CENTRO STUDI LUCA D'AGLIANO

WWW.DAGLIANO.UNIMI.IT

CENTRO STUDI LUCA D'AGLIANO
DEVELOPMENT STUDIES WORKING PAPERS

N. 347

November 2012

On the Pro-Trade Effects of Immigrants

*Massimiliano Bratti**
*Luca de Benedictis***
*Gianluca Santoni****

* University of Milan
** University of Macerata
*** Fondazione Manlio Masi and University of Rome "Tor Vergata"

On the pro-trade effects of immigrants

Massimiliano Bratti[a,*], Luca De Benedictis[b], Gianluca Santoni[c]

[a] *DEMM, Università degli Studi di Milano, via Conservatorio 7, 20122 Milan (MI), Italy*
[b] *DED, Università degli Studi di Macerata, Macerata (MC), Italy*
[c] *Fondazione Manlio Masi and DE Università di Roma "Tor Vergata", Rome (RM), Italy*

Abstract

In this paper we investigate the causal effect of immigration on trade flows. We exploit the very favorable set-up offered by the Italian panel data — the fine geographical disaggregation (provinces, i.e., *Nomenclature of territorial units for statistics* 3 level — NUTS-3,) the very high number of countries of origin of immigrants ('super-diversity',) the high heterogeneity of social and economic characteristics of Italian provinces, and the absence of cultural (e.g. language) or historical (colonial ties) attractors for immigration — to deal with the possible distortions generated by the choice of the areal unit (the so-called Modifiable Areal Unit Problem — MAUP,) comparing estimates at the NUTS-2 and NUTS-3 geographical level; with unobserved heterogeneity, controlling for a wide set of fixed effects; with the endogeneity of immigrants' location choices, using instruments based on immigrants' *enclaves*. We find that immigrants have a significant positive effect on both exports and imports, much larger for the latter. The pro-trade effects of immigrants tend to decline in space, and even turn negative when large ethnic communities are located too far away from a specific province (via a trade-diversion effect). Finally, we give evidence of a substantial heterogeneity in the effects of immigrants: the impact on trade tends to be larger for immigrants coming from low-income countries, for earlier waves of immigrants and for the less advanced provinces of Southern Italy.

JEL Classification F10 · F14 · F22 · R10

Keywords: Immigration, Trade, Gravity model, Super-diversity, MAUP, Transplanted-home bias effect, Business and social networks effects

1. Introduction

At the turn of the century, 4.6% of world population was born in a different country from the one where it currently lived. Similarly, more than 20% of the value of world production was sent to a different country

* Corresponding author
Email addresses: massimiliano.bratti@unimi.it (Massimiliano Bratti), debene@unimc.it (Luca De Benedictis), sntglc01@uniroma2.it (Gianluca Santoni)

December 17, 2012

from the one where it was produced. This paper is about the link between these two 'facts,' controlling for possible common determinants and focusing on the causal effect that immigrants have on the international trade of the host country.

To properly deal with the issue of causality, the analysis takes a single-country perspective examining the Italian case during the 2000s. Looking at this specific case has not only an interest *per se*, but offers many insights on a more general basis. First, Italy shares some common features with many (OECD and non-OECD) immigration countries: in the year 2000, the percentage of the Italian population that was born abroad was 4.1%, and it grew steadily between 2000 and 2009, reaching the total number of 4.2 million foreign-born residents, meaning that 7% of Italian residents were born in a different country (ISTAT, 2011). Second, the large heterogeneity of the countries of origin of immigrants in Italy makes the Italian case relevant for other countries as well. Third, the data collected have some special features in terms of the characteristics of the countries involved in trade and migration flows, and of the time and space of the evolution of the two phenomena in hand. Moreover, the very fine geographical disaggregation of the Italian longitudinal data on trade and migration allows us to adopt a reliable empirical strategy for the identification of the effect of immigration on international trade. For all these features, we think that our analysis offers some widely applicable indications on the interplay between immigration and international trade, giving a general contribution to the ongoing policy debate on the issue.

The Italian immigration case is characterized by what anthropologists call "*super-diversity*," (Vertovec, 2007) a notion intended to emphasize the level and kind of complexity in immigrants' social and economic participation in national everyday life, way above anything the country had previously experienced. The relevant ingredients of super-diverse immigration are "…the increased number of new, small and scattered, multiple-origin, transnationally connected, socio-economically differentiated and legally stratified immigrants who have arrived over the last decade" (Vertovec, 2006). This seems to fit Italy quite well. In Italy, the phenomenon of massive immigration is quite recent. Italy was a land of emigrants at least until the 1960s. It is only in the 1970s that the migration balance started showing a positive sign. To the traditional ethnic groups coming from North Africa, often on a temporary basis, a new diaspora of permanent (essentially housemaid) workers entered Italy from the Philippines, Cape Verde and Sri Lanka. In the 1980s, immigrants coming from Central Africa (Senegal, Nigeria, Côte d'Ivoire, Burkina Faso), South America (Peru, Dominican Republic), the Indian sub-continent (India, Pakistan and Sri Lanka again) and Asia (China) settled permanently in Italy. The more recent wave of immigration took place in the 1990s. It started in 1991 with the dramatic outflow from Albania and became even more numerically relevant with the fall of the Berlin Wall and the entrance of Poland, first, and Romania, afterwards, in the European Union. The 2000s — the period covered by our analysis — is a period of growing immigration characterized by the emergence of new ethnic groups and the downsizing of others. The resulting geographical coverage of the data used is remarkable, allowing to account for 187 countries of origin of immigrants.[1]

In our empirical analysis, we turn to our advantage the minimal participation of Italy to colonialism. As emphasized by Briant *et al.* (2009), in country-level analyses there are very good reasons to suspect that the correlation between trade and immigration might depend on one or more omitted common determinants (such as colonial ties, common language or cultural proximity) or be spoiled by the reverse causality inherent to the fact that immigrants generally move to countries where formal or informal links were already established and where trade with their homeland was already present. In the Italian case, differently from other cases such as the UK (and the London area in particular) or France and the U.S. (and the New York area in particular), the super-diversity of the many ethnicities now living in Italy is largely unrelated to colonial heritage, linguistic or genetic proximity or institutional and cultural similarity. This characteristic of the Italian case is therefore particularly convenient for the identification of the causal effect that immigrants have on trade flows in and out of Italy.[2]

The fine geographical detail of our data is advantageous from an empirical standpoint. In line with recent contributions (Wagner *et al.* (2002); Dunlevy (2006); Bandyopadhyay *et al.* (2008); Briant *et al.* (2009); Peri and Requena-Silvente (2010); Herander and Saavedra (2005), see Section 2 on that) we test the relationship between trade and immigration over fine spatial units.[3] The choice of the appropriate spatial unit of analysis is of primary relevance if there is evidence of a Modifiable Areal Unit Problem (MAUP). As in the case of the 'ecological fallacy' (Jargowsky, 2005) or the 'Simpson's paradox' (Samuels, 1993), the bias associated with MAUP depends on the loss in variation due to data aggregation in arbitrary zonal units. In general, if the grouping process that guided the zonal aggregation is related to an omitted variable which is correlated with the variables of interest, the solution is to control for that omitted variable (see Briant *et al.* (2010) on the issue). In our case, we are able to tackle the MAUP at the origin, using both regional (NUTS-2)[4] – 20 regions

[1]The Italian dataset guarantees the most extensive countries' coverage among those considered in the empirical literature, reducing the risk that the selection of specific countries may bias the estimates of the elasticity of trade to immigration.

[2]Colonial origins and linguistic proximity can both influence trade — and so they do in the traditional analyses of bilateral trade based on the gravity model (see Head *et al.* (2010); Helliwell (1999); Debaere *et al.* (2012) and De Benedictis and Taglioni (2011) and Anderson (2011) for a review of the gravity model in international trade) — and immigration and, therefore, they can confound the relationship between immigrants and trade flows.

[3]To the best of our knowledge, the Italian provinces are the smallest geographical entities used so far to investigate the link between immigration and trade. Briant *et al.* (2009) analyze 96 French *départements* which are almost 30 times tinier than U.S. states (Dunlevy, 2006) and more than 100 times smaller than Canadian provinces (Wagner *et al.*, 2002).

[4]For the unfamiliar reader, NUTS stands for Nomenclature of Territorial Units for Statistics and is a European Union geocode standard for referencing the subdivisions of countries for statistical purposes. There are three zonal levels, NUTS-1,

of an average size of 14000 square km – and provincial (NUTS-3) Italian data – namely 107 Italian provinces of an average size of 2800 square km — and comparing the results of the analyses.[5] The evidence goes in favor of the use of provincial data. This areal unit is also relatively more appropriate from a theoretical viewpoint. Indeed, the most popular explanations for the pro-trade effects of immigrants (see section 2) are based on interactions and knowledge flows between natives and immigrants. These interactions are likely to depend on the distance between individuals, and are accordingly more precisely captured if the geographical units of analysis are small areas such as NUTS-3, rather than countries, states or NUTS-2 regions.

Our analysis makes several improvements over the existing literature. First, the risk of a spurious correlation between trade and immigration is minimized owing to the very fine geographical scale of our analysis. This also allows us to investigate geographical spillovers of immigrants on trade, and to analyze the implications of the MAUP for the trade-immigration link. Second, the extensive country coverage of our dataset ensures that any sample selection bias stemming from the specific choice of the foreign countries entering the analysis has been avoided. Third, to further rule out the possibility of an endogeneity bias that could inflate our coefficients of interest, we control for omitted common determinants including time-varying country-specific, time-varying region-specific and trading-pair fixed effects in the regressions, and, especially, we make use of an Instrumental Variables (IVs, hereafter) approach *à la* Altonji and Card (1991), where the geographical distribution of immigrants' residence permits in 1995 (the earlier year for which comparable geographical data are available) and immigrants' flows at the nationwide level serve to compute an instrument (the imputed stock of immigrants).[6] Fourth, we bring to the data the two main explanations highlighted in the literature: the *business and social network effect à la* Rauch (2001) (i.e. immigrants foster both bilateral imports and exports because of their superior knowledge of, or preferential access to, market opportunities in their home-country) and the *transplanted home-bias effect* (Gould, 1994; White, 2007) (i.e. immigrants promote imports of their home-country consumption-goods to satisfy their different consumption tastes). In order to do that, we use both export and import flows, and we depart from the most recent literature which,

NUTS-2 and NUTS-3, which for Italy correspond to the country, region (*regione*) and province (*provincia*) levels, which also correspond to the three main administrative units of the country.

[5]To be more precise, the mean area of Italian provinces is 2,816 square km with a coefficient of variation of 0.17, almost 57 times tinier than American states (162,176 square km, when Alaska and Washington DC are included), and more than 200 times smaller than Canadian provinces (606,293 square km when Nunavut, North-West and Yukon territories are excluded). These administrative units are also much smaller and more regular in size with respect to French metropolitan *départements* and Spanish provinces. The mean area of French *départements* is 5,666 square km with a coefficient of variation of 0.33 (when Corsica and overseas French regions are excluded), whereas the related figures for Spanish provinces are 10,118 square km with a standard deviation of 0.47 (excluding Ceuta and Melilla).

[6]To date, very few studies have attempted to address endogeneity when investigating the immigration-trade link. To the best of our knowledge only Briant *et al.* (2009) and Peri and Requena-Silvente (2010) have done so. Moreover, this is the first paper to investigate geographic spillovers of immigrants on trade and possible heterogeneous effects taking into account endogeneity.

examining exports but not imports, focused only on the first causal pathway (see section 2). Finally, we give evidence of the heterogeneous effects of immigrants on trade according to the level of per capita income of their country of origin, the timing of arrival (distinguishing between ethnic groups participating to the first or the second wave of immigration) and their geographical location in Italy (distinguishing between Northern and Southern Italy).

The remainder of the paper is organized as follows. Section 2 discusses the literature on the pro-trade effects of immigrants and highlights the traditional mechanisms behind this positive effect. Section 3 presents the data used in the analysis (which is also fully described in Appendix 1) and discusses both Italian 'super-diversity' and geographical heterogeneity. Section 4 includes the benchmark empirical results; in subsection 4.1 we run a simple OLS regression and we discuss the possible shortcomings of this approach; in subsection 4.2 we discuss the role played by geographical spillovers in the effect of immigrants on bilateral trade; in subsection 4.3 we describe the strategy used to tackle the endogeneity issue and report the causal effect resulting from two-stage least squares (2SLS) estimates. In section 4.4 we investigate treatment effects' heterogeneity. Section 5 summarizes the conclusions and the Appendices include the full description of the data used and of the foreign countries considered in the analysis, a discussion of the role of fixed-effects in saturating the empirical model, and a description of empirical attempts related to the inclusion of zero-trade flows in the analysis, and the different role played by large Italian cities such as Rome and Milan.

