depending on the estimator.

Hence, our identification is based on two main sources of variation: the cross-sectional variation across provinces within the same region-country pair; and the within-region-country-pair time variation that is not explained by country- and province-specific shocks. We are confident that this approach captures the relevant cross-sectional and time variation in the phenomenon that we are interested in while controlling for most of the confounding factors that would pose threats to our internal validity—most importantly, that the more economically dynamic provinces within a given region are at the same time the strongest exporters and the strongest attractors for migrants.

The resulting gravity model with three-way fixed-effects described above (see Anderson and van Wincoop, 2003; Head et al., 2010; Cheng and Wall, 2005; Baldwin and Taglioni, 2007) is augmented with our variables of interest. The stock of immigrants from country j living in province i at time t and the stock of emigrants from province i living in country j at time t are included as logs. In order not to lose observations, we add one unit to them, $(\text{Immi}_{ijt}+1)$ and $(\text{Emi}_{ijt}+1)$. Furthermore, in order to control for possible non-linearities associated with this transformation, we add two further dummy variables: "No immigrant" (NI_{ijt}) and "No Emigrant" (NE_{ijt}), equal to one if the immigrant (emigrant) stocks from (to) country j to (from) province i in year t are equal to zero, and zero otherwise. Our benchmark econometric model is thus the following:

$$X_{ijt} = (\text{Y}_{it-1} \times \text{Y}_{jt-1})^{b_1} \text{Dist}_{ij}^{b_2} (\text{Immi}_{ijt-1}+1)^{b_3} (\text{Emi}_{ijt-1}+1)^{b_4} \times \text{NI}_{ijt-1}^{b_5} \text{NE}_{ijt-1}^{b_6} e^{(\boldsymbol{\omega_{it}} + \boldsymbol{\theta_{jt}} + \boldsymbol{\eta_{rj}} + \varepsilon_{ijt})} \quad (2)$$

where, besides the variables that we have already defined, ε_{ijt} is a random error term with standard properties.

As the literature has highlighted, the choice of the estimator with which to estimate Eq.2 is not trivial given the intertwining of a set of problematic issues, which we will try to address as in the following.

3.2.1 Zero trade flows and heteroskedasticity: OLS vs. PPML estimators

As we have noticed above, an appreciable number of bilateral trade data of our sample are zeros. A standard OLS log-linearised specification, yielding the values of the migration elasticities, would only be estimated on positive trade values, posing a problem of selection bias. Among the alternative ways proposed to overcome this problem, the adoption of the Poisson Pseudo-Maximum

Likelihood estimator (PPML) would appear the most suitable (Santos-Silva and Tenreyro, 2006), given its estimation of the dependent variable in levels, retaining any non-negative value of it, including zero.[16] Furthermore, as also noted by Santos-Silva and Tenreyro (2006), the use of dependent variables in levels (Bosquet and Boulhol, 2010) is preferable when the error terms are heteroskedastic, as it might occur in our application. Log-linearizing the gravity equation to estimate it by OLS would actually introduce a bias in this case (see also Manning and Mullahy, 2001).[17] Finally, the PPML estimator has been shown to be consistent even by over-or under- dispersion (Wooldridge, 2002).

On the other hand, Head and Mayer (2014) have recently shown that relatively common misspecifications of the conditional mean, such as taking as linear an effect that is actually non-linear, can lead to severe bias in the PPML estimates too, due to the higher weight posed on larger observations by this estimator. In this case, the more flexible distributional assumptions of the Gamma PML (GPML) would be more suitable.

In front of the choice between a Poisson and a Gamma PML estimator, a rigorous analysis of the underlying distribution of the errors is of fundamental importance for an accurate estimation of Eq.2. In order to proceed with such an evaluation, the same equation will be estimated first by OLS (Eq.3) – taking

[16]Previous studies addressed the issue by opting for a Tobit model, with an arbitrary zero or an estimated threshold (Wagner et al., 2002; Herander and Saavedra, 2005). In more recent work, Santos Silva and Tenreyro (2011) highlighted the consistency of the PPML estimator even when the share of zeros is substantial.

[17]Indeed, a violation of the homoskedasticity assumption will in general lead to the fact that, in the log-linear transformation of the gravity model, the expected value of the log-linearized error term depends on the covariates. In other words, the conditional mean of the log of the errors will depend on both their mean and on the higher-order moments of their distribution, which by heteroskedasticity will be correlated with the covariates, leading to inconsistent OLS estimates.

the natural logarithm of the sole positive values of the dependent variable, $ln(X_{ijt})$ – and then by PPML and GPML (Eq.4) – employing the trade values in levels X_{ijt}, including zeros, and their strictly positive values, $e^{(ln(X_{ijt}))}$, respectively, assuming that the errors are, respectively, Poisson and Gamma distributed.

$$ln(X_{ijt}) = \beta_1 ln(Y_{it-1} \times Y_{jt-1}) + \beta_2 ln(\mathrm{Dist}_{ij}) + \beta_3 \ln(\mathrm{Immi}_{ijt-1} + 1) + \\ + \beta_4 \ln(\mathrm{Emi}_{ijt-1} + 1) + \beta_5 \mathrm{NI}_{ijt-1} + \beta_6 \mathrm{NE}_{ijt-1} + \omega_{\mathbf{it}} + \theta_{\mathbf{jt}} + \eta_{\mathbf{rj}} + \varepsilon_{ijt} \quad (3)$$

$$X_{ijt} = e^{\beta_1 ln(Y_{it-1} \times Y_{jt-1}) + \beta_2 ln(\mathrm{Dist}_{ij}) + \beta_3 \ln(\mathrm{Immi}_{ijt-1} + 1) + \beta_4 \ln(\mathrm{Emi}_{ijt-1} + 1) + \beta_5 \mathrm{NI}_{ijt-1} + \beta_6 \mathrm{NE}_{ijt-1} + \omega_{\mathbf{it}} + \theta_{\mathbf{jt}} + + \eta_{\mathbf{rj}} + \varepsilon_{ijt}} \quad (4)$$

A Park test for heteroskedasticity will be then performed on the OLS estimates. Should this reveal evidence of heteroskedasticity, the choice of the estimator will be then performed through the diagnostic “MaMu test”, discussed by Head and Mayer (2014), drawing on Santos-Silva and Tenreyro (2006) and Manning and Mullahy (2001), which we apply for the first time to actual data in the analysis of the migration-trade link. Focusing on the relationship between the variance and the conditional mean of the residuals – $\mathrm{var}[X_{ij}|\mathbf{z}_{ij}] = h\mathrm{E}[X_{ij}|\mathbf{z}_{ij}]^{\lambda}$, where $\mathbf{z}_{ij}$ is the vector of covariates – the test estimates the following equation:

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

Equation 5 is estimated by OLS when applied to OLS residuals, by PPML when applied to the PPML residuals and by GPML when applied to the GPML residuals (Santos-Silva and Tenreyro, 2006; Manning and Mullahy, 2001). In the obtained estimates, values of λ close to 2 would reflect 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 thus the homoskedastic OLS on logs – which is the MLE, if the homoskedasticity assumption is reasonable – and the Gamma PML.[18] In this same situation, according to Weidner and Zylkin (2018), the Gamma PML with three-way fixed effects will also be consistent. If λ is instead closer to 1, generalizing the Poisson distributional assumptions (Manning and Mullahy, 2001), the Poisson PML should be preferred, as OLS will suffer from a heteroskedasticity bias and Gamma PML from an incidental parameters problem.