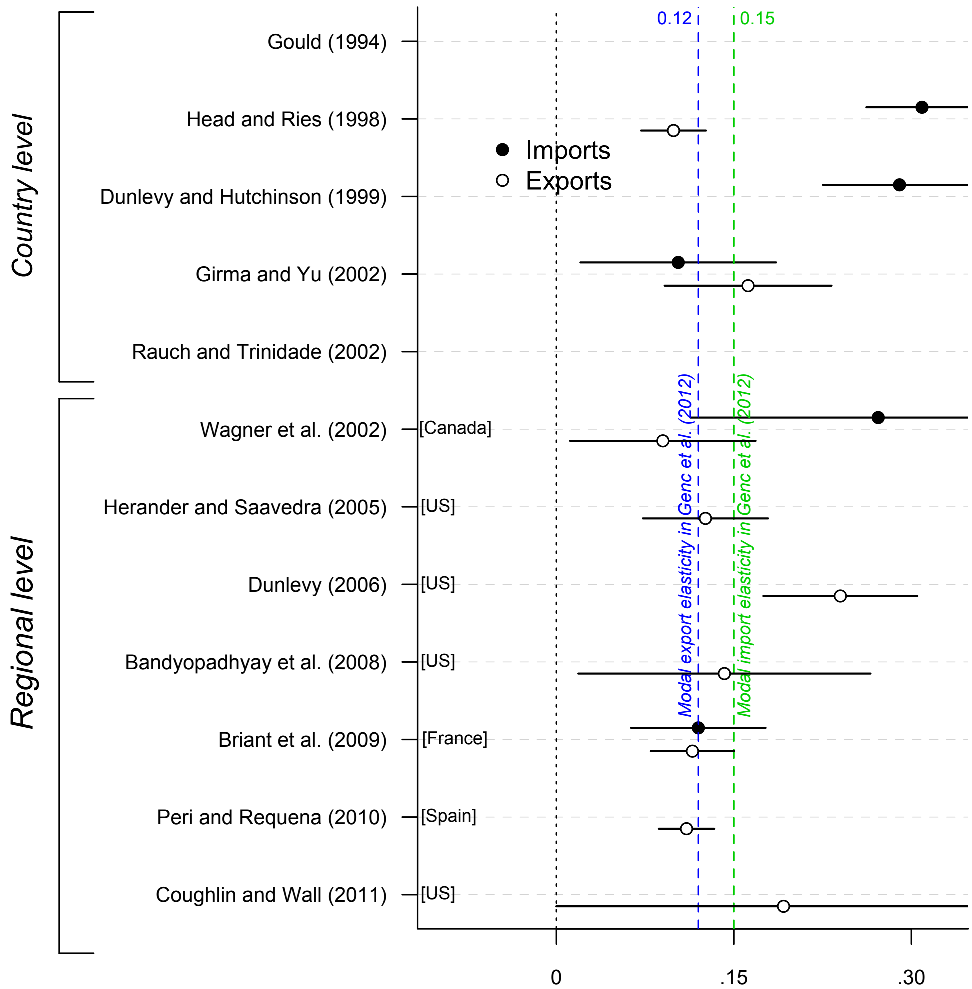
2. The pro-trade effects of immigrants

The international trade literature based on the estimate of a gravity equation[7] — where trade flows between a regional entity i and its international counterpart j are positively associated with economic attractors, such as the GDPs of i and j, and negatively associated with obstacles to international trade, such as distance — has generally found a strong association between immigration and trade. The presence in i of immigrants from j can be considered as an attracting force, fostering international trade between i and j.

Different studies (Head and Ries, 1998; Dunlevy and Hutchinson, 1999; Rauch and Trinidate, 2002; Girma and Yu, 2002; Coughlin and Wall, 2011), for different samples, periods and estimation techniques have generally reported a strong positive association between immigrants and trade. Some recent papers have also made attempts to qualify such association as causal using IVs methods (Briant *et al.*, 2009; Peri and Requena-Silvente, 2010).

[7] See De Benedictis and Taglioni (2011) for an empirically oriented review of the literature on the gravity model in international trade.

Figure 1: Summary of the literature. Estimated elasticity of trade to immigrants: Imports (black) Exports (white)



Note: The figure plots data obtained from several contributions to the literature on the effect of immigration on trade. Black dots indicate the elasticity of imports to immigrants, white dots that of exports, segments indicate 95% confidence intervals. The two vertical lines correspond to the meta-modal elasticity estimated in Genc *et al.* (2012), which is 0.12 for exports and 0.15 for imports. The horizontal axis is trimmed for visual purposes. The seminal contribution by Gould (1994) did not use a gravity equation framework and the estimated elasticities are not comparable and omitted from the plot. The complete list of papers is included in the References.

In figure 1 we summarize the results of a sample of relevant contributions to the literature in terms of the estimated elasticity of trade to the stock of immigrants. Black dots depict imports, white dots exports, and horizontal segments indicate 95% confidence intervals. The two vertical lines correspond to the meta-modal elasticity estimated in Genc *et al.* (2012), which is 0.12 for exports and 0.15 for imports. As it is evident, the estimates show a high degree of variability. Between the seminal contribution by Gould (1994), that basically started this new stream of research, and 2002, the literature was dominated by cross-country studies. Wagner *et al.* (2002) set the standard in the subsequent set of contributions, highlighting the role of two fundamental ingredients: (a) country-fixed effects, to control for omitted variable bias; (b) the use of regional data, to exploit cross-sectional variation on trade and immigration at the sub-national level (for Canada, in their original analysis) and to deal with the endogeneity bias discussed in section 1. Since Wagner *et al.* (2002) the variability in the estimates reduces substantially.

One first important evidence of the literature is that the elasticity of imports to immigrants is higher than that of exports, and that both are positive and generally significantly different from zero. Why?

The mechanisms at the basis of the common explanations of the pro-trade effects of immigrants are twofold. The main explanation is rooted in the idea that information costs play a major role in the fixed cost that firms pay to enter foreign markets. In the seminal contributions of Rauch (1999, 2001), ethnic networks related to migration flows are likely to reduce some of these information costs. Cross-border networks of people sharing the same country of origin can substitute or integrate organized markets in matching international demand and supply. Several studies have explored the role of ethnic networks in international trade since Rauch (2001).[8]

A further point associated with this first explanation is related to the characteristics of immigrants and how these characteristics can reduce the fixed cost of exporting. Language, specific knowledge of homeland institutions and norms, familiarity with homeland (excess) demand, can bridge the home-country and the host-country, if these assets are positively valued and acquired by firms producing in the country were immigrants settled (Wagner *et al.*, 2002; Peri and Requena-Silvente, 2010). Moreover, "immigrant networks may provide contract enforcement through sanctions and exclusions, which substitutes for weak institutional rules and reduces trade costs" (Briant *et al.*, 2009). Following Rauch (2001), this explanation has been named the *business and social network effect* of immigrants on trade.

The second, less explored explanation, is that immigrants are characterized by different habits in con-

[8]See, among others, Rauch and Trinidate (2002), Epstein and Gang (2004), Felbermayr *et al.* (2010) and Coughlin and Wall (2011).

sumption with respect to natives, and they may slowly modify their original home-biased demand after settling in the host-country (see Gould (1994) for an innovative discussion of the issue). Since homeland goods are more costly in the host-country, immigrants have an incentive to buy those goods from the home-country itself. Proper empirical evidence on what has been called the *transplanted home-bias effect* of immigrants on trade was, until recently, basically non existent (White, 2007; White and Tedesse, 2007). The significance and magnitude of the effect was generally inferred from the difference between the estimated immigrant-elasticity of imports (to which both effects were contributing) and exports (not affected by the transplanted home-bias effect). Since, as it is evident from figure 1, the immigrants elasticity of imports tends to be higher than that of exports, this was interpreted by deductive reasoning as supporting the idea that there should be something forcing the two elasticities to be different, and this 'something' was attributed to a persistent difference in tastes between immigrants and natives.

Recently, some more clear evidence of the relevance of the transplanted home-bias effect has been provided by Bronnenberg *et al.* (2012), Atkin (2010) and Mazzolari and Neumark (2012). Bronnenberg *et al.* (2012) looking at the consumption behavior of U.S. consumers migrating across state borders, find that in choosing between the two top brands in a category of a very specific product, past experiences are an important driver of current consumption. Consumers migrating from a certain U.S. state tend to partially adapt to local habits to a certain extent, but in spite of the difference in price and in brand availability, they still tend to persist in consuming according to the prevalent choices in the U.S. state they migrated from. Even after 50 years migrants still consume 'differently' from locals.

The same evidence is found for India in Atkin (2010), where it is shown that inter-state migrants carry their food tastes with them, consuming food bundles less similar to those consumed in their destination state and more similar to those consumed in their state of origin. Migrants originating from rice-producing states keep consuming rice instead of wheat, notwithstanding rice being relatively more expensive than wheat on the local market. This habit persistence dissipates with time, disappearing slowly and lasting four generations after migration.

The willingness to pay high prices for goods similar to those one consumed in the home-country is also found in Mazzolari and Neumark (2012), where immigration is associated with increased ethnic diversity of restaurants.

While more recent contributions have disregarded the effect of immigrants on imports (see the regional-level estimates in table 1), in the following analysis we look at both the export and import elasticities to immigrants, so as to give quantitative content to both the *business and social network effect* and the *transplanted home-bias effect* of immigrants on trade.

3. Data and descriptive statistics

The data used in the analysis come mainly from two publicly available data sources collected by the Italian National Statistical Institute (ISTAT). Trade flow data refer to the value of imports and exports of 107 Italian provinces (NUTS-3) with 210 countries, over the period 2002-2009.[9] The trade data are originally measured in Euros, and report export and import flows between the Italian province of shipment, i.e. the province where the custom transaction was registered, and the foreign country of destination (for exports) and of origin (for imports).[10] Information on the number of foreign born residents by Italian province or region and foreign country of origin is obtained from ISTAT as well, and covers the same period. Our explanatory variable of interest is the stock of legal immigrants by country of origin (home-country) and province (or region) of destination in Italy, defining immigrants as residents born abroad with a foreign nationality.[11]

Of the 187 ethnic groups included in the dataset, table 1 shows the top 20 countries of origin of immigrants in 2009. The top five countries by the number of immigrants are Romania, Albania, Morocco, China and Ukraine, accounting for about 50 percent of the total foreign-born population. Comparing the rank of these top 20 countries of origin, and especially the average growth rate over the period, gives an idea of the change in the composition of immigrants by the country of origin. In 2009, the majority of the foreign-born population came from Eastern Europe (Romania, Ukraine, Rep. of Moldova, Poland), the area which experienced also the highest growth rate of immigration over the period. The change in the ranking between 2002 and 2009 is reported in figure 2 which visualizes some of the ethnic 'big movers' (only 20 out of the total 187 country of origin are plotted). Moldova and Ukraine, for instance, gained 32 and 23 positions, respectively, while Senegal lost 9 positions.

An interesting feature of the immigration pattern in Italy is the uneven distribution of immigrants across

[9] More precisely, we consider 103 provinces until 2006 and 107 afterwards. The number of Italian provinces changed in recent times, as reported by ISTAT. In the mid 1990s the number of Italian provinces was 103. In 2001 the Sardinia autonomous region established 4 new provinces, that became operative during 2005. In 2004 the Italian Parliament established 3 new provinces that became operative in 2009. The total actual number of provinces is 110. Since our dataset does not include observations for the years after 2009, we do not consider these latter changes in the number of Italian provinces.

[10] The information of Extra-EU transactions are based on the "Documento Amministrativo Unico" (DAU), for the intra-EU exchanges the custom system has been replaced, since 1993, by the Intrastat standard. The original values of trade flows, in Euros, have been converted in U.S. dollars using the nominal exchange rate from the World Development Indicators (WDIs on-line database) in order to make them consistent with GDP data used in the gravity equations. The conversion is not influencing the results, since in the multivariate regression in Section 4 we use country time-varying fixed effects.

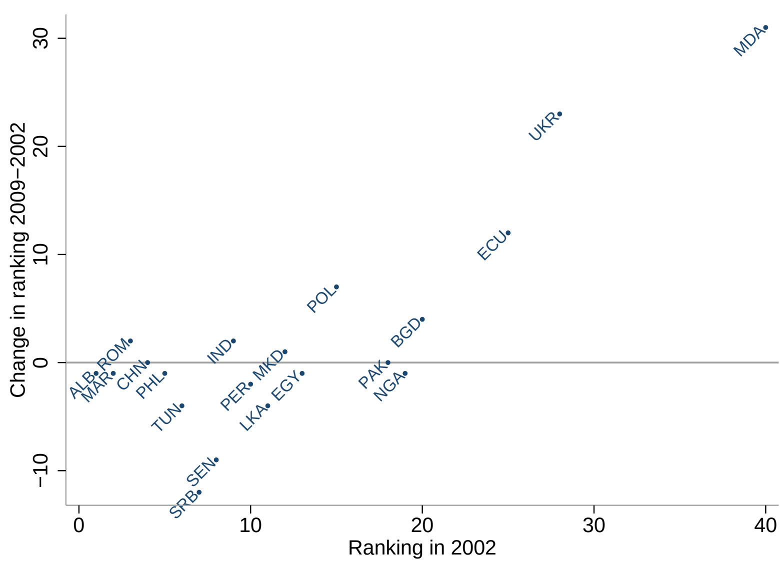
[11] Like all previous papers on the topic, we only consider legal immigrants. Direct information on the stocks of immigrants with illegal status is not available.

Table 1: Immigrants by country of origin (first 20)

Ranking in 2009	Country of origin	ISO3 Country code	Number of immigrants in 2009	% of total immigrants in 2009	Annual growth rate, 2002/2009 (%)	Ranking in 2002
(1)	Romania	ROM	887,763	20.96	40.45	(3)
(2)	Albania	ALB	466,684	11.02	11.76	(1)
(3)	Morocco	MAR	431,529	10.19	10.51	(2)
(4)	China	CHN	188,352	4.45	15.51	(4)
(5)	Ukraine	UKR	174,129	4.11	68.99	(28)
(6)	Philippines	PHL	123,584	2.92	9.67	(5)
(7)	India	IND	105,863	2.50	16.99	(9)
(8)	Poland	POL	105,608	2.49	20.04	(15)
(9)	Moldova Rep.	MDA	105,600	2.49	60.20	(40)
(10)	Tunisia	TUN	103,678	2.45	8.33	(6)
(11)	Macedonia	MKD	92,847	2.19	16.25	(12)
(12)	Peru	PER	87,747	2.07	14.60	(10)
(13)	Ecuador	EQU	85,940	2.03	32.67	(25)
(14)	Egypt	EGY	82,064	1.94	13.82	(13)
(15)	Sri Lanka	LKA	75,343	1.78	11.99	(11)
(16)	Bangladesh	BGD	73,965	1.75	20.27	(20)
(17)	Senegal	SEN	72,618	1.71	10.24	(8)
(18)	Pakistan	PAK	64,859	1.53	16.72	(18)
(19)	Serbia	SRB	57,877	1.37	1.19	(7)
(20)	Nigeria	NGA	48,674	1.15	12.97	(19)
	Top 20 countries		3,434,724	81.1	20.66	
	TOTAL		4,223,154	100	14.9	

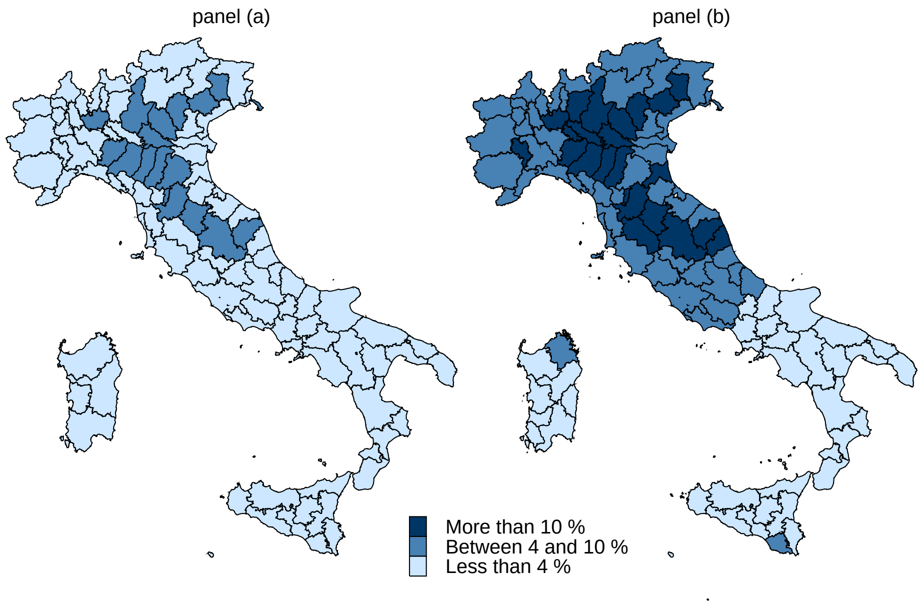
Source: ISTAT

Figure 2: Ranking of immigrants by country of origin (first 20)



Note: The figure plots the top 20 countries of origin of immigrants in 2009 vs the rate of change between 2002 and 2009. Countries are identified by ISO3 country codes, published by the International Organization for Standardization (ISO) to represent countries, dependent territories, and special areas of geographical interest with a three-letter code.

Figure 3: Percentage of foreign-born population across Italian provinces. Year 2002 (panel a) and year 2009 (panel b)



Italian provinces. Figure 3 shows a map of Italy where provinces are colored according to the share of foreign-born population in the total population, with 'darker' provinces hosting a higher share of immigrants. While in 2002 none of the 103 provinces registered a share higher than 10 percent, in 2009 twenty three provinces exceeded the level of 10 percent of foreign born residents, mainly in the Center and the North of the country. The map of Italy also reveals some spatial clustering of immigrants: immigrant-abundant provinces are more likely to be close to each other. We address the potential consequences of this issue for our analysis in Section 4.2.

Although the distribution of foreign residents across provinces reveals a relative concentration in Northern Italy, the number of provinces with zero immigrants from a particular country of origin is rather small. This is an instance of the Italian 'super-diversity'. Table 2 reports the mean number of nationalities registered in each province at the beginning and at the end of the period under study: the value is around 111 in 2002, and about 125 in 2009. Provinces with less coverage of nationalities are in the South of Italy, as is evident from figure 3. If we look at the distribution of immigrants from the perspective of each ethnic community, immigrants from the same country of origin located, on average, in 82 out of the 103 Italian provinces in 2002, and in 90 out of 107 Italian provinces in 2009. The distribution is far from being uniform: some immigrant communities are concentrated in a limited number of provinces (the minimum number of provinces for an

Table 2: Migrants' location by province and country of origin

	Mean	Std. Dev.	Min	25p	Median	75p	Max
Year: 2002							
Foreign nationalities per province	111.92	23.39	49	97	112	128	175
Provinces per Foreign nationality	82.07	24.37	1	68	95	102	103
Year: 2009							
Foreign nationalities per province	124.91	20.46	58	113	126	139	179
Provinces per Foreign nationality	90.23	22.35	1	83	100	106	107

Note: 25p stands for 25th percentile, while 75p stands for 75th percentile. The total number of Italian provinces is 103 (107 from 2006) while the total number of foreign nationalities is 187.

immigrant community is just 1), others are spread all over Italy (the observed maximum always hits the theoretical one, i.e. the number of provinces). Focusing on the twenty most numerically relevant nationalities, we register huge differences in their geographical settlement. The most concentrated groups are from Egypt, Ecuador and the Philippines. The distribution exhibits in 2009 a coefficient of variation[12] from 40% to 80% higher than the median value. On average, around 70% of these communities is located in the first five provinces of residence. The degree of concentration is relatively high compared to Albania, Morocco and Tunisia, the most evenly distributed nationalities. In this case the first five locations account for less than 25% of total residents. The most widely represented country of origin (Romania) records over 139,000 residents just in the province of Rome. The provinces of Rome and Milan play an attractive role that make them different form other provinces. We control for this peculiarity in the multivariate analysis reported in Section 4. The incidence of zeros in the full dataset is relatively high. On average we observe zero flows for 48% of imports and 29% of exports from or to province i coming from or directed toward the foreign country j, while for immigrants' stocks, the percentage of zeros is 39%.We also discuss the zero-trade-flows issue in Section 4.

A full account of the characteristics of the trade and migration data is included in Appendix 1. The sample summary statistics are tabulated in the same Appendix (see Table 9).

[12]The coefficient of variation refers to the distribution of the province's share of the total number of foreign residents by nationality.

4. Empirical results

In this section we put the various pieces of the analysis together. Having the results of previous contributions in mind and relying on the quality of the Italian provincial data on trade and immigration, we look for evidence of pro-trade effects of immigrants. The main steps of our empirical strategy are dictated by the necessity to control for possible reverse causality (from trade to immigration) and the omission of common variables affecting both trade and immigration in the same direction. We operate in sequence. In subsection 4.1 we assume that, after controlling for a wide range of fixed effects, immigrant stocks are exogenous with respect to trade, and use OLS. In subsection 4.2 we allow for geographic spillovers, still retaining the assumption of exogeneity. In subsection 4.3 we deal with the remaining potential endogeneity of immigration, instrumenting the stock of immigrants with an imputed variable related to past ethnic diasporas in Italian provinces and report 2SLS estimates. Finally, subsection 4.4 sheds light on the possible heterogeneous effects of immigrants.

4.1. Ordinary least squares and MAUP

Our starting point is the following Anderson and van Wincoop (2003) theoretically founded specification of the gravity equation:[13]

$$\ln(1 + X_{ijt}) = \boldsymbol{\delta}_{rj} + \boldsymbol{\theta}_{jt} + \boldsymbol{\phi}_{rt} + \alpha \ln(Y_{it-1}\ Y_{jt-1}) + \beta\ \ln(1 + IMM_{ijt-1}) + \\ + \gamma \ln(distance_{ij}) + \rho\ \ contiguity_{ij} + \epsilon_{ijt} \tag{1}$$

where i is the subscript for Italian provinces (NUTS-3), r is the subscript for the region where province i is located (NUTS-2), j indicates the foreign country (i.e. the country of origin of immigrants), and t stands for time. $\boldsymbol{\delta}_{rj}$ are $region \times country$ (trading-pair) fixed effects, $\boldsymbol{\theta}_{jt}$ are $country \times year$ fixed effects, and $\boldsymbol{\phi}_{rt}$ $region \times year$ fixed effects. X_{ijt} is trade (exports or imports) between province i and country j at time t. Y_{it-1}, Y_{jt-1} are province and foreign countries GDPs at time $t-1$, IMM_{ijt-1} is the stock of immigrants from country j located in province i, acting as a trade-enhancing force in contraposition with $distance_{ij}$, which is the great-circle distance between province i and country j. $Contiguity_{ij}$ is a dummy for contiguity between the Italian province i and the foreign country j, included to take into account possible nonlinearities in distance. Trade flows cover the 2003-2009 period and immigration stocks the 2002-2008 period. Covariates

[13] The Anderson and van Wincoop (2003) specification of the gravity equation can be derived from micro-foundations, and results from an expenditure function that takes into account the fundamental role of general equilibrium effects in trade: aka, the multilateral resistance index. See De Benedictis and Taglioni (2011) and Anderson (2011) on the theoretical foundation of the gravity equation.

are predetermined and time-lagged, and ϵ_{ijt} is an error term clustered at the country by province level.

Being consistent with the previous literature, we use a log-log version of the gravity model, and to retain observations with zeros in either trade flows or immigrant stocks, we follow Dunlevy (2006); Peri and Requena-Silvente (2010); Coughlin and Wall (2011); Artal-Tur *et al.* (2012) adding a constant equal to one to both bilateral trade flows and the stock of immigrants. Since trade is measured in dollar units (rather than thousands, millions or billions of dollars) this is likely to introduce only a small measurement error in the observations with zero trade (see Liu, 2009). Previous contributions treated zero trade observations in different ways. Bandyopadhyay *et al.* (2008), for instance, restrict the analysis to the observations with positive trade. However, in general, by considering only observations with positive trade one is likely to neglect the effect of immigrants on the extensive margin of trade (i.e. the existence of a trade link), which is likely to be a relevant dimension of the trade-creation effect of immigration.

It is worth noting that we are not including in the gravity equation *province* × *country* and *province* × *year* fixed effects because they would absorb almost all variation in both trade and immigration, leaving virtually nothing else to be explained (this is a crucial point in our analysis, and we give full account of it in Appendix 2). By contrast, we use larger geographical units in Italy (NUTS-2 administrative areas) to define *region* × *country* fixed effects, $\boldsymbol{\delta}_{rj}$, and *region* × *year* fixed effects, $\boldsymbol{\phi}_{rt}$. This enables us to control for unobserved heterogeneity at the *regional* level and still be able to exploit within-region variation across provinces (i.e. cross-sectional variation between provinces located in the same region) in both trade and immigration to identify the effect of interest. A similar approach has been used by Wagner *et al.* (2002), which exploits cross-sectional variation between Canadian provinces to investigate the effect of immigration on trade and controls for foreign countries' fixed effects.As the authors state that '. . . this approach enables us to capture most of the advantages of fixed effects, since the special relationships that affect both trade and immigration likely occur politically at the national level. Yet, by using province-level data, we can still make use of cross-sectional variation and need not rely solely on temporal variation.' (p. 515). We bring their original idea a step further in the spatial dimension. Moreover, another reason to opt for a specification including region-country fixed effects is related to the IVs-based identification strategy we use in Section 4.3. As we report in that Section, when using a log-log specification and instruments based on a shift-and-share analysis, in which the lagged distribution of immigrants across provinces is used to build the instrument, if then we would include *province* × *country* fixed effects instead of *region* × *country* fixed effects, the 2SLS estimates would be identified by non-linearity only, and not by an exclusion restriction.