Finally, in order to corroborate the choice of the estimator based on the previous test, we will run the Ramsey (1969)'s RESET-tests on each estimation method, aiming to detect possible misspecifications in the conditional means such as non-constancy in the covariates. All of these info will be retained in our final evaluation of the estimators (Santos-Silva and Tenreyro, 2006).

In concluding the discussion of the optimal estimator, it should be stressed that, in our context, its choice is further complicated by the inclusion of three-

[18] This, and their similar first-order conditions, explain why the Gamma and OLS estimates are often quite similar (Head and Mayer, 2014).

way fixed effects. Until recently, given its high computational burden, a consistent estimation of three-way FE models was only possible by OLS. With respect to the PPML, these limitations have however been recently overcome by Correia et al. (2019b) with the `ppmlhdfe` algorithm implemented in Stata. Similarly, the `gpmlhdfe` algorithm recently developed by Weidner and Zylkin (2018) has permitted to overcome the even more substantial computational problems that were affecting the estimation of FE models by Gamma PML. [19]

3.2.2 Endogeneity and instrumental variable approach

A second issue that affects our econometric strategy relates to possible problems of endogeneity in the migration variable. Even if we include large sets of fixed effects, omitted bilateral time-varying variables, as well as reverse causality, may in principle affect our estimates. Indeed, people could migrate in response to established trade routes, rather than the other way round. In this respect, it could be argued that migration flows are more exogenous than trade. As noticed by Gould (1994), migration is generally driven by more structural factors than trade, like family reunifications, wage differentials, and pre-existing co-ethnic communities: an argument that has found ample confirmation afterward (Munshi, 2003; Mayer, 2004; Jayet and Ukrayinchuk, 2007). Furthermore, our sub-national approach should attenuate the strength of the two endogeneity sources. Still, the problem cannot be ruled out *a priori* and further action is needed to confirm that the results can be interpreted as causal

[19] On the other hand, Weidner and Zylkin (2018) have recently shown that the consistency of the Gamma PML with fixed effects (Greene, 2004) only applies if the underlying distribution of the errors satisfies the assumption that the conditional variance equals the square of the conditional mean. When this assumption is not satisfied, the Gamma PML is found to suffer from an incidental parameters problem, unlike the Poisson PML.

effects.

Following the extant literature, we will then address this issue by searching for appropriate instrumental variables for our two focal regressors (immigration and emigration) and complement our core econometric strategy with a 2SLS procedure. To instrument for immigration stocks, we can benefit from an already standard procedure in the labor economics and economic geography literature (see, e.g., Card, 2001; Ottaviano and Peri, 2006). Considering that immigrant settlements are very persistent over time, due to their tendency to direct in enclaves where previous co-nationals have already settled, past immigration can be used to impute its current stocks (Card, 2001). In particular, the pre-determined distribution of immigrants across provinces can be used to build up weights for each country-province pair ij, and these weights then multiplied by the overall migration stocks from country j to Spain in year t.[20]

As regards emigration, however, the lack of historical data on Spain prevents us from constructing a similar instrument for emigration and requires an alternative procedure. Reversing the logic of the recent works by Beine and Coulombe (2018) and Basile et al. (2018),[21] we propose an original procedure that imputes bilateral emigration employing data on residential cancellations (available from INE disaggregated by province of cancellation or by country of new residence) to measure "push" and "pull" factors of emigration, as in Burchardi et al. (2018). On the one hand, we take the overall cancellations from province i to any other country to account for the province-specific fac-

[20]This is the procedure recently followed by Bratti et al., 2014.

[21]Their proposal is to instrument the overall flows of immigrants (from any countries) within a given province by aggregating the bilateral flows of migrants estimated through a gravity model of international migration.

tors that "push" emigration away from it.[22] On the other hand, the overall cancellations from any Spanish province to country j are taken to account for the factors that "pull" Spanish emigrants as a whole towards it.[23]

Denoting with w_{it} the share of residential cancellations from province i over total cancellations at time t, and with w_{jt} the share of residential cancellations directed to country j over total cancellations still at time t, we use w_{it} and w_{jt} to reweigh the total stocks of emigrants Emi_t (from any province, to any country, in year t). Furthermore, in order to avoid perfect multicollinearity between the imputed measurement and the time-varying fixed effects, and to mitigate concerns arising from the computation of the overall stocks based on possibly endogenous bilateral stocks, we subtract bilateral stocks of emigrants from Emi_t, and instrument emigration stocks as follows:

$$\mathrm{Emi}_{ijt}^{\mathrm{imputed}} = w_{it} w_{jt} (\mathrm{Emi}_t - \mathrm{Emi}_{ijt}) \tag{6}$$

While our instrument for emigration includes both "push" and "pull" factors, data constraints impede us to integrate any of the "recursive" factors that Burchardi et al. (2018) use, accounting for persistent emigrant settlements from a specific province to a specific country over a long period of time. Yet, drawing on the panel structure of our data, as in all our main specifications, we include in our 2SLS approach the three sets of province-year, country-year and region-country effects (along with the products of per-capita GDPs and distance). Hence, time-invariant ties between specified region-country dyads

[22] In this respect, the relative size of the (unobserved) cancellations targeting country j should be negligible relative to overall cancellations from i.

[23] Again, we assume that the relative weight of the unobserved bilateral flows over total emigration to j will be negligible.

are accounted for.

4 Results

4.1 The pro-export effect of migration

Following our econometric strategy, Table 3 reports the results of the three "candidate" estimators of the gravity model, by including both immigrants and emigrants, and by allowing for heterogeneous MRT at the province level.[24]

Starting with the building blocks of the gravity-model, the product of per-capita GDPs is collinear with the province-year and country-year fixed effects and is thus omitted.[25] As for the distance variable, it has the expected negative effects in the OLS and GPML estimates, while its effect is insignificant, conditional on the bilateral region-country effects, according to the Poisson estimates.

Coming to our focal migration variables, their point estimates are similar across the different estimators and specifications.

[24] All the inference reported in this section is based on standard errors that are clustered at the pair level. The results are robust to multi-way clustering of the errors at the province-year, country-year and region-country levels.

[25] In practice, the regression output of some applications—it is not our case—may still yield a coefficient for this variable, but this will refer to the coefficient for the reference category of the regression, which will typically not be under the analyst's control when high dimensional fixed effects are included. Hence, Head and Mayer (2014) recommend that the inclusion of such terms be avoided.

Table 3: Pro-export effect of immigrants and emigrants: sub-regionally heterogenous MRT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		PPML			OLS			GPML	
$ln(\text{Immi}_{ijt-1}+1)$	0.077**		0.053	0.064***		0.052**	0.036*		0.023
	(0.039)		(0.037)	(0.022)		(0.022)	(0.018)		(0.019)
NI_{ijt-1}	-0.104		-0.133	0.046		0.022	-0.022		-0.047
	(0.092)		(0.091)	(0.068)		(0.068)	(0.054)		(0.054)
$ln(\text{Emi}_{ijt-1}+1)$		0.115***	0.096**		0.103***	0.091***		0.101***	0.097***
		(0.041)	(0.038)		(0.026)	(0.027)		(0.023)	(0.023)
NE_{ijt-1}		0.005	-0.000		0.043	0.037		0.064*	0.063*
		(0.067)	(0.068)		(0.043)	(0.043)		(0.036)	(0.036)
$ln(\text{Dist}_{ij})$	-0.117	0.009	0.022	-1.427***	-1.323***	-1.386***	-1.020***	-0.927***	-0.952***
	(0.350)	(0.362)	(0.363)	(0.364)	(0.363)	(0.361)	(0.319)	(0.319)	(0.315)
N	54,250	54,250	54,250	50,210	50,210	50,210	50,210	50,210	50,210
province-year effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
country-year effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
region-country effects	yes	yes	yes	yes	yes	yes	yes	yes	yes

Standard errors clustered at the province-country level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

When included separately, both immigrants and emigrants positively and significantly affect provincial exports.[26] On the other hand, results change when they are included jointly, as we claim they should. With the sole exception of the OLS estimates a significant and positive effect on provincial Spanish export is revealed only by their emigrants, as that of immigrants turns out non-significant. More precisely, a 10% increase in the emigration stock towards a certain country increases exports to it by about 0.9% on average, across all the different estimators.

This is a first interesting result of our application, and indeed an original one with respect to previous studies at the country-level, which have found that the "impact of emigrant networks on exports coexists with a positive and significant impact of immigration on exports", and that "the elasticity of manufacturing exports to emigration is robust and of similar size as the effect of immigration." (Hiller, 2014). As we will say in the following, different conceptual interpretations can be provided for this result. In the meantime, the following two sections will show how its difference with respect to the extant literature could also be due to the methodological choices that we deem to better capture the phenomenon.

4.1.1 Do sub-national heterogenous MRT matter?

As we have claimed in Section 2.3, working with heterogeneous exporting capacity across provinces appears recommendable for different reasons. Yet,

[26]Importantly, the positive yet comparatively small correlation between immigration and emigration stocks implies that the coefficient of each of these variables is somewhat overestimated when they are included separately, supporting our approach of including both directions of migration. The insignificant or only mildly significant effects of NI_{ijt-1} and NE_{ijt-1} also reassure us that having added one unit to the migration variables does not alter the results.

given the relatively small size of Spanish provinces, one may question whether this theoretically-founded yet quite demanding approach yields statistically different implications from ones that assume the exporting capacity to be homogeneous country-wide or within the same NUTS2 region. In order to address this issue, we compare our previous results with those of two alternative specifications. In the first one, sub-national MRT are assumed equal across provinces pertaining to the same region, similarly to Bratti et al. (2014). In a second specification, any subnational heterogeneity in the exporting capacity is ruled out, while region-country FE and country-year FE are still included. The corresponding PPML, OLS, and GPML estimates are reported in the upper and bottom panels of Table 4, respectively.

As we did expect, the assumption of homogeneous export capacities among the provinces of the same region (upper panel of Table 4) strongly overestimates the effects of both our focal variables.[27] Across the three estimators, both the immigration and the emigration coefficients increase. This is quite an important change, suggesting a correlation between bilateral stocks of migrants and the exporting capacity of provinces, which are omitted when the sub-regional heterogeneity is assumed out. Indeed, some provinces concentrate most of the exporting capacity, as well as most of immigrants' and emigrants' stocks. Allowing for heterogeneous exporting capacity at the relatively aggregated NUTS2 level, indeed, may confound these effect. This interpretation is supported when looking at the coefficient for distance, which now becomes positive and significant.

[27] Omitting the province-year effects, the product between province and country per-capita GDPs can now be estimated and results positive and significant.

Table 4: Pro-export effect of immigrants and emigrants: alternative FE specifications

	PPML	OLS	GPML
$ln(Y_{jt-1} \times Y_{it-1})$	5.730***	7.182***	6.331***
	(0.375)	(0.230)	(0.192)
$ln(\text{Immi}_{ijt-1}+1)$	0.194***	0.323***	0.275***
	(0.050)	(0.023)	(0.017)
$ln(\text{Emi}_{ijt-1}+1)$	0.639***	0.615***	0.571***
	(0.046)	(0.028)	(0.024)
NI_{ijt-1}	-0.271**	-0.224***	-0.132**
	(0.117)	(0.078)	(0.062)
NE_{ijt-1}	0.267***	0.160***	0.203***
	(0.081)	(0.053)	(0.044)
$ln(\text{Dist}_{ij})$	1.170**	0.158	1.181***
	(0.460)	(0.422)	(0.397)
N	54,250	50,210	50,210
province-year effects	no	no	no
country-year effects	yes	yes	yes
region-country effects	yes	yes	yes
region-year effects	yes	yes	yes
$ln(Y_{jt-1} \times Y_{it-1})$	5.389***	6.866***	5.958***
	(0.360)	(0.223)	(0.189)
$ln(\text{Immi}_{ijt-1}+1)$	0.191***	0.322***	0.277***
	(0.050)	(0.023)	(0.017)
$ln(\text{Emi}_{ijt-1}+1)$	0.636***	0.614***	0.565***
	(0.046)	(0.028)	(0.024)
NI_{ijt-1}	-0.265**	-0.227***	-0.134**
	(0.117)	(0.078)	(0.062)
NE_{ijt-1}	0.248***	0.154***	0.186***
	(0.080)	(0.053)	(0.045)
$ln(\text{Dist}_{ij})$	1.128**	0.089	1.114***
	(0.456)	(0.419)	(0.393)
N	54,250	50,210	50,210
province-year effects	no	no	no
country-year effects	yes	yes	yes
region-country effects	yes	yes	yes
region-year effects	no	no	no

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

In Spain, the regional capitals are typically located in the more geographically central provinces within their regions, hence they are on average more distant to any trading partner. Unsurprisingly then, when omitting the province-year effects, more distant provinces within their regions turn out to export more[28].

Similar results are obtained when the region-time effects are excluded all together (lower panel of Table 4), which corresponds to an assumption of homogeneous exporting capacity throughout Spain. Somehow surprisingly, when the province-year effects are omitted, the results are similar, irrespectively from region-time effects are included or not. Once more, this suggests that the main source of heterogeneity is actually to be found at a quite refined geographic level, like the NUTS3 we are considering.

Overall, our assumption of sub-national heterogeneous MRT does actually make a difference in the appreciation of trade effects of migration. In particular, the pro-export effects of emigrants we have detected as our main result gets apparently inflated by homogeneous MRT, as the relative coefficient increases by between almost 7 and almost 9 times.[29]

4.1.2 Picking up the right estimator

Before searching for the estimator that better suits the current structure of our data, let us observe that, according to Head and Mayer (2014), a scenario like the one presented in Table 3, in which the three estimators yield largely similar results, is reassuring: our models do not signal major misspecification

[28]Similar results are obtained when omitting the migration variables from the specification.

[29]The immigration coefficient instead increases by between 4 and 6 times.

concerns. In particular, including or excluding zero trade flows leaves the results virtually unaffected.