The specification of equation (1) is to the best of our knowledge one of the most comprehensive used in the literature in terms of the set of fixed effects included. Early papers did not include importer or exporter dummies (see, for instance, Head and Ries, 1998; Girma and Yu, 2002). Several subsequent contributions to the literature included importer and/or exporter fixed effects (Dunlevy, 2006; Wagner *et al.*, 2002; Santos Silva and Tenreyro, 2006; Briant *et al.*, 2009). Some recent papers using sub-national level data include *trading-pair* and *year* fixed effects (Bandyopadhyay *et al.*, 2008) or *trading-pair* and *country* $\times$ *year* fixed effects (Peri and Requena-Silvente, 2010; Coughlin and Wall, 2011) but *region* $\times$ *year* fixed effects are generally omitted. Yet, the inclusion of the complete set of fixed effects attenuates the potential endogeneity of migration flows. Indeed, in our specific case, *trading-pair* fixed effects are likely to account for factors that may spur trade between an Italian region and a specific country such as cultural proximity or historical ties (e.g., past Italian emigration from a given province towards a certain foreign country,) *country* $\times$ *year* dummies are likely to capture time-variant foreign countries' economic, social and political events (e.g., the entry in the EU, wars or economic crises) which are likely to affect both trade and migration flows towards many Italian provinces, and *region* $\times$ *year* dummies capture features such as the state of the local economy which may affect both trade and immigration flows from several foreign countries. Hence, the focus on sub-national variation within the same country coupled with the inclusion of a wide range of fixed effects is likely to attenuate the potential endogeneity problems of immigration with respect to trade. For this reason, we use in this section the OLS estimator, clustering standard errors at the *province* $\times$ *country* level.

Before commenting on the results obtained using the *benchmark* specification described in equation (1), we report the results from specifications that include fewer fixed effects, to stress the importance of controlling for unobserved heterogeneity. In these specifications we also included (for comparison with previous analyses at the European level) a dummy for EU or EFTA participating countries, whenever it is not absorbed by the fixed effects included (or not dropped due to collinearity). Panel (a) of Table 3 reports the results for exports. In column (1) we report the results of a specification only including time dummies (*year* fixed effects): the estimated elasticity of exports with respect to immigration is very high, at 0.506, meaning that a 1% increase in the stock of immigrants settled in a specific Italian province raises exports from that province to the country of origin of immigrants by about 0.5%.[14] Column (2) reports a specification including *trading-pair* and *year* fixed effects, from which the elasticity drops to 0.084. In column (3), we control for *trading-pair* and *region* $\times$ *year* fixed effects, and the elasticity does not vary (0.083). In column (4), we report a specification including *trading-pair* and *country* $\times$ *year* fixed effects but excluding *region* $\times$ *year*

[14] All provinces of Sardinia are omitted from the analysis in 2006. This depends on the fact that, as we said above, four new provinces were created in Sardinia and we do not have lagged values for the independent variables for 2006.

Table 3: Gravity equations for exports and imports (OLS) — province level

	(1)	(2)	(3)	(4)	(5)	(6)
(a) *Exports*						
$\ln(Y_{it-1}Y_{jt-1})$	1.638***	2.108***	2.101***	2.174***	2.167***	2.296***
	(0.011)	(0.027)	(0.027)	(0.029)	(0.029)	(0.032)
$\ln(1+IMM_{ijt-1})$	0.506***	0.084***	0.083***	0.059***	0.058***	0.121***
	(0.013)	(0.019)	(0.019)	(0.020)	(0.020)	(0.021)
EU_{jt}, $EFTA_{jt}$	0.108	-0.282***	-0.280***			
	(0.072)	(0.053)	(0.052)			
$\ln(distance_{ij})$	-1.270***	-1.832***	-1.829***	-1.876***	-1.873***	-1.588***
	(0.032)	(0.386)	(0.384)	(0.389)	(0.388)	(0.392)
$contiguity_{ij}$	-1.105***	0.116	0.117	0.141	0.142	0.048
	(0.411)	(0.306)	(0.305)	(0.318)	(0.316)	(0.356)
Fixed effects:						
year	Yes	Yes				
region × year			Yes		Yes	Yes
country × year				Yes	Yes	Yes
trading-pair		Yes	Yes	Yes	Yes	Yes
Sample:						
including Rome and Milan	Yes	Yes	Yes	Yes	Yes	
N. observations	135,586	135,586	135,586	135,586	135,586	132,982
R-squared	0.61	0.79	0.79	0.80	0.80	0.80
N. clusters	20,009	20,009	20,009	20,009	20,009	19,635
(b) *Imports*						
$\ln(Y_{it-1}Y_{jt-1})$	1.649***	2.017***	2.012***	2.105***	2.100***	2.127***
	(0.013)	(0.034)	(0.034)	(0.036)	(0.036)	(0.038)
$\ln(1+IMM_{ijt-1})$	0.796***	0.363***	0.362***	0.344***	0.344***	0.347***
	(0.016)	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)
EU_{jt}, $EFTA_{jt}$	1.929***	-0.367***	-0.366***			
	(0.088)	(0.086)	(0.086)			
$\ln(distance_{ij})$	-0.581***	-2.837***	-2.834***	-2.899***	-2.897***	-2.696***
	(0.039)	(0.592)	(0.592)	(0.595)	(0.595)	(0.594)
$contiguity_{ij}$	0.236	-0.301	-0.301	-0.284	-0.283	-0.315
	(0.555)	(0.339)	(0.338)	(0.357)	(0.356)	(0.360)
Fixed effects:						
year	Yes	Yes				
region × year			Yes		Yes	Yes
country × year				Yes	Yes	Yes
trading-pair		Yes	Yes	Yes	Yes	Yes
Sample:						
including Rome and Milan	Yes	Yes	Yes	Yes	Yes	
N. observations	135,586	135,586	135,586	135,586	135,586	132,982
R-squared	0.63	0.76	0.76	0.77	0.77	0.76
N. clusters	20,009	20,009	20,009	20,009	20,009	19,635

*, **, ***, significant at the 10%, 5% and 1% statistical level

Note. The dependent variable is $\ln(1+export_{ijt})$ for panel (a) and $\ln(1+import_{ijt})$ for panel (b), i.e. export (import) flows of province i to (from) country j at time t. *Trading-pair* fixed effects are defined at the *region × country* level. Export and import flows cover the period 2003-2009. Standard errors are clustered at the province by (importer or exporter) country level. In column (6) we exclude Rome and Milan from the sample.

fixed effects.[15] In this case the drop in the elasticity is smaller than in column (3). Moving to our *benchmark* specification does not generate a major change in the estimated elasticity, which becomes 0.058 in column (5). Hence, it seems that *trading-pair* and *country* $\times$ *year* fixed effects captures most of the unobserved heterogeneity in our data, causing a large drop in the elasticity of exports to immigrants with respect to the initial specification. In any case we consider the model including *trading-pair*, *country* $\times$ *year* and *province* $\times$ *year* fixed effects (in column (5)) as our preferred specification in terms of completeness, and limitation of endogeneity problems in the IVs strategy that we subsequently use and report in Section 4.3. In column (6), we estimate the *benchmark* specification excluding Rome and Milan from the sample, in order to evaluate the potentially different role of large cities. The elasticity rises to 0.121, showing that the conditional correlation between the export flows from the Italian province i to the foreign country j and the presence of an ethnic community from country j hosted in the province i is positive and much larger when we do not consider the two major Italian urban hubs, in political and administrative (Rome), and economic (Milan) terms. We will return to this specific and important point later on. Finally, in all specifications included in table 3 the coefficient on immigration is statistically significant at least at the 1% level.

When we consider imports — in panel (b) of Table 3 — we find a similar fall in the estimated elasticities by progressively adopting richer specifications in terms of fixed effects. The elasticity of imports with respect to immigrants is 0.796 in column (1). It falls to 0.363 when adding *trading-pair* fixed effects, in column (2). Also in this case, comparison of column (2) with the following columns shows that, as for exports, most unobserved heterogeneity is captured by the *trading-pair* fixed effects, and the estimated elasticity turns out to be only marginally affected by the inclusion of the other fixed effects. In the *benchmark* specification in column (5) the elasticity of imports with respect to immigration is 0.344, and rises only marginally to 0.347 in column (6) when Rome and Milan are excluded from the sample. Also for imports, as for exports, the coefficient on immigration is always statistically significant at least at the 1% level.

Thus, as predicted by economic theory and confirmed in most of the previous literature (reviewed in section 2), our estimate of the elasticity of imports with respect to immigrant stocks is much larger than that of exports. This stems from the fact that while both the *transplanted-home bias effect* and the *business and social networks effects* are at work for imports, only the second causal pathway affects exports.

The first step done, we deal now with possible concerns regarding OLS. To begin with, the log-log version of the gravity model has been recently subject to some criticism by Santos Silva and Tenreyro (2006). The

[15] This is equivalent to the 'basic' specification in Peri and Requena-Silvente (2010) (see column 1 of Table 4 in their article) or the preferred specification in Bandyopadhyay *et al.* (2008).

debate on the most appropriate nonlinear estimator to be applied when zeros are a relevant proportion of trade flows is still very open (see De Benedictis and Taglioni (2011) on this specific point of the gravity literature). In the present case, the use of the Pseudo Poisson Maximum Likelihood estimator proposed by Santos Silva and Tenreyro (2006) clashes with the inclusion of many fixed effects. The very high number of fixed effects prevent us from using any other nonlinear estimator or from applying the Heckit estimator as in Helpman *et al.* (2008) or the threshold Tobit model of Eaton and Tamura (1994) to account for zero-trade observations, since both require estimating a Probit model which suffers from an incidental parameters problem.[16] In conclusion, controlling for unobserved heterogeneity through $region \times country$ fixed effects, makes the log-log specification the sole feasible option among the many possible different alternatives. For the sake of brevity, a full account of our attempts to deal with zero-trade observations is reported in Appendix 3. More in general, the present case is indeed paradigmatic in terms of the trade-off between accounting for unobserved heterogeneity in trade data (through fixed effects) and using nonlinear models to estimate the gravity equation.

We checked nonetheless the sensitivity of our results to alternative transformations of the dependent variables (exports and imports) which allow us to retain the zero-trade observations in the estimation. We tried (i) adding to exports and imports 0.1 (i.e. 10 cents of a dollar) instead of one dollar before taking natural logarithms; and (ii) using an inverse hyperbolic sine (IHS) transformation (Burbidge *et al.*, 1988).[17] Using the benchmark model of column (5), in the first case we obtain point estimates of 0.027 and 0.369 for exports and imports respectively, while in the second case (IHS) the estimates are 0.049 and 0.352. In both cases, the estimates were not statistically different from those reported in column (5) of Table 3.[18]

A second issue, which we mentioned above, involves the peculiarity of Rome and Milan, the two largest cities in Italy. Their specificity is associated with the size, and density of economic activity, the peculiar characteristic of a capital city and the associated historical presence of ethnic *enclaves*. Indeed, thanks to the presence of agglomeration economies, firms located in these two cities are likely to be the most productive and efficient, and export their goods irrespective of the presence of immigrants. In other words, they may act as outliers in the trade-immigration relationship. This is indeed confirmed in Appendix 4, in which we

[16]In general studies using a poisson specification or other non-linear models adopt a much less richer set of fixed effects. Just to take two examples, Helpman *et al.* (2008) include separate importer, exporter and year fixed effects, while Eaton and Tamura (1994) include separate region, sub-continent and year fixed effects.

[17]The inverse hyperbolic sine (IHS) transformation consists of replacing X_{ijt} with $ln(X_{ijt} + (X_{ijt}^2 + 1)^{1/2})$. In this case, as in the traditional logarithmic transformation, if the values of X_{ijt} are not too small, the coefficients of the covariates can be interpreted as elasticities.

[18]As for all the attempts enumerated in Appendix 3, all these estimates are available from the authors upon request.

estimate a negative association between immigrants and trade for Rome and Milan. For this reason, in column (6) of Table 3 we have reported the estimates excluding these provinces. Interestingly enough, while the elasticity of imports remains virtually unchanged, the elasticity of exports to immigrants doubles and is in line with that estimated in the past literature. On the basis of this evidence, and the peculiarity of Rome and Milan, we put more emphasis on the estimates excluding these provinces. Accordingly, in later Sections we will only present the 2SLS estimates excluding Rome and Milan from the estimation samples. Inclusion of these provinces generally tends to lower the magnitude and significance level of the elasticity of exports with respect to immigration, while it has no effect for imports.[19]

The third concern regards the appropriate choice of the spatial unit in case there is evidence of a MAUP. As we explained above, a unique feature of our dataset is that it gives us the opportunity to deal with the MAUP in a very straight way, assessing the impact of the size of the areal units chosen on the magnitude and the significance of trade elasticities to immigration. In particular, since both trade and immigrant stocks are available at a very disaggregated geographical level in our data, i.e. Italian provinces, we can collapse them at the regional level, estimating, for the sake of comparison, a gravity equation using region-level data.

The estimated gravity equation becomes:[20]

$$\ln(1 + X_{rjt}) = \boldsymbol{\delta}_{rj} + \boldsymbol{\theta}_{jt} + \boldsymbol{\phi}_{rt} + \alpha \ln(Y_{rt-1}\ Y_{jt-1}) + \beta\ \ln(1 + IMM_{rjt-1}) + \epsilon_{rjt} \tag{2}$$

where r is the subscript for Italian regions (NUTS-2), j indicates the foreign country, and t is the subscript for time. $\boldsymbol{\delta}_{rj}$ are *region* $\times$ *country* (*trading-pair*) fixed effects, $\boldsymbol{\theta}_{jt}$ are *country* $\times$ *year* fixed effects, and $\boldsymbol{\phi}_{rt}$ are *region* $\times$ *year* fixed effects. X_{rjt} are exports (or imports) from region r towards (from) country j at time t. IMM_{rjt-1} is the stock of immigrants originating from country j and located in the Italian region r; Y_{rt-1}, Y_{jt-1} are region and foreign countries' GDPs at time $t-1$, and ϵ_{rjt} is an error term clustered at the region by country level. As the main aim of the region-level analysis is to provide a term of comparison to assess the existence of a potential MAUP, equation (2) is estimated using OLS.