Still, as we mentioned above, the log-linear OLS estimates may suffer from a heteroskedasticity bias. Indeed, a standard Park regression of the squared OLS residuals on the covariates (Table 5) confirms this suspect: the variance of the residuals actually increases with both immigrants and emigrants stocks, and decreases with distance; furthermore, the variance is also on average smaller for provinces with no immigrants. In principle, this implies that Poisson and Gamma PML estimators should be preferred to OLS when estimating the effects of migration on export with our data. This is actually confirmed by the "MaMu test" we have discussed in Section 3.2.1.

As Table 6 shows, the estimated value of λ in Equation 5 is about 1.7 with respect to OLS, about 1.3 with respect to PPML, and 1.9 with GPML. In the latter case, the confidence intervals for the $\hat{\lambda}$ includes 2. This implies that, provided that the conditional mean is well specified, the high-dimensional fixed-effects Gamma model that we implemented does not suffer from an incidental parameter problem, similarly to the fixed-effects Poisson (Weidner and Zylkin, 2018). Moreover, while the $\hat{\lambda}$ for OLS and PPML are neither precisely 2 nor precisely 1, they can be regarded to reasonably well satisfy the distributional assumptions of the estimators they are drawn from. According to these estimates, the GPML would seem to be the most efficient estimator for our data; on the other hand, because the $\hat{\lambda}$ estimated for the OLS and PML residuals are both significantly below 2, the PPML estimator should be preferred over the OLS (see Head and Mayer, 2014). In brief, the results of the MaMu test would support the implementation of either the PPML or the GPML.

Table 5: Park test

	(1)
$ln(\text{Immi}_{ijt-1} + 1)$	0.291***
	(0.038)
$ln(\text{Emi}_{ijt-1} + 1)$	0.540***
	(0.051)
NI_{ijt-1}	-0.348***
	(0.112)
NE_{ijt-1}	-0.062
	(0.078)
$ln(\text{Dist}_{ij})$	-3.311***
	(0.795)
N	50,210
province-year effects	yes
region-country effects	yes
country-year effects	yes

Standard errors clustered at the province-country level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In the bottom panel of Table 6, we report the coefficients and p-values associated with a set of Ramsey (1969) RESET-tests on each estimation method, again following Santos-Silva and Tenreyro (2006). The null hypothesis of the correct specification of the conditional mean cannot be rejected in the case of the Poisson PML, while it is rejected in the case of the OLS estimates and in that of the Gamma PML too[30].

[30] We leave to future research the further exploration of the sources of mis-specification in the GPML model, which are apparently not driven by the incidental parameters problem (Greene, 2004; Weidner and Zylkin, 2018) nor by functional form mis-specification. According to Santos-Silva and Tenreyro (2006), the reason may be found in the larger weight that the GPML gives to smaller observations.

Table 6: MaMu and RESET tests

Model	PPML residuals[a]	OLS residuals[b]	GPML residuals[b]
Manning and Mullahy (MaMu) test on the underlying distribution of the errors			
$ln(\hat{\mu})$	1.298***	1.697***	1.904***
	(0.047)	(0.005)	(0.059)
RESET-tests			
Squared linear prediction	-0.007	-0.020	-0.216
P-value	0.2418	0.001	0.000
N	54,250	50,210	50,210

Heteroskedasticity-robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

In conclusion, on the basis of the previous diagnostic test, we feel confident in selecting the Poisson PML estimator as the most suitable one for addressing our focal issue. Sticking to this last estimator, we restate our main results so far as follows. Conditional on emigrant stocks, immigrants into Spanish provinces do not significantly increase their exports towards the respective home countries. On the other hand, we detect a positive and significant effect on export only of emigrants. Conditional on immigration stocks, a 10% increase in the emigration stocks of Spanish provinces increases their export to their origin countries by almost 1% (Table 3). Because the emigrants' effects on export incorporate both a network effect (information and enforcement) and a preference effect, this magnitude actually compares with previous estimates of the immigrants effects on imports, which are generally larger than those for exports and would lead to a corresponding increase of about 1.5% (see Genc et al., 2012). Hence, compared with these studies, our estimates are comparable but relatively small. As mentioned, this relatively conservative result is found by allowing for differential exporting capacity of provinces. Erroneously ruling out this heterogeneity would lead to much larger estimates of both the immigrants' and emigrants' effects (Table 4).

As we have anticipated, the result we found about the exclusive pro-export effect of emigrants is original and amenable of different interpretations. First of all, along a negative phase of the business cycle like the one we are considering, the relative incidence of emigration flows compared to immigration ones can make the former the unique channel through which trade-sensitive business network can be effectively built up. While more refined data would be needed to ascertain it with accuracy, in the same period, the demand for home-country products (preference effect) expressed by emigrants may be the greatest, if not the only relevant channel through which migrants can affect trade. In brief, due to the multiple sources of a pro-export effect, emigrants are perhaps not surprisingly more effective in promoting exports than immigrants.

Before taking these results as conclusive, two further steps are required to ensure that our estimates are consistent: address possible remaining sources of endogeneity; and study whether there is significant heterogeneity in the migrants' effects, which could challenge the underlying assumption of constant elasticity in the Poisson PML model (Head and Mayer, 2014).

4.1.3 Instrumenting migration stocks

Using the instrumentation methodology discussed in Section 3.2.2, Table 7 reports the results of our baseline model following a standard instrumental

Table 7: 2SLS estimates

	First Stage		Second Stage
Dep. var.	$ln(\text{Immi}_{ijt-1}+1)$	$ln(\text{Emi}_{ijt-1}+1)$	$ln(\text{X}_{ijt})$
$ln(\text{Immi}^{\text{imputed}}_{ijt-1}+1)$	0.136***	0.053***	
	(0.011)	(0.007)	
$ln(\text{Emi}^{\text{imputed}}_{ijt-1}+1)$	0.029	0.705***	
	(0.041)	(0.023)	
$ln(\text{Immi}_{ijt-1}+1)$			0.032
			(0.100)
$ln(\text{Emi}_{ijt-1}+1)$			0.327***
			(0.085)
$ln(\text{Dist}_{ij})$	1.215***	-0.264	-1.262***
	(0.373)	(0.262)	(0.388)
N	49,389	49,389	49,389
F-statistic	76.97	593.92	
province-year effects	yes	yes	yes
country-year effects	yes	yes	yes
region-country effects	yes	yes	yes

Standard errors clustered at the province-country level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

variable/2SLS estimation.[31]

At the outset, the F-statistics of our first stage regressions allow dismissing concerns about the potential weakness of the instruments. This is a common result for instrumental variables such as our Altonij-Card instrument for immigration. More importantly, it is also reassuring about the strength of our less standard instrument for emigration, which actually results very strong. Our

[31]Data for the instrumental variables are available for a slightly smaller subset of trade pairs than the official trade data, which leads to a small reduction in the number of observations. Results for this smaller set of dyads are fully robust and are thus unreported.

equation is just identified, so we have to assume that the exclusion restrictions hold for our instruments.

The second stage estimates confirm the results of the Poisson PML and highlight an even larger role for emigrants than we identified in the baseline estimates, similarly to Bratti et al. (2014, 2018)[32]. A possible explanation they suggest for this result is a measurement error, which could apply here too: as discussed, not all Spanish nationals residing abroad appear in the electoral registries of their host counties, so that our emigration variable likely underestimates the actual stocks. By contrast, even if not disaggregated by province-country dyad, data on residence cancellations may more accurately represent the movements of Spanish nationals abroad than the electoral registries. Overall, the results of Table 7 suggest that the positive effect of emigrants on trade detected in the previous section is robust to endogeneity concerns.

4.2 Sources of heterogeneity in the pro-export effect of migration

As the extant literature has largely shown (e.g. Wagner et al., 2002; Girma and Yu, 2002), the trade effect of migrants usually interacts with standard trade determinants: institutional and language commonality above all. As for the former, previous studies have argued and found[33] that migrants affect trade more in the presence of large institutional distance between partners – e.g. arbitration tribunals of different quality and/or different rules stimulating

[32]Interestingly, the 2SLS estimate of our emigrants' effect is closer than our baseline estimate to the immigration elasticities of imports detected by Bratti et al. (2014).

[33]For example, Girma and Yu (2002) with respect to the UK, Dunlevy (2006) with respect to the US, and Briant et al. (2014) with respect to French departments.

predatory behaviors on both sides – as this is the kind of obstacles they can be expected to attenuate through the enforcement mechanisms of their social and business networks (Briant et al., 2014); conversely, the trade effect of migrants would diminish/vanish in front of institutionally similar partners, with respect to which the costs of starting a trade relationship would be lower (Peri and Requena-Silvente, 2010).

Similar considerations were argued (Dunlevy, 2006; Briant et al., 2014) for the role of language commonality: the lack of a common linguistic and cultural background should make the role of migrant more significant in reducing the entailed transaction and fixed costs of trade.

As a way to test the previous arguments with respect to our sample, in the following we report the estimates of our gravity model by inserting the interaction terms between its focal variables, $ln(Immi_{ijt})$ and $ln(Emi_{ijt})$, and two mutually exclusive dummies for institutional similarity (proxied by EU membership, EU), and its complement to one (NEU), and for language commonality ($Spanish-Speaking$), and its complement to one ($non-Spanish-Speaking$) (this is line with Girma and Yu, 2002).[34]

The results of the PPML estimates are reported in Table 8, while we still report the OLS and GPML estimates as benchmarks for comparison and for robustness checks in the Appendix (Table A.12.)

[34]In so doing, we instead omit dealing with the complexity of the traded goods. This would have entailed a further effort of data mining, which we have decided to postpone to future research, after having singled out the main implications of our empirical methodology. Hopefully, this could already emerge even without dealing with the issue of good complexity.

Table 8: Sources of heterogeneity in the migrants' effects

Institutional Similarity		Language Similarity		Geographic Proximity	
$ln(\text{Immi}_{ijt}^{EU})$	0.061	$ln(\text{Immi}_{ijt}^{Spa})$	0.125	$ln(\text{Immi}_{ijt-1}+1)$	0.086*
	(0.045)		(0.103)		(0.052)
$ln(\text{Immi}_{ijt}^{non-EU})$	0.029	$ln(\text{Immi}_{ijt}^{non-Spa})$	0.048	$ln(\text{ImmiOut}_{ijt-1})$	0.601
	(0.040)		(0.037)		(0.492)
$ln(\text{Emi}_{ijt}^{EU})$	0.032	$ln(\text{Emi}_{ijt}^{Spa})$	0.163*	$ln(\text{Emi}_{ijt-1}+1)$	0.138***
	(0.045)		(0.089)		(0.042)
$ln(\text{Emi}_{ijt}^{non-EU})$	0.168***	$ln(\text{Emi}_{ijt}^{non-Spa})$	0.087**	$ln(\text{EmiOut}_{ijt-1})$	1.221**
	(0.043)		(0.039)		(0.618)
NI_{ijt-1}	-0.117	NI_{ijt-1}	-0.141	NI_{ijt-1}	-0.110
	(0.090)		(0.092)		(0.099)
NE_{ijt-1}	0.055	NE_{ijt-1}	-0.013	NE_{ijt-1}	0.003
	(0.071)		(0.068)		(0.066)
$ln(\text{Dist}_{ij})$	-0.022	$ln(\text{Dist}_{ij})$	0.002	$ln(\text{Dist}_{ij})$	-0.018
	(0.362)		(0.363)		(0.356)
N	54,250	N	54,250	N	54,250
province-year effects	yes	province-year effects	yes	province-year effects	yes
country-year effects	yes	country-year effects	yes	country-year effects	yes
region-country effects	yes	region-country effects	yes	region-country effects	yes

PPML estimates. Standard errors clustered at the province-country level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2.1 Institutional similarity and language commonality

With respect to institutions, the left panel of Table 8 shows that, consistently with previous literature, the pro-export effect of emigrants is unambiguously driven by emigrants towards extra-EU countries, and consistently so across estimators.[35]

[35] In line with our baseline results (Table 3), OLS estimates also highlight a positive and significant effect for immigrants this time, with no significant differences apparently driven by the EU membership of the partner country.

With respect to language communality, instead, the mid panel of Table 8 shows that emigrants towards Spanish-speaking countries increase exports towards them to a greater extent – about twice as much – than emigrants towards non-Spanish speaking ones. The difference, however, is not significant. Yet, the relative magnitudes of the effects hold across estimators. This positive pro-export effect holds even if the export enabling role of a common language is already captured by the fixed effects of the model.

While not conclusive, these results suggest that, in our sample, institutions and language do not represent the two sides of the same coin in transnational networking, and that, unlike common institutions, language commonality adds up to the emigrants' ability to promote export. Drawing on the random encounter model proposed by Wagner et al. (2002),[36] we could think that, while the trade barriers may be lower, language commonality eases the access to information about the home or host country opportunities. In short, language commonality increases the probability that an emigrant has the capacity to facilitate the correspondent export flow.

Overall, while some heterogeneity in the effects emerges, the results of our estimators are similar to each other and do not raise substantial concerns that the larger weight given by the PPML to larger observations is biasing the results.

[36] In brief, this model investigates the probability that, given a set of realizable trade opportunities, the immigrant actually realizes them.

4.2.2 The geography of migration networks: direct and indirect migration effects

As we said in Section 2, migrants could affect exports at the sub-national level of analysis both within the relevant province and outside of it (Herander and Saavedra, 2005; Bratti et al., 2014). Furthermore, geographical proximity should make a localized effect more likely than a wide-ranging one. In order to investigate this argument, in Table 8 we report the results of one specification of our gravity model, which integrates two additional variables: $ln(ImmiOut_{ijt})$, measuring the total stock of immigrants from country j living in provinces other than i; and $ln(EmiOut_{ijt})$, referring to the total stock of emigrants registered in provinces other than i, who had migrated to country j. According to the argument at stake, these variables could account for extra-province networks of immigrants and emigrants, on which a focal province could draw for detecting and exploiting new export opportunities.