As we already did for the province-level analysis, we also report specifications including fewer fixed effects to stress the importance of controlling for unobserved heterogeneity. The OLS estimates are reported in table 4. The model in column (1), which excludes *trading-pair* fixed effects, for consistency with model (1) of table 3 also includes log distance and the contiguity dummy. As is clear, the initial large positive elasticities for

[19] The complete set of estimates is available from the corresponding author upon request.

[20] Note that because *trading-pair* fixed effects are now defined at the *region* $\times$ *country* level ($\boldsymbol{\delta}_{rj}$), when they are included in the specification, the variable distance ($\ln(distance_{rj})$) and the contiguity dummy are dropped from the regression.

exports and imports estimated in column (1) tend to fall when fixed effects are included. In the fifth and sixth column where *trading-pair*, *region* $\times$ *year* and *country* $\times$ *year* fixed effects are included, and where the regions containing Rome and Milan (Lazio and Lombardy) are excluded from the sample, respectively, these elasticities even turn out to be negative, although they are never statistically significant.

It is interesting to compare the *benchmark* estimates in the province-level analysis (columns (5) and (6) of table 3) and in the region-level analysis (column (5) and (6) of table 4). As is clear from the comparison, the use of a finer geographic unit of analysis increases the estimated elasticity. One potential explanation is that considering regions instead of provinces may produce an *aggregation bias* associated with the spatial unit in use. Just to fix ideas, let's think of two regions A and B which are equally populated both in terms of natives and immigrants. The two regions have two provinces each, 1 and 2. Let's now assume that the main source of the pro-trade effects of immigration is the information conveyed through the interaction between natives and immigrants (i.e. the *business and social network effect à la* Rauch and Trinidate (2002)). Imagine that in region A natives and immigrants are equally split between provinces (A_1 and A_2), while in region B there is perfect segregation, that is all immigrants live in province B_1 and all natives live in province B_2. If we use provinces as the unit of analysis, the magnitude of the trade-creating effect of immigrants is estimated correctly and due only to provinces A_1 and A_2, where immigrants live with natives. However, if regions are used as the unit of analysis this creates a sort of measurement error in the immigration variable. In particular, it appears as if in both regions the level of interaction between natives and immigrants was the same (as both regions have the same number of immigrants) causing a downward bias in the elasticity of trade with respect to the stock of immigrants. Excluding from the sample the NUTS-2 regions including the provinces of Rome and Milan, Lazio and Lombardy respectively, as reported in column (6) of Table 4, unlike in the province-level estimates, does not change the elasticities neither for exports nor for imports.

In summary, the comparison of the province-level with the region-level results suggests that the MAUP is an important issue in the analysis of the pro-trade effects of immigrants. This implies that the geographical aggregation of the areal units makes a difference when estimating the relation of interest, and accordingly that geographic distance between natives and immigrant networks is a key factor. Indeed, were these effects not affected by distance, we would not observe any difference in the coefficient estimates when assessing the link at the province or at the region level, i.e. using an aggregation of areal units positively correlated with distance. For this reason, in the next section we investigate more directly the potential importance of geographical spillovers and inter-provincial distance for the trade-immigration nexus.

Table 4: Gravity equations for exports and imports (OLS) — regional level

	(1)	(2)	(3)	(4)	(5)	(6)
(a) *Exports*						
$\ln(Y_{rt-1}Y_{jt-1})$	1.541***	0.800***	-0.332	0.852***	0.627*	0.640
	(0.024)	(0.145)	(0.833)	(0.144)	(0.363)	(0.398)
$\ln(1+IMM_{rjt-1})$	0.345***	0.028	-0.029	0.016	-0.042	-0.068
	(0.022)	(0.064)	(0.078)	(0.065)	(0.078)	(0.082)
EU_{jt}, $EFTA_{jt}$	-0.363***	-0.039	0.988	-0.041	-0.405	-0.441
	(0.129)	(0.067)	(0.733)	(0.069)	(0.371)	(0.401)
$\ln(distance_{rj})$	-1.260***					
	(0.061)					
$contiguity_{rj}$	-1.246***					
	(0.424)					
Fixed effects:						
year	Yes	Yes				
region × *year*			Yes		Yes	Yes
country × *year*				Yes	Yes	Yes
trading-pair		Yes	Yes	Yes	Yes	Yes
Sample:						
including Lazio and Lombardy	Yes	Yes	Yes	Yes	Yes	
N. observations	25,853	25,853	25,853	25,853	25,853	23,249
R-squared	0.66	0.90	0.91	0.90	0.91	0.90
N. clusters 3740	3740	3740	3740	3740	3740	3366
(b) *Imports*						
$\ln(Y_{rt-1}Y_{jt-1})$	1.718***	0.654***	-0.159	0.693***	-0.201	-0.248
	(0.027)	(0.192)	(0.955)	(0.195)	(0.564)	(0.624)
$\ln(1+IMM_{rjt-1})$	0.647***	-0.117	-0.084	-0.129	-0.096	-0.077
	(0.030)	(0.085)	(0.097)	(0.086)	(0.098)	(0.103)
EU_{jt}, $EFTA_{jt}$	0.687***	0.245	0.351	0.245	0.258	-0.078
	(0.167)	(0.153)	(1.430)	(0.153)	(0.818)	(0.907)
$\ln(distance_{rj})$	-0.742***					
	(0.077)					
$contiguity_{rj}$	-0.602					
	(0.611)					
Fixed effects:						
year	Yes	Yes				
region × *year*			Yes		Yes	Yes
country × *year*				Yes	Yes	Yes
trading-pair		Yes	Yes	Yes	Yes	Yes
Sample:						
including Lazio and Lombardy	Yes	Yes	Yes	Yes	Yes	
N. observations	25,853	25,853	25,853	25,853	25,853	23,249
R-squared	0.67	0.89	0.90	0.89	0.90	0.90
N. clusters 3740	3740	3740	3740	3740	3740	3366

*, **, *** statistically significant at the 10, 5 and 1 percent level, respectively.

Note. The dependent variable is $\ln(1+export_{rjt})$ for panel (a) and $\ln(1+import_{rjt})$ for panel (b), i.e. export (import) flows of region r to (from) country j at time t. *Trading-pair* fixed effects are defined at the *region* × *country* level. Export and import flows cover the period 2003-2009. Standard errors are clustered at the region by (importer or exporter) country level. In column (6) we exclude from the sample the regions of Lazio and Lombardy, containing the cities of Rome and Milan, respectively.

4.2. Geographic spillovers of immigration on trade: the role of geographic proximity

Most of the papers existing in the literature, and the specification of equation (1), do not allow for geographic spillovers from immigration on trade. However, given our previous discussion on the role of the appropriate areal unit of analysis, taking into account potential spillovers is important. Province i may, for instance, have a low stock of immigrants originally coming from a specific country, but may be surrounded by provinces in which immigrants from that country are located in large quantity. Now, depending on the geographic scale of ethnic networks, this province may also benefit from proximity to immigrant-abundant provinces. Moreover, if there are inter-province spillovers, and the provinces' immigrant stocks are spatially correlated, then the coefficient on a province's own immigration may also pick up the effect of immigrants located in other provinces, generating an omitted variables bias.

This issue has been investigated by Herander and Saavedra (2005) for the U.S.. The authors compare the estimated effects of local immigrant populations on U.S. states' exports to the effects of out-of-state populations of the same immigrant group and find the former effect to be greater than the latter. They interpret this result as consistent with the importance of proximity to immigrant networks. A similar approach is adopted in a recent article by Artal-Tur *et al.* (2012) who investigate the same issue for Italy, Spain and Portugal using NUTS-3 level data. As a proxy of spillovers, they consider the immigrants outside the province, and the immigrants in adjacent and non-adjacent provinces in a log-log specification estimated with OLS and find that all these variables, unlike the province's own stock of immigrants, are not significantly associated with trade.

Here, we exploit the very fine spatial disaggregation of Italian trade and immigration data to shed light on the importance of geographic proximity for the trade-creating effect of immigrants. In our specification, spillovers are captured by the number of immigrants of a given ethnicity located outside the province within a distance of 25 km, between 25 and 50 km, between 50 and 100 km, between 100 and 200 km, and over 200 km.[21] The simultaneous inclusion of all these controls allows us to have a clearer idea of the existence of, and of the spatial decay of the spillovers related to ethnic networks. Indeed, we expect the amount of contacts between individuals living in different provinces, and knowledge spillovers accordingly, to decline

[21] Distances are computed using the provinces' centroids. More in detail, the spillover variables were computed by aggregating the stocks of immigrants of all provinces falling within the radius d.

with distance. The specification becomes

$$\ln(1+X_{ijt}) = \boldsymbol{\delta}_{rj} + \boldsymbol{\theta}_{jt} + \boldsymbol{\phi}_{rt} + \alpha \ln(Y_{it-1}\ Y_{jt-1}) + \beta\ \ln(1+IMM_{ijt-1}) + \\ + \sum_{d=1}^{5} \lambda^d ln(1+IMM^d_{ijt-1}) + \gamma \ln(distance_{ij}) + \rho\ \ contiguity_{ij} + \epsilon_{ijt} \tag{3}$$

and d its a radius that takes values from 1 to 5, for immigrants of ethnicity j living in a different province within a distance of 25 km, 25-50 km, 50-100 km, 100-200 km and more than 200 km from province i, respectively.

The results from the OLS estimation are reported in table 5 and refer to the sample excluding Rome and Milan.[22] Column (1) shows the results for exports. Immigrants of the ethnic group j, located within 50 km from province i have a positive effect on export flows from province i to the foreign country j, yet their effect is much smaller than that of the stock of immigrants from j living in the province i. At longer distances the positive effect weakens, and even turns significant and negative for ethnic networks which are located too far away from province i. We interpret this as evidence of a *trade-diversion* effect. Indeed, the existence of provinces sufficiently distant from province i which are abundant in immigrants from country j puts them in a better position than province i as far as trading with country j is concerned. In simple terms, provinces seem to compete for the trade enhancing factor embedded in immigrants. A similar pattern is observed for imports in column (2). Positive spillovers are observed within a distance of 100 km, while they change sign for longer distances. Taken together these results are broadly consistent with the findings of Herander and Saavedra (2005) and Artal-Tur *et al.* (2012) about the importance of geographic proximity in ethnic networks, but we also find a trade-diversion effect (negative spillovers) when large stocks of immigrants are located 'too far away', and other provinces benefit from them.

4.3. Endogeneity and two-stage least squares

The OLS results in section 4.1 confirm the evidence in the raw data and suggest that immigrants may have a positive effect on both imports and exports, but larger for the former. However, a potential pitfall with the OLS estimates is that even after controlling for *trading-pair* fixed effects immigrant inflows may be endogenous with respect to export or import flows. The endogeneity problem could be determined, for instance, by trading-pair *time-variant* unobservables which simultaneously affect immigrant flows and trade. To the best of our knowledge there have been few attempts to address this issue in the literature. Combes *et al.* (2005) study the role of domestic migrations in shaping trade between French regions. They seek to

[22] We have already discussed in the previous section and in Appendix 4 the reasons for doing this.

Table 5: Geographic spillovers of immigration on trade (OLS) — province level

	(1) Exports	(2) Imports
$\ln(Y_{it-1}Y_{jt-1})$	2.317***	2.160***
	(0.032)	(0.038)
$\ln(1+IMM_{ijt-1})$	0.118***	0.345***
	(0.020)	(0.028)
$\ln(1+IMM_{ijt-1}^{<25})$	0.038***	0.037***
	(0.004)	(0.007)
$\ln(1+IMM_{ijt-1}^{25-50})$	0.032***	0.035***
	(0.005)	(0.007)
$\ln(1+IMM_{ijt-1}^{50-100})$	0.013*	0.033***
	(0.008)	(0.008)
$\ln(1+IMM_{ijt-1}^{100-200})$	-0.001	-0.034***
	(0.008)	(0.009)
$\ln(1+IMM_{ijt-1}^{>200})$	-0.067***	-0.069***
	(0.021)	(0.019)
$\ln(distance_{ij})$	-1.569***	-2.695***
	(0.395)	(0.594)
$contiguity_{ij}$	0.100	-0.264
	(0.360)	(0.372)
N. observations	132,982	132,982
R-squared	0.797	0.766

*, **, ***, significant at the 10%, 5% and 1% statistical level

Note. The dependent variables are $\ln(1 + export_{rjt})$ and $\ln(1 + import_{rjt})$, i.e. export (import) flows of region r to (from) country j at time t. The regressions include all sets of fixed effects as in column (5) of Table 3. *Trading-pair* fixed effects are defined at the *region* × *country* level. Export and import flows cover the period 2003-2009. Standard errors are clustered at the province by (importer or exporter) country level. Rome and Milan have been excluded from the estimation samples.

address the potential endogeneity of migrations by using the lagged stock of immigrants (15 years before). A similar instrument (lagged immigrant stock) is used by Briant *et al.* (2009) who focus on the trade-creating effect of foreign migrants in French departments. In both cases, the authors find very similar elasticities when using OLS and 2SLS. An important feature of their analysis is that they use cross-sectional data and are not able to control for unobserved heterogeneity at the trading-pair level.[23] An attempt to address the endogeneity of immigrants in panel data is made by Peri and Requena-Silvente (2010). The authors adopt a more complex way of building a potential instrument, based on supply-push factors and motivated by the presence of historical immigrant enclaves, as in Altonji and Card (1991). We apply the same strategy in the present case. The presence of a community of immigrants from a given country in a certain Italian province is likely to decrease immigration costs and increase returns to migration for new immigrants of the same nationality that settle in that province. Indeed, co-nationals already present in a province may offer hospitality, financial support or help new migrants to find a job in the local labor market. For these reasons, we expect the stock of immigrants to be highly correlated with the inflow of new immigrants. Accordingly, we adopt the following procedure to build an instrumental variable. We compute the total stock of immigrants by country for Italy as a whole in each year, and we allocate it to each province according to the distribution of immigrants by nationality across provinces in 1995. This 'base' year for computing weights is chosen on the grounds that before 1995 in Italy there were only 95 provinces, and considering early years the weight would be zero for eight provinces between 2002 and 2006, and for twelve provinces between 2006 and 2009.[24] Since province-level data on the stock of immigrants by nationality are not provided by ISTAT for 1995, we use the distribution of immigrants' requests for residence permits ('permessi di soggiorno') provided by the Ministry of Interior. In this way, we compute an *imputed stock of immigrants*, which is used as an instrument for the observed stock.

More in detail, define IMM_{ijt} the number of immigrants from country j located in province i at time t, and IMM_{jt} the total stock of immigrants from country j at time t in Italy. Then the share of total migrants of the nationality j residing in province i at time t can be defined as

[23] Indeed, Combes *et al.* (2005) only consider 1993, while Briant *et al.* (2009) average trade flows over three years (1998, 1999 and 2000) for each *département*-country pairs. Hence, both these studies are unable to account for trading-pair unobserved heterogeneity.

[24] As we said above, in 2006 four new provinces were created (in Sardinia), raising the total number of provinces from 103 to 107. Hence, even fixing the base year at 1995, the instrument assumes value zero for these four provinces. We avoided imputing weights based on subjective assumptions, but checked the sensitivity to this issue by dropping observations for Sardinia after 2006, and did not obtain notable differences in the 2SLS estimates.

$$w_{ijt} = \frac{IMM_{ijt}}{IMM_{jt}}. \tag{4}$$

After considering the lagged distribution of immigrants by nationality across provinces w_{ij95} and having defined IMM_{j0} the total stock of immigrants from country j at the first year of the time interval we consider (time zero, that is 2002), the imputed stock of immigrants becomes

$$\widehat{IMM}_{ijt} = w_{ij95}IMM_{j0} + w_{ij95}\sum_{q=0}^{t} F_{jq} = w_{ij95}(IMM_{j0} + \sum_{q=0}^{t} F_{jq}) = w_{ij95}IMM_{jt} \tag{5}$$

where F_{jq} is the total net inflow of immigrants from country j in Italy at time q. The instrument is then given by the product of two terms, the first (w_{ij95}) exhibits trading-pair variation, and the second (IMM_{jt}) country by time variation. Thanks to the product of the two terms, the imputed stock of immigrants varies simultaneously at the province, country and time level. Imagine now that as in our paper the immigrant stock is measured in logarithms. After taking logs, equation (5) becomes

$$ln(\widehat{IMM}_{ijt}) = ln(w_{ij95}) + ln(IMM_{jt}) \tag{6}$$

that is linear in two terms, one varying at the trading-pair level and the other at the country-year level. This means that if one includes in the gravity equation *province* $\times$ *country* and *country* $\times$ *year* fixed effects, they completely absorb the instrument's variation. Adding one to the stock of migrants before taking logs makes it no longer possible to write the imputed stock of immigrants as a linear function of $ln(w_{ij95})$ and $ln(N_{jt})$, that is

$$ln(1 + \widehat{IMM}_{ijt}) = ln(1 + w_{ij95}IMM_{jt}) \neq ln(w_{ij95}) + ln(IMM_{jt}) \tag{7}$$

and the variation in the log of the modified imputed stock of immigrants (i.e., the imputed stock after adding one) is now not completely absorbed by *trading-pair* and *country* $\times$ *year* fixed effects. However, in the 2SLS estimates identification does stem only from the nonlinearity of $ln(1+\widehat{IMM}_{ijt})$ in these fixed effects, which may be too weak in many cases (see Table 10 where only few residual variation remains after controlling for *trading-pair* fixed effects at the *country* $\times$ *province* level).[25]

As we said in section 4.1, we do not only exploit this non-linearity, but also cross sectional variation between provinces within the same region, as we are using *trading-pair* defined at the *region* $\times$ *country* level.

[25] Peri and Requena-Silvente (2010) estimate, for instance, the impact of immigrants on trade of Spanish regions, and include in the gravity equation both *region* $\times$ *country* and *country* $\times$ *year* fixed effects. First, they add one to both trade and the immigrant stock to retain observations with zeros in their logarithmic specification, and, on top of that, in the 2SLS specification using the imputed stock of immigrants as the excluded instrument, they omit *trading-pair* dummies from the first stage.

The main threat to identification comes from time-varying trading-pair unobserved factors during the period observed which simultaneously affect provinces' trade with a given country and the stock of immigrants from that country. In this respect, the main determinants of the imputed stock of immigrants in equation (5) are presumably exogenous, i.e. uncorrelated with such unobservables. Indeed, the net immigration flows by country to overall Italy in each year 2002-2008 (F_{jz}) and the stock of immigrants by nationality in Italy in 2002 (IMM_{j0}), referring to the whole country, should not be affected by trading-pair *shocks*, especially when shocks are related to very small geographical units, such as Italian provinces. As for the remaining component, the distribution of residence permits by nationality across provinces in 1995 (w_{ij95}), conditional on *trading-pair* fixed effects, *region* $\times$ *year* and *country* $\times$ *year* fixed effects, should not be theoretically correlated with any trading-pair time variant shock taking place during the estimation period (2003-2009), especially given that our geographic units of analysis (provinces) are relatively small. It may happen, for instance, that immigrants decided to locate in specific provinces in 1995 because they were predicting more local opportunities for trade growth with their home countries. However, controlling for *trading-pair* (*region* $\times$ *country*) fixed effects helps to address this potential criticism, as the variation in the instrument which we are effectively exploiting is only the one between provinces within the same region and for the same country of origin of immigrants. For the instrument to fail it must be the case, for instance, that immigrants from a specific foreign country back in 1995 chose to locate in province A_1 rather than in province A_2 (both in region A) because A_1 offered the best opportunities for them to trade with their home country. However, this is very unlikely as provinces within the same region present the same (or very similar) institutional and, often, socio-economic characteristics, and therefore immigrants would have the same opportunity to trade with their home country irrespective of the specific choice of province of initial location. This is clearly an advantage of using very small geographical units of analysis, as this assumption would be much more difficult to maintain in country-level or even in region-level analyses. Moreover, the small size of provinces is important as it makes it less likely that migration flows from a specific foreign country toward a single Italian province account for most immigration from that country towards Italy.

The results of the 2SLS estimation are reported in Table 6. Columns (1) and (2) report the estimates for the model without spillovers. In the first-stage, the instrument turns out to be very strong, with an F-test well above the threshold of 10 suggested by Staiger and Stock (1997) to detect a potential weak instrument problem.[26] Column (1) shows that the export elasticity to immigrants is 0.13, similar to the one estimated

[26] The coefficient on the predicted stock of immigrants in the first stage is 0.418 with a t-value of 57.42.

Table 6: Gravity equations for exports and imports (2SLS) — province level

	without spillovers		with spillovers	
	(1) Exports	(2) Imports	(3) Exports	(4) Imports
Second stage				
$\ln(1+IMM_{ijt-1})$	0.128***	0.594***	0.101**	0.580***
	(0.040)	(0.057)	(0.040)	(0.056)
$\ln(1+IMM^{<25}_{ijt-1})$			0.044***	0.026***
			(0.005)	(0.008)
$\ln(1+IMM^{25-50}_{ijt-1})$			0.019**	0.041***
			(0.007)	(0.011)
$\ln(1+IMM^{50-100}_{ijt-1})$			-0.031	0.035
			(0.024)	(0.027)
$\ln(1+IMM^{100-200}_{ijt-1})$			-0.075*	-0.184***
			(0.041)	(0.044)
$\ln(1+IMM^{>200}_{ijt-1})$			-0.727***	-0.457**
			(0.214)	(0.207)
$\ln(Y_{it-1}Y_{jt-1})$	2.291***	1.954***	2.316***	2.018***
	(0.041)	(0.051)	(0.042)	(0.052)
$\ln(distance_{ij})$	-1.603***	-2.621***	-1.570***	-2.630***
	(0.377)	(0.576)	(0.380)	(0.577)
$contiguity_{ij}$	0.035	-0.596	0.136	-0.492
	(0.352)	(0.384)	(0.358)	(0.395)
First Stage F-test$^{(a)}$				
$\ln(1+IMM_{ijt-1})$	3,297.02	3,297.02	577.25	577.25
$\ln(1+IMM^{<25}_{ijt-1})$			8,424.22	8,424.22
$\ln(1+IMM^{25-50}_{ijt-1})$			2,431.37	2,431.37
$\ln(1+IMM^{50-100}_{ijt-1})$			590.88	590.88
$\ln(1+IMM^{100-200}_{ijt-1})$			278.21	278.21
$\ln(1+IMM^{>200}_{ijt-1})$			338.49	338.49
N. observations	132,982	132,982	132,982	132,982
R-squared	0.175	0.138	0.166	0.134

*, **, ***, significant at the 10%, 5% and 1% statistical level

$^{(a)}$ We have reported for the sake of brevity only the F-tests for the excluded instruments. In the spillover specifications there are five first stages, one for each of the six endogenous variables.

Note. The dependent variables are $\ln(1+export_{ijt})$ and $\ln(1+import_{ijt})$, i.e. export and import flows of province i from country j at time t, respectively. The regressions also control for $region \times country$, $country \times year$ and $region \times year$ fixed effects. Trade flows cover the period 2003-2009. Robust standard errors are clustered at the $province \times country$ level and are reported in parenthesis. Rome and Milan have been excluded from the estimation sample.

with OLS (0.12). The estimated elasticity of imports is instead 0.59, larger than the one obtained with OLS, and larger than for exports. Thus our estimates of the pro-trade effect of immigration are totally in line with the literature for exports (see section 2), but larger for imports.[27] However, it must be noted that most previous studies have used OLS, and often included a much smaller set of fixed effects.

In column (3) and (4) we report the specifications augmented with the spillover variables (see section 4.2), which have been instrumented using the same shift-and-share procedure used to instrument a province's own stock of immigrants. The 2SLS estimates show that both for imports and for exports positive spillovers have a quick spatial decay, and disappear above 100 km. Also when using 2SLS, like for OLS, we find large negative spillovers for immigrants located above 200 km of distance. Importantly, comparison of column (1) with (3), and (2) with (4), suggests that the estimated effect of a province's immigrant stock is only marginally affected by the inclusion of the spillover variables.

4.4. Heterogenous effects

When the effect of immigrants is heterogeneous, the IVs estimates reported in the previous section can be interpreted as Local Average Treatment Effects (LATE), i.e. as the effect on trade of the immigrants whose stocks are affected by the excluded instrument (the imputed stock of immigrants). In the presence of heterogenous effects, the LATE is an 'average' effect and may hide substantial differences across immigrants' countries of origin, or along other dimensions. In this section, we exploit the two-way geographical heterogeneity of our data to explore differential effects in the trade-creating effect of immigrants.

As we have already stressed, Italy provides two interesting features: (i) the 'super-diversity' of immigration; (ii) a remarkable socio-economic diversity across provinces, namely a strong North-South divide where Southern provinces lag behind in terms of economic development. As far as the first point is concerned, if immigrants are likely to reduce information costs, and contribute to enforcing contracts, this is likely to happen the higher the level of development of the countries where they come from. To test this hypothesis, we have classified immigrants' countries in four groups by level of Gross National Income per capita (in U.S. dollars): low, middle-low, middle-high and high income. The classification was made by the World Bank, using an Atlas conversion factor to reduce the effect of exchange rate fluctuations on cross-country income comparison and the list of countries in each category is included in Appendix 5. The results of the 2SLS estimation in the four samples are reported in columns (1)-(4) of Table 7. For the high-income countries no

[27] In the IZA Discussion Paper's version we reported the estimates for the sample including Rome and Milan. When including these cities the elasticity of exports falls to 0.004 and ceases to be statistically significant (s.e. = 0.038), while the elasticity of imports is only marginally affected (0.548, s.e. = 0.053). This issue is discussed in Section 4.1 and Appendix 4.

significant effect neither on exports nor on imports emerge. This is consistent with those countries having well developed markets and institutions, which are likely to provide rich trade-related information and ensure contract enforcement without having to rely on their citizens abroad. Moreover, models of consumption are likely to be very similar across high-income countries, and we do not expect a strong home-bias effect in consumption. Switching to the middle-high income category, we still do not observe any effect on exports, but a positive and significant effect on imports emerges. We tend to rationalize this effect as the consequence of partial differences in models of consumption between high-middle income countries and Italy, a high-income country. At lower levels of income, we find statistically significant effects both on exports and imports. For these countries, where market institutions are not well developed, immigrants are more likely to provide a valuable source of information and contract enforcement thanks to their persisting links with the home country. For middle-low and low-income countries, a home bias is also likely to emerge, as the kind of goods immigrants were used to consume at home (e.g., of low quality, low price) are less likely to be produced by Italian firms.