Table A.13 shows that geographical proximity actually matters in conveying the pro-export effects of emigration networks. Somewhat in line with the findings by Herander and Saavedra (2005) on the US, a weakly significant pro-export effect of immigrants does emerge for province-level networks. On the other hand, the emigrants' effect seems driven by both a localized and a country-wide component. This is an interesting and important specification of our main result about the pro-export effect of emigrants. As argued by Rauch (2001), the exchange of trade-relevant information occurs mainly within networks of proximity. Indeed, this allows the exchange also of a tacit and embodied kind of trade-related information. On the other hand, a significantly positive and much larger effect emerges from $ln(EmiOut_{ijt-1})$. This result suggests that the exports of a given province i to a country j rely not only on the pro-trade effects of emigrants from those provinces, but also and above all from all other provinces. Just to make an example, not only is the export of the Alicante province to China affected by a larger network of Alicantinos moving to China, but also by larger stocks of emigrants moving to China from any other Spanish province than Alicante. This effect could be driven by a composite network effect, operating between migrants from different provinces meeting in the same foreign country; or by the preference the expatriates of a given province have for Spanish products as a whole, including from provinces other than that of their origin. Also in this last respect, both preference and information effects could be at play and a distinction between the two is unfortunately impossible with our available data. One way or the other, these out-of-province emigrants could help province i to span and diversify the spectrum of its international activities towards their destination country j: precisely because they refer to other local contexts of emigration departure.

In front of what is, to the best of our knowledge, the first application of the proximity argument to the emigration side, this is for sure a tentative interpretation, which is need of further evidence and theoretical reflection. While other interpretations could be put forward, the different scale of the networks through which immigrants and emigrants exert their pre-export effect in the case of Spanish provinces – as we said, local and non-local, respectively – represents an important result on which future research should focus.

4.2.3 Migrants' effects over time

The dataset we use for our empirical analysis starts with the burst of the global financial downturn, usually identified in the late 2008, and extends over the following recession period. This crisis has heavily affected Spanish exports, which faced a substantial drop in 2009, and has been accompanied by a rise in the unemployment rates throughout Spain. It is reasonable to expect that these peculiar dynamics have affected the migrants composition, decreasing the incentives to stay in Spain and increasing those to expatriate, leading to a negative selection of the "stayers" and deteriorating, ultimately, their ability to facilitate trade. In Table 9, we provide an original way to address this issue, by enriching the three-way fixed effects PPML regression with a full set of interaction effects between each of our migration variables and each year of the sample[37] (Figure 5 plots the results of the PPML estimates graphically).

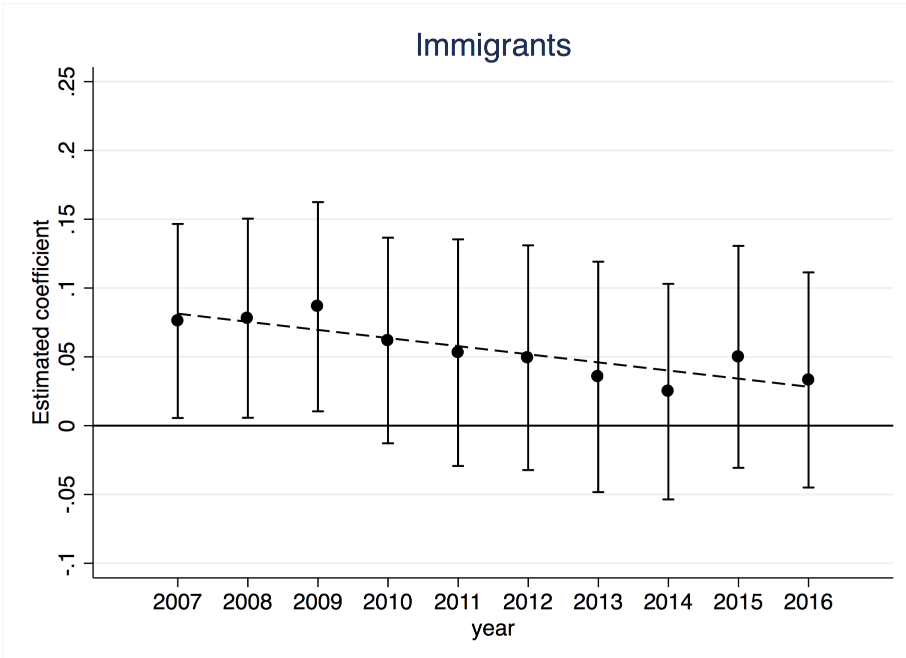
[37]It should be retained that the magnitude of the standard interaction effects in non-linear models does not equal the marginal effects of the interaction between the two (Ai and Norton, 2003). Therefore, for the ease of the interpretation, we do not report the standard main effect along with the interaction effects, but rather a set of mutually exclusive interaction terms for each of the years under scrutiny. The identified pattern corresponds to the one of the migration elasticities calculated algebraically on the basis of the estimated marginal effects of the interaction term for each year.

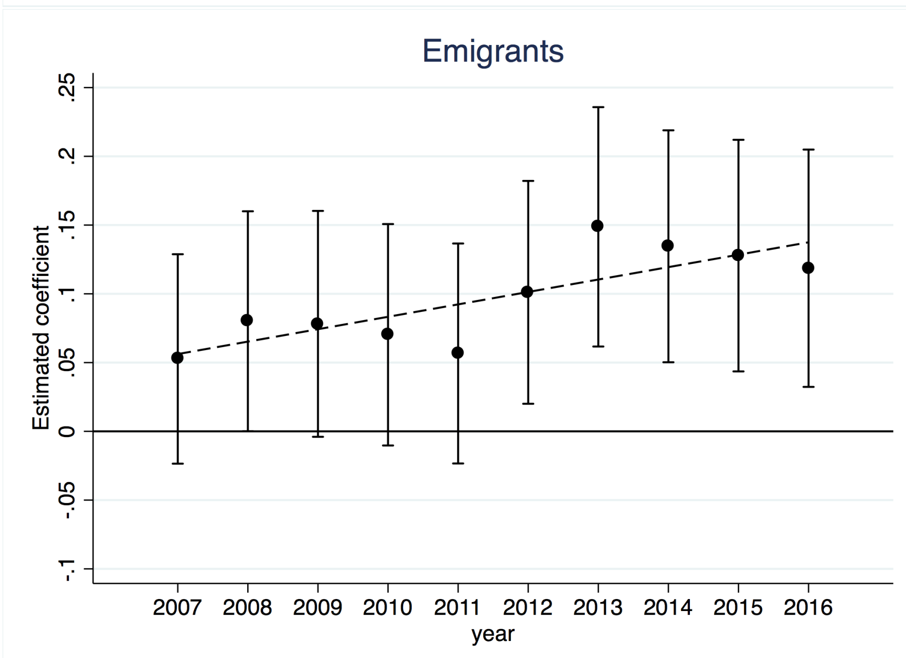
Table 9: Immigrants' effects over time

	Coef. (SE)		Coef. (SE)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2007)$	0.076** (0.036)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2007)$	0.053 (0.039)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2008)$	0.078** (0.037)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2008)$	0.080** (0.041)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2009)$	0.086** (0.039)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2009)$	0.078* (0.042)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2010)$	0.062 (0.038)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2010)$	0.070* (0.041)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2011)$	0.053 (0.042)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2011)$	0.057 (0.041)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2012)$	0.049 (0.042)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2012)$	0.101** (0.041)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2013)$	0.035 (0.043)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2013)$	0.149*** (0.044)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2014)$	0.025 (0.040)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2014)$	0.135*** (0.043)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2015)$	0.050 (0.041)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2015)$	0.128*** (0.043)
$ln(\text{Immi}_{ijt-1}+1) \times (\text{year}=2016)$	0.033 (0.040)	$ln(\text{Emi}_{ijt-1}+1) \times (\text{year}=2016)$	0.119*** (0.044)

$N = 54,250$. Included, but unshown covariates are: NI_{ijt-1}, NE_{ijt-1}, $ln(\text{Dist}_{ij})$, as well as province-year, country-year and region-country effects. PPML estimates Standard errors clustered at the province-country level in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 5: Time patterns of the migration effects





While coefficients do not result significantly different from each other, they display a visible pattern. In the case of immigrants, their pro-export effect is found to decrease and become insignificantly different from zero from 2010 onwards. The pro-trade effect of emigrants is instead increasingly positive over time, and more markedly so since 2012. These results are important in at least two respects. On the one hand, these results help reconcile our findings with those of previous studies and with the results by Peri and Requena-Silvente (2010) in particular. On the other hand, and relatedly, they suggest that a change in the composition of Spanish migrants could have actually occurred over time, following the stronger reaction to the crisis of the more qualified immigrants and emigrants, who are better able to facilitate the creation of a trade tie.