Another source of heterogeneity that we investigate is between 'new' (i.e. recently arrived) and 'old' communities of immigrants. We defined two waves of immigration. The 'first wave' includes the first twenty countries of origin of immigrants, ordered according to the share in overall foreign born population in Italy at the beginning of the period (2002). The 'second wave' is defined similarly, but considering the first twenty countries of origin at the end of the period (2008), excluding those that were in the first wave's sample to avoid overlapping between the two groups. The list of countries is reported in Appendix 5. The results are reported in column (5) and (6) of Table 7. Interestingly, only first-wave immigrants seem to spur trade. A possible rationalization for this evidence is that knowledge flows from immigrants to natives need time to materialize, and accordingly recently arrived communities of immigrants are unable to spur trade. As for imports, even if there is a home-bias in consumption, and new immigrants would prefer to consume their home goods, it takes time to organize import activities. For instance, imports from origin countries may require immigrants to create import-export firms (ethnic firms), or it may take time for 'native' firms to evaluate the existence of a potential market for 'ethnic goods', and its size. We leave further investigation of these hypotheses for future work.

Italy is characterized by a very strong North-South divide. Firms operating in the South are less productive (Aiello and Scoppa, 2000) due to the inefficiencies in the public sector, the presence of organized crime, credit constraints, less developed institutions, which along with other factors make them also less likely to export (Castellani, 2002). Firms located in the North of Italy are instead more productive and well

Table 7: Heterogeneous effects (2SLS) – Second stages

	Income Level				Migration Wave		Geographic
	High	High-Mid.	Low-Mid.	Low	First	Second	interactions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Second stage Exports:							
$\ln(1+IMM_{ijt-1})$	-0.024	-0.042	0.313***	0.479***	0.462***	0.275	0.470***
	(0.085)	(0.087)	(0.063)	(0.106)	(0.083)	(0.170)	(0.073)
$\times North$							-0.477***
							(0.069)
$\times Centre$							-0.529***
							(0.089)
$\ln(Y_{it-1}Y_{jt-1})$	1.841***	2.346***	2.091***	2.758***	1.430***	1.909***	2.276***
	(0.096)	(0.088)	(0.069)	(0.105)	(0.117)	(0.193)	(0.041)
$\ln(distance_{ij})$	-0.711*	-2.579***	-3.765	-1.129	-1.307*	-2.139**	-1.457***
	(0.423)	(0.709)	(2.522)	(4.368)	(0.767)	(0.862)	(0.381)
$contiguity_{ij}$	0.360				-1.000*		0.219
	(0.248)				(0.516)		(0.320)
N. observations	21,450	36,053	38,299	22,880	14,300	14,300	132,982
R-squared	0.302	0.188	0.150	0.189	0.220	0.211	0.174
Second stage Imports:							
$\ln(1+IMM_{ijt-1})$	-0.061	0.439***	0.859***	0.634***	0.618***	-0.096	0.678***
	(0.111)	(0.134)	(0.094)	(0.134)	(0.134)	(0.244)	(0.087)
$\times North$							-0.067
							(0.089)
$\times Centre$							-0.270**
							(0.127)
$\ln(Y_{it-1}Y_{jt-1})$	2.141***	2.348***	1.813***	1.833***	2.062***	2.893***	1.944***
	(0.128)	(0.123)	(0.087)	(0.123)	(0.186)	(0.279)	(0.051)
$\ln(distance_{ij})$	-1.981***	-2.789**	-11.193***	-0.225	-2.766**	-2.558*	-2.602***
	(0.448)	(1.322)	(3.593)	(4.902)	(1.099)	(1.534)	(0.578)
$contiguity_{ij}$	0.291				-1.897**		-0.612
	(0.313)				(0.954)		(0.389)
N. observations	21,450	36,053	38,299	22,880	14,300	14,300	132,982
R-squared	0.244	0.154	0.133	0.107	0.218	0.185	0.138

*, **, ***, significant at the 10%, 5% and 1% statistical level

Note. The dependent variables are $\ln(1+export_{ijt})$ and $\ln(1+import_{ijt})$, i.e. export and import flows of province i from country j at time t, respectively. The regressions also control for $region \times country$, $country \times year$ and $region \times year$ fixed effects. Trade flows cover the period 2003-2009. Robust standard errors are clustered at the $province \times country$ level and are reported in parenthesis. Rome and Milan have been excluded from the estimation samples.

Table 8: Heterogeneous effects (2SLS) – First stages

	Income Level				Migration Wave		Geographic interactions		
	High	High-Mid.	Low-Mid.	Low	First	Second			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First stage:									
$\ln(1+\widehat{IMM}_{ijt-1})$	0.309***	0.351***	0.484***	0.427***	0.420***	0.229***	0.408***	-0.0582***	-0.014***
	(0.007)	(0.006)	(0.006)	(0.008)	(0.009)	(0.008)	(0.011)	(0.002)	(0.001)
$\times North$							0.027*	0.549***	-0.000
							(0.014)	(0.010)	(0.000)
$\times Centre$							-0.018	0.017***	0.475***
							(0.018)	(0.002)	(0.016)
$\ln(Y_{it-1}Y_{jt-1})$	0.591***	0.430***	0.327***	0.418***	0.584***	0.733***	0.363***	0.187***	0.046***
	(0.011)	(0.008)	(0.007)	(0.010)	(0.017)	(0.015)	(0.008)	(0.006)	(0.003)
$\ln(distance_{ij})$	-0.491***	-0.666***	0.340	-1.442**	-0.855***	-0.827***	-0.439***	0.023	-0.085*
	(0.104)	(0.147)	(0.361)	(0.570)	(0.154)	(0.186)	(0.148)	(0.105)	(0.047)
$contiguity_{ij}$	0.684***				0.501***		0.517**	0.519**	0.006
	(0.106)				(0.136)		(0.236)	(0.241)	(0.012)
N. observations	21,450	36,053	38,299	22,880	14,300	14,300	132,982	132,982	132,982
R-squared	0.568	0.467	0.479	0.399	0.604	0.514	0.460	0.462	0.358
F-test instruments$^{(a)}$	410.62	728.65	1,177.59	621.16	469.41	153.04	1,108.04	1,529.90	456.06

*, **, ***, significant at the 10%, 5% and 1% statistical level

Note. $^{(a)}$ In column (7) we have reported the first stage for $\ln(1+IMM_{ijt-1})$, while in column (8) and (9) we have reported first stages for interacted variables, respectively, $\ln(1+IMM_{ijt-1})\times North$, and $\ln(1+IMM_{ijt-1})\times Centre$. The regressions also control for $region \times country$, $country \times year$ and $region \times year$ fixed effects. Robust standard errors are clustered at the $province \times country$ level and are reported in parenthesis. Rome and Milan have been excluded from the estimation samples.

integrated in European markets, as reflected by the average trade openness; over the sample period, in fact, the ratio between exports plus imports over GDP for the provinces located in the North of the country is about 0.51, while for the Center it is 0.39, and for the South and Islands it is 0.24 — far below the country's average 0.39. Thus immigrants may spur firms' exporting activities especially in the South. To investigate this hypothesis, the benchmark specifications of the export and import equations have been modified to include interaction terms between the stock of immigrants and macro-area dummies (North, Center, South and Islands).[28] The results are reported in column (7) of Table 7. In line with the theoretical predictions, a significant positive effect of immigrants on exports is estimated only for Southern Italy. The effect on imports turns out instead to be positive irrespective of the macro-region considered, suggesting again an important role for home-bias in fostering import flows.

The first-stage results of all models are reported in Table 8 and show no evidence of a weak instruments' problem.

5. Concluding remarks

This paper seeks to contribute to the literature on the causal effect of immigrants on international trade. The Italian case provides an ideal setting for shedding light on the trade-immigration link thanks to the large number of ethnic groups (187) present in the Italian territory, i.e. the Italian "*super-diversity*," the very fine geographical disaggregation of the immigration and trade data (at the province level, NUTS-3), and the lack of colonial, language or cultural ties with immigrants' origin countries. The potential endogeneity of immigration is addressed using a shift-and-share approach based on the existence of immigrant *enclaves*. We have a number of interesting results. First, distance seems to be a key factor for the ethnic networks relevant to international trade. Indeed, although a province's own stock of immigrants is the most important factor for its international trade, we find positive spillovers of immigrants within a distance of 50 km for both exports and imports. Interestingly enough, the spillovers change in sign when large communities of immigrants are too far away and other provinces are presumably benefiting from them (trade-diversion effect). Second, our IVs estimates generally point to a larger effect of immigrants on imports than on exports. This is in line with the theoretical expectations, since the former are affected both by the *transplanted home-bias effect* and the *business and social network effect* of immigrants on trade, while the latter only by the second effect. Using IVs, we estimate an elasticity of exports with respect of immigrants of 0.128 and of imports of 0.594.

[28] These interaction terms have been instrumented with interaction terms between macro-region dummies and the predicted stocks of immigrants.

These figures imply that one more immigrant coming from a specific foreign country, settling in a specific province would directly increase the average value of that province's exports toward that very country by 32,865 U.S. dollars and of imports from the same country by 14,049 U.S. dollars (effects computed at the sample mean, excluding Rome and Milan). Last but not least, we show that the average effect of immigrants on trade hides a substantial heterogeneity. Consistent with the idea of immigrants providing a valuable source of information and contract enforcement when market institutions are not well developed, we find the pro-trade effect of immigrants to be larger for immigrants coming from low-income countries, and for Southern Italy's provinces, in which firms are less productive and efficient. Moreover, trade creation appears to be larger for 'first-wave' immigrants, those who settled in Italy earlier, suggesting that it may take time for information to flow from immigrants to natives. An alternative rationalization of this last result is that the pro-trade effect of immigrants may originate not from information spillovers to natives but from immigrants exploiting their superior knowledge of the home country by organizing themselves import-export activities through ethnic firms. We leave further investigation of this hypothesis for future work.

Acknowledgements. The research project started while Massimiliano Bratti and Luca De Benedictis were visiting UC Berkeley in 2010. Massimiliano Bratti wishes to thank IRLE and CLE and Luca De Benedictis the ARE Department for their hospitality. We thank Rosalia Alessi, Antonella Ciccarese, Elena Mazzeo, and Alessia Proietti for their generous help with the data, and Massimo Tamberi for the references on super-diversity in migration. We also thank all the people who gave comments on previous versions of the paper, in particular Antonio Accetturo, Paolo Buonanno, Tommaso Frattini, Hubert Jayet, Giovanni Pica and participants in seminar and conference presentations at the Polytechnic Institute of Milan, ERSA 2011 (Barcelona), ETSG 2011 (Copenhagen), ITSG 2011 (Milan), AIEL 2011 (Milan), SIE 2011 (Rome), 'Brucchi Luchino' 2011 (Rome), CIE 2012 (Granada) and EGI 2012 (Bari). A previous and partly different version of this paper circulated under the same title as IZA Discussion Papers 6628. The usual disclaimer applies.

Appendixes

Appendix 1: Main variables' description and descriptive statistics

Trade data. Trade data are taken from the public available database of the Italian Institute of Statistics (ISTAT). Trade flows refer to the value of imports and exports of 107 Italian provinces (NUTS-3) with more than 200 trading partners around the world, over the period 2003-2009. Data are measured so that exports and imports are associated with the province of shipment, i. e. the province where the custom transaction was registered. Information of Extra-EU transactions are based on the *Documento Amministrativo Unico* (DAU) which is filled in for each commercial transaction, for the intra-EU exchanges the custom system has been replaced, since 1993, by the Intrastat standard. Imports and exports from each province are reported in current euros, we than express them in U.S. current dollars, using the nominal exchange rate from WDIs[29].

Immigrants. Data on foreign born residents by province (NUT-3) are taken from the demographic portal of ISTAT, and report the stock of foreign-born residents per province at 31 December of each year (from 2002-2009). In the IV analysis we use data on *residence permits* in 1995, in this period Italian migration policy was mostly based on the residence permit, which allows the holder to legally stay in the country for a given period of time. Residence permits are issued by Italian Ministry of Interior, but data are also collected by the Italian National Institute of Statistics [30].

GDP. Data on country Gross Domestic Product are taken from the *World Development Indicators* (WDIs), and are expressed in current U.S. dollars. The GDP of Italian provinces are taken from ISTAT and then re-scaled to match the value of nominal Italian GDP, as reported in WDIs[31].

EU and *EFTA*. It is a dichotomous variable which equals one in case a country is a EU or EFTA member and zero otherwise.

Distance. It is the country-province distance computed as the geodesic distance between the foreign country's capital and the province's centroid.

Contiguity. It is a dichotomous indicator which equals one if a country and a province are contiguous and zero otherwise.

[29] Trade data by Province are available at `http://www.coeweb.istat.it/english/default.htm`

[30] Foreign born resident data, as well as residence permits records, are available at `http://demo.istat.it/index_e.html`

[31] GDP figures for italian provinces are available at `http://dati.istat.it/?lang=en` ; while for other countries at `http://databank.worldbank.org/ddp/home.do?Step=12&id=4&CNO=2`

Table 9: Sample summary statistics — province-level analysis

	All Sample			Without Rome – Milan		
	N.	Mean	SD	N.	Mean	SD
$\ln(1+export_{ijt})$	135,586	9.811	6.748	132,982	9.707	6.745
$\ln(1+import_{ijt})$	135,586	7.230	7.274	132,982	7.110	7.238
$\ln(Y_{it-1}Y_{jt-1})$	135,586	46.646	2.530	132,982	46.594	2.504
$\ln(1+IMM_{ijt-1})$	135,586	1.851	2.116	132,982	1.800	2.070

Appendix 2: Fixed-effects and model saturation

A significant portion of the variation in our data comes from the *province* $\times$ *country* dimension. It is therefore worth exploring this characteristic in an explicit way, in relation with the variables we are focusing on in the core of the analysis. Therefore, in Table 10 we report the R-squared of different linear regressions having $1+exports$ (in logs), $1+imports$ (in logs) and $1+immigrants$ (in logs) regressed on a different set of dummies. Including the dummies *province*, *country* and *time* separately, allows us to account already for between 73% and 82% of the variability in the data. If instead we substitute the *province* dummy with a *region* dummy, allowing all NUTS-3 regions that are part of the same NUTS-2 region to take the same effect, the R-squared decreases by 5 percentage points. The inclusion of the full set of fixed effects captures almost 90% of the variation in both export and import flows and over 98% of immigration stocks, our main independent variable of interest, when using *province* $\times$ *country* fixed effects in panel (a).

Table 10: R-squared for models including different sets of fixed effects

	Immigrants	Imports	Exports
Panel (a)			
$province, country, year$	.819	.728	.765
$province \times country, year$	.978	.886	.889
$provinccdcde \times country, province \times year$	.979	.887	.890
$province \times country, country \times year$	.983	.889	.892
$province \times country, province \times year, country \times year$	.984	.890	.893
Panel (b)			
$region, country, year$	.766	.679	.711
$region \times country, year$	.836	.728	.760
$region \times country, region \times year$	.837	.729	.761
$region \times country, country \times year$	.842	.731	.762
$region \times country, region \times year, country \times year$	.843	.732	.763

Note: The table reports the adjusted-R^2 obtained regressing, alternatively, imports, exports, and immigrants on different sets of dummies. In all cases the dependent variable is $ln(1+x)$, were x are exports, imports or immigrants (to retain zeros in the estimation).

By contrast, the R-squared is between 73% and 84% when *region* $\times$ *country* fixed effects are included, in panel (b). Any attempt to identify the effect of immigrants on trade, controlling for country-of-origin and provincial heterogeneity using *province* $\times$ *country* dummies, or even more demanding specifications also including time-varying versions of country and province fixed effects (as reported in Table 10), clashes with the very little variation left in the data. In the empirical analysis that follows we therefore opt for the inclusion of a *region* $\times$ *country* specification of the *trading-pair* dummies.

Appendix 3: Zeros in trade

The log-log version of the gravity model in equation (1) has been recently subject to some criticism by Santos Silva and Tenreyro (2006). Since this seminal contribution, the debate on the most appropriate nonlinear estimator to be applied when zeros are a relevant proportion of trade flows is still very open (see De Benedictis and Taglioni (2011) on this specific point of the gravity literature). In the present analysis, the use of the Pseudo Poisson Maximum Likelihood estimator proposed by Santos Silva and Tenreyro (2006) clashes with the inclusion of many fixed effects. As the trading-pairs are defined at the *region* $\times$ *country* level (to exploit province-level variation within regions), they cannot be 'partialled out' using a fixed effect poisson model (e.g., `xtpoisson` in STATA), as there are repeated time observations within country-region pairs. This forces us to include the fixed effects 'manually', i.e. one by one. However, the very high number of fixed effects included in equation (1) causes serious numerical instability to the PPML estimator (and to other possible nonlinear alternatives such as a Negative Binomial Model, a Zero-Inflated Poisson Model or a Zero-Inflated Negative Binomial Model, see Cameron and Trivedi (2005) on the use of count models in case of high percentage of zeros). In particular, we tried to estimate an exponential version of equation (1) with the `poisson` command in STATA, which failed to reach convergence. Then we switched to the `ppml` command of Santos Silva and Tenreyro (2011). After rescaling the dependent variable and centering all the continuous variables, to ensure that the estimates exist `ppml` drops more than 1,000 fixed effects (mainly region-country fixed effects), and has convergence problems. This is likely to be caused by the inclusion of many dichotomous indicators (especially *trading-pair* fixed effects) which take value one for a very low proportion of the estimation sample. Similar problems arise when using a Zero-Inflated Poisson Model or Zero-Inflated Negative Binomial Model.

As mentioned above, the very high number of fixed effects prevent us from applying the Heckit estimator as in Helpman *et al.* (2008) or the threshold Tobit model of Eaton and Tamura (1994) to account for zero-trade observations, since both require estimating a Probit model which suffers from an incidental parameters problem. In conclusion, controlling for unobserved heterogeneity through *region-country* fixed effects, makes the log-log specification the sole feasible option among the many possible different alternatives.

Appendix 4: Large cities (Rome and Milan)

In section 4.1 we have put forward the idea that the provinces of the two main Italian trade hubs, Rome and Milan, may behave very differently from the rest of Italian provinces. In table 11 we present the OLS estimates for the provinces of Rome and Milan.

The table 11 shows that the coefficient of immigrants on export is negative, while that on imports is positive, but in both cases they are not precisely estimated. In the main text we have put forward the idea that one potential explanation for the absence of an effect of immigrants on exports for provinces including the two mega cities may be related to the fact that firms located in Rome and Milan are likely to be the most productive and efficient, thanks to the strong agglomeration economies present in these cities, and to export a large share of their production irrespective of the presence of ethnic networks.

Table 11: Regression Results for Rome and Milan (OLS) — province level

	(1) Exports	(2) Imports
$\ln(Y_{it-1}Y_{jt-1})$	0.721***	0.654
	(0.220)	(0.690)
$\ln(1 + IMM_{ijt-1})$	-0.084	0.428
	(0.356)	(0.485)
N. observations	2,604	2,604
R-squared	0.942	0.935

*, **, ***, significant at the 10%, 5% and 1% statistical level. Note. The dependent variables are $\ln(1 + export_{ijt})$ $\ln(1 + import_{ijt})$, i.e. export (import) flows of province i to (from) country j at time t. The regressions also control for $region \times country$, $country \times year$ and $region \times year$ fixed effects. Geodesic distance, the EU and EFTA dummy and the contiguity dummy are dropped owing to multicollinearity. Export and import flows cover the period 2003-2009. Standard errors are clustered at the province by (importer or exporter) country level.

Appendix 5: Country classifications

In section 4.4 countries are classified by income level according to the World Bank's classification:

High Income: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Japan, Republic of Korea, Luxembourg, Netherlands, New Zealand, Norway, Svalbard, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Liechtenstein, USA, United kingdom.

Upper Middle Income: Albania, Algeria, Antigua and Barbuda, Argentina, Azerbaijan, Belarus, Bosnia-Hercegowina, Botswana, Brazil, Bulgaria, Chile, China, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, Gabon, Grenada, Iran (Islamic Republic of), Jamaica, Jordan, Kazakhstan, Latvia, Lebanon, Libyan Arab Jamahiriya, Lithuania, Macedonia (the former Yugoslav Republic), Malaysia, Maldives, Mauritius, Mexico, Namibia, Palau, Panama, Peru, Romania, Russian Federation, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Seychelles, South Africa, Suriname, Thailand, Tunisia, Turkey, Uruguay, Venezuela.

Lower Middle Income: Angola, Armenia, Belize, Bhutan, Bolivia, Cameroon, Cape Verde, Congo, Cote d'Ivoire, Djibouti, East Timor, Egypt, El salvador, Fiji, Georgia, Ghana, Guatemala, Guyana, Honduras, India, Indonesia, Iraq, Kiribati, Lao People's Democratic Republic, Lesotho, Marshall Islands, Mauritania, Micronesia (Federated States of), Moldova (Republic of), Mongolia, Morocco, Nicaragua, Nigeria, Pakistan, Papua New Guinea, Paraguay, Philippines, Samoa, Sao Tome and Principe, Senegal, Solomon Islands, Sri lanka, Sudan, Swaziland, Syrian arab republic, Tonga, Turkmenistan, Tuvalu, Ukraine, Uzbekistan, Vanuatu, Vietnam, Yemen, Zambia.

Low Income: Afghanistan, Bangladesh, Benin, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Eritrea, Ethiopia, Gambia, Guinea, Guinea-bissau, Haiti, Kenya, Korea (Democratic People's Republic of), Kyrgyzstan, Liberia, Madagascar, Malawi, Mali, Mozambique, Myanmar, Nepal, Niger, Rwanda, Sierra Leone, Somalia, Tajikistan, Tanzania (United Republic of), Togo, Uganda, Zimbabwe.

In the same section the first and second waves of migration are identified as the 20 most important ethnic groups of foreign born residents, at the beginning of the period – year 2002 for the first wave – and at the end of the period – year 2009 for the second wave. The two classes are non overlapping, the second wave is identified excluding the ethnic groups of the first wave.

First Wave: Albania, Morocco, Romania, China, Philippines, Tunisia, Serbia and Montenegro, Senegal, India, Peru, Sri lanka, Macedonia, Egypt, Germany, Poland, Ghana, France, Pakistan, Nigeria, Bangladesh. Ordered by the share on overall foreign born residents.

Second Wave:United Kingdom, Brazil, Croatia, Bosnia-Hercegowina, Ecuador, USA, Spain, Ukraine, Algeria, Dominican Republic, Colombia, Russian Federation, Argentina, Cote d'ivoire, Cuba, Bulgaria, Turkey, Rep. of Moldova, Eritrea, Burkina Faso.

References

Aiello F. and Scoppa V. (2000) Uneven regional development in italy: Explaining differences in productivity levels., *Giornale degli Economisti e Annali di Economia*, 59, 2, 270–298.

Altonji J. and Card D. (1991) The effects of immigration on the labor market outcomes of less-skilled natives, in: *Immigration, Trade and the Labor Market*, Abowd J. and Freeman R., eds., University of Chicago Press, Chicago.

Anderson J.E. (2011) The gravity model, *Annual Review of Economics*, 3, 133–160.

Anderson J.E. and van Wincoop E. (2003) Gravity with gravitas: A solution to the border puzzle, *American Economic Review*, 93, 1, 170–192.

Artal-Tur A., Pallardó-López V.J. and Requena-Silvente F. (2012) The trade-enhancing effect of immigration networks: New evidence on the role of geographic proximity, *Economics Letters*, In Press.

Atkin D. (2010) Trade, tastes and nutrition in India, *Yale University Growth Center Discussion Papers*, 986.

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

Briant A., Combes P.P. and Lafourcade M. (2009) Product complexity, quality of institutions and the pro-trade effect of immigrants, *CEPR Discussion Papers*, 7192.

Briant A., Combes P.P. and Lafourcade M. (2010) Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?, *Journal of Urban Economics*, 67, 3, 287–302.

Bronnenberg B.J., Dube J.P.H. and Gentzkow M. (2012) The evolution of brand preferences: Evidence from consumer migration, *American Economic Review*, 102, 6, 2472–2508.

Burbidge J.B., Magee L. and Robb A.L. (1988) Alternative transformations to handle extreme values of the dependent variable, *Journal of the American Statistical Association*, 83, 123–127.

Cameron C. and Trivedi P. (2005) *Microeconometrics: Methods and Applications*, Cambridge University Press, Cambridge.

Castellani D. (2002) Export behavior and productivity growth: Evidence from Italian manufacturing firms, *Review of World Economics (Weltwirtschaftliches Archiv)*, 138, 605–628.

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

Coughlin C.C. and Wall H.J. (2011) Ethnic networks and trade: Intensive versus extensive margins, *Economics Letters*, 113, 1, 73–75.

De Benedictis L. and Taglioni D. (2011) The Gravity Model of International Trade, in: *The Trade Impact of European Union Preferential policies: An Analysis through Gravity Models*, De Benedictis L. and Salvatici L., eds., Springer, chapter 4, forthcoming.

Debaere P., Lee H. and Lee J. (2012) Language, ethnicity and intra-firm trade, *CEPR Working Papers*, , 9029.

Dunlevy J.A. (2006) The influence of corruption and language on the protrade effect of immigrants: Evidence from the American States, *Review of Economics and Statistics*, 88, 1, 182–186.

Dunlevy J.A. and Hutchinson W.K. (1999) The impact of immigration on American import trade in the late nineteenth and twentieth centuries, *Journal of Economic