5 Conclusions

Although it represents a research issue with a long-standing tradition already, the analysis of the pro-trade effects of migrants at the sub-national level is still open to ameliorations. In particular, recent advances in the gravity model literature offers a battery of new analytical tools with which previous knowledge about the quantity and quality of these effects can be proven and possibly enriched. This is primarily the case of the heterogeneity of regional multilateral resistance terms, in addition to region-country dyadic time-invariant fixed effects and time-varying country-level effects. While the inclusion of time-varying effects for the traders 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. As we have argued, multiple motivations urge the inclusion of these controls, whose neglect, we have shown, entails serious distortions in the results. As we have also shown, additional important refinements can be obtained by taking stock of computational advances at the frontier of gravity models estimation via Pseudo Maximum Likelihood (PML) estimators with multi-way fixed effects in panel data (Head and Mayer, 2014). Through our diagnostic tests, implemented for the first time to the analysis of the migration-trade link and with panel data, we compared OLS, GPML and PPML estimators. The Poisson PML (PPML) estimator has emerged as the most suitable in the analysis of the pro-export effect of migrants for Spanish provinces. While the magnitudes of the PPML, the Gamma PML (GPML) and the OLS estimates have resulted comparable with each other, suggesting that the model is well specified and not substantially affected by the zeroes in the dependent variable, the OLS estimator was discarded on grounds of heteroskedasticity, with important implications for the obtained results.

By exclusively relying on the OLS estimates, we would have concluded that immigrants exert a significant pro-trade effect along with emigrants, consistently with previous studies. However, neither of our PML models supports such inference in the considered timeframe. Overall, our estimates instead suggest and robustly confirm a positive effect on the export of Spanish provinces of emigrants only. The identified magnitude of the effect implies that a 10% increase in the emigrant population from a given province to a certain country would increase its exports towards it by almost 1%. Disregarding the heterogeneous exporting capacity of provinces would have led to an undue overestimation of both immigrants' and emigrants' effects. When compared to the few previous studies that include both immigrants and emigrants in affecting exports, the exclusive effect of the latter represents an interesting result. This is supported when an original instrumentation approach is applied in search of a causal interpretation of the effects. Different interpretations could be provided for it. The particular phase of the business cycle and its impact on the immigrants/emigrants balance and composition have apparently eroded the immigrants' ability to promote trade.

Importantly, moreover, the insignificant immigrants' effects bear implications about the mechanism underlying the pro-trade effects. Indeed, the emigrants' effect could be either an information, an enforcement or a preference effect. Considering that no robust evidence could be found in support of the immigrants' effects, which is exclusively driven by the first two, our results may in principle reveal an exclusive preference effect, at least in the specific period that we considered. Omitting the outward side of migration would have made this distinction impossible and may have wrongly attributed to immigrants a role that actually accrues to emigrants. Further research may seek to confirm whether this is also the case during more positive phases of the cycle and, more generally, should investigate the relationship between business cycle and the migration-trade link.

Original results also emerge when the new methodological setting that we have proposed is applied to investigate the nuances of the trade-migration link. Consistently with previous literature, the effect of emigration results stronger in the trade linkage provinces establish with more institutionally distant countries, i.e. with non-EU countries in our setting. Such an institutional distance does actually represent a transaction cost, which the business networks created by emigrants can contribute to alleviate. On the contrary, an opposite result emerges with respect to language commonality, as Spanish emigrants to Spanish-speaking countries have a magnifying, rather than reduced, pro-export effect. This result suggests that, while language differences could also represent a trade bareer that emigrant networks could contribute to reduce, language commonality could be a leverage for better detecting and exploiting trade opportunities. In other words, among the potential trade opportunities that a migrant could facilitate, some could be lost due to language differences, which would have the effect of reducing the emigrant capacity to promote trade (cfr. the random encounter model in Wagner et al., 2002). Although the magnitudes of the effects for Spanish-speaking and non-Spanish-speaking countries, were not statistically different from each other, rendering our evidence not conclusive, this is an additional original result that deserves further scrutiny in future research.

Additional insights are obtained when, mimicking what has been previously done with respect to immigrants, for the first time – still to the best of our knowledge – the networks of expatriates is retained to affect provincial exports both through their local and nation-wide networks. Results show that not only both matter, but the existence of a network of Spanish expatriates in the same country is more important than that of their provinces of origin, with a very large elasticity, likely indicating that an increase in the emigrants' stocks triggers a demand effect that not only promotes the trade of their origin province, but of Spain as a whole. The intertwining of local and non-local effects of emigration on trade also represent a newly emergent piece of evidence, on which future research should concentrate.

Bibliography

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1):123–129.

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

Anderson, J. E. and van Wincoop, E. (2004). Trade costs. Working Paper 10480, National Bureau of Economic Research.

Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102:94–130.

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of international Economics*, 71(1):72–95.

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

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

Basile, R., De Benedictis, L., Durban, M., and Faggian, A. (2018). The Impact of Immigration on the Internal mobility of Natives and Foreign-born Residents: Evidence from Italy.

Beine, M. and Coulombe, S. (2018). Immigration and internal mobility in Canada. *Journal of Population Economics*, 31(1):69–106.

Bergstrand, J. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Review of Economics and Statistics*, 67(3):174—-181.

Bosquet, C. and Boulhol, H. (2010). 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., Benedictis, L. D., and Santoni, G. (2018). Immigrant entrepreneurs, diasporas and exports. *Journal of Regional Science*.

Bratti, M., De Benedictis, L., and Santoni, G. (2014). On the Pro–Trade Effects of Immigrants. *Weltwirtschaflisches Archiv/Review of World Economics*, 150:557–594.

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

Burchardi, K., Chaney, T., and Hassan, T. (2018). Migrants, Ancestors, and Foreign investments. *The Review of Economic Studies*, 86(4):1448–1486.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

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

Chaney, T. (2014). The network structure of international trade. *American Economic Review*, 104(11):3600–3634.

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

Correia, S., Guimarães, and Zylkin, T. (2019a). ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects.

Correia, S., Guimarães, and Zylkin, T. (2019b). Verifying the existence of maximum likelihood estimates for generalized linear models.

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.

Dunlevy, J. and Hutchinson, W. (1999). The impact of immigration on American import trade in the late nineteenth and twentieth centuries. *Journal of Economic History*, 59:1043–62.

Feenstra, R. (2004). *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton, New Jersey.

Felbermayr, G., Grossmann, V., and Kohler, W. (2015). Migration, international trade, and capital formation. *Handbook of the Economics of International Migration*, 1:913–1025.

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

Genc, M., Gheasi, M., Nijkamp, P., Poot, J., et al. (2012). The impact of immigration on international trade: A Meta-Analysis. *Migration impact assessment: New horizons*, 301.

Girma, S. and Yu, Z. (2002). The Link between Immigration and Trade: Evidence from the United Kingdom. *Weltwirtschaftliches Archiv/Review of World Economics*, 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.

Greene, W. (2004). Fixed effects and bias due to the incidental parameters problem in the Tobit model. *Econometric Reviews*, 23(2):125–147.

Hatzigeorgiou, A. (2010). Migration as Trade Facilitation: Assessing the Links between international trade and migration. *The B.E. Journal of Economic Analysis & Policy*, 10.

Head, K. and Mayer, T. (2014). Gravity Equations: Workhorse, Toolkit, and Cookbook. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*. Elsevier.

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

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

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

Hiller, S. (2014). The export promoting effect of emigration: Evidence from denmark. *Review of Development Economics*, 18(4):693–708.

Jayet, H. and Ukrayinchuk, N. (2007). La localisation des Immigrants en France: Une Première Approche. *Revue d'Economie Régionale et Urbaine*, pages 625–649.

Kaufmann, D., Kraay, A., and Zoido-Lobaton, P. (1999). Governance matters. Policy Research Working Paper 2196, The World Bank.

Larch, M., Wanner, J., Yotov, Y. V., and Zylkin, T. (2019). Currency unions and trade: A ppml re-assessment with high-dimensional fixed effects. *Oxford Bulletin of Economics and Statistics*, 81(3):487–510.

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

Mayer, T. (2004). Where do foreign firms locate in france and why? *EIB Papers*, 9(2):38–61.

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics*, 118(2):549–599.

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

Openshaw, S. (1983). *The modifiable areal unit problem.* Geo Books, Norwick.

Ottaviano, G. and Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, 6:9–44.

Ottaviano, G. I. P., Peri, G., and Wright, G. C. (2013). Immigration, offshoring, and american jobs. *American Economic Review*, 103(5):1925–59.

Parsons, C. and Vézina, P.-L. (2018). Migrant networks and trade: The vietnamese boat people as a natural experiment. *The Economic Journal*, 128(612):F210–F234.

Parsons, C. R. (2012). *Do migrants really foster trade? The trade-migration nexus, a panel approach 1960-2000.* The World Bank.

Peri, G. and Requena-Silvente, F. (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. (2001). Business and social networks in international trade. *Journal of Economic Literature*, 39(4):1177–1203.

Rauch, J. and Casella, A. (2001). *Networks and Markets.* Russel Sage Foundation, New York.

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

Santos-Silva, J. and Tenreyro, S. (2006). The Log of Gravity. *Review of Economics and Statistics*, 88:64–58.

Santos Silva, J. and Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222.

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

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

Weidner, M. and Zylkin, T. (2018). Bias and consistency in three-way gravity models.

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

A Appendix

Table A.10: **Main Data Sources**

Var.	Description	Source
X_{ijt}	Value of the exports from province i to country j in year t (thousands of €)	Datacomex, `http://datacomex.comercio.es/principal_comex_es.aspx`. Full database at 3-digit disaggregation requested and received by email.
Y_{jt}	GDP of Country j in year t (billions of US$)	IMF (International Monetary Fund) World Economic Outlook Database, `http://www.imf.org/external/pubs/ft/weo/2012/01/weodata/index.aspx`
Y_{it}	Gross product of province i in year t (thousands of €)	INE (*Instituto Nacional de Estadóística*) - "PIB a precios de Mercado precios Corrientes", `http://www.ine.es/jaxi/menu.do?L=0&type=pcaxis`
Immi_{ijt}, NI_{ijt}	Foreign residents with country j nationality residing in province i in year t	INE - "Población extranjera por sexo, comunidades y provincias y nacionalidad" `http://www.ine.es`
Emi_{ijt}, NE_{ijt}	Spanish expatriates registered in province i and residing in country j in year t	Censo Electoral de españoles residentes en el extranjero (CERA) por provincia de inscripcióón y paíís de residencia, `http://www.ine.es/ss/Satellite?c=Page&cid=1254735793323`
Dist_{ij}	Great-circle distance between the main center of province i and the capital of country j	Google Maps

Table A.11: **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.12: Robustness checks. OLS and Gamma PML estimates

	OLS	GPML
Institutional similarity		
$ln(\text{Immi}_{ijt}^{EU})$	0.049	0.022
	(0.032)	(0.031)
$ln(\text{Immi}_{ijt}^{non-EU})$	0.051**	0.022
	(0.024)	(0.020)
$ln(\text{Emi}_{ijt}^{EU})$	-0.023	0.017
	(0.040)	(0.037)
$ln(\text{Emi}_{ijt}^{non-EU})$	0.130***	0.125***
	(0.029)	(0.025)
NI_{ijt-1}	0.026	-0.043
	(0.068)	(0.054)
NE_{ijt-1}	0.047	0.069*
	(0.043)	(0.036)
$ln(\text{Dist}_{ij})$	-1.360***	-0.930***
	(0.366)	(0.319)
N	50,210	50,210
province-year effects	yes	yes
region-country effects	yes	yes
country-year effects	yes	yes
Language commonality		
$ln(\text{Immi}_{ijt}^{Spa})$	0.096**	0.046
	(0.046)	(0.038)
$ln(\text{Immi}_{ijt}^{non-Spa})$	0.046**	0.020
	(0.023)	(0.019)
$ln(\text{Emi}_{ijt}^{Spa})$	0.131**	0.117***
	(0.053)	(0.042)
$ln(\text{Emi}_{ijt}^{non-Spa})$	0.064**	0.084***
	(0.028)	(0.025)
NI_{ijt-1}	0.010	-0.054
	(0.068)	(0.054)
NE_{ijt-1}	0.006	0.047
	(0.044)	(0.038)
$ln(\text{Dist}_{ij})$	-1.359***	-0.937***
	(0.362)	(0.317)
N	50,210	50,210
province-year effects	yes	yes
country-year effects	yes	yes
region-country effects	yes	yes

Standard errors clustered at the province-country level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: Robustness checks. OLS and Gamma PML estimates. Geographic proximity

	OLS	GPML
$ln(\text{Immi}_{ijt-1}+1)$	0.035	0.001
	(0.025)	(0.020)
$ln(\text{ImmiOut}_{ijt-1})$	-0.745*	-0.871**
	(0.380)	(0.360)
$ln(\text{Emi}_{ijt-1}+1)$	0.090***	0.095***
	(0.028)	(0.024)
$ln(\text{EmiOut}_{ijt-1})$	-0.080	-0.098
	(0.370)	(0.296)
NI_{ijt-1}	0.008	-0.066
	(0.070)	(0.055)
NE_{ijt-1}	0.038	0.065*
	(0.043)	(0.036)
$ln(\text{Dist}_{ij})$	-1.418***	-0.994***
	(0.367)	(0.315)
N	50,210	50,210
province-year effects	yes	yes
country-year effects	yes	yes
region-country effects	yes	yes

Standard errors clustered at the province-country level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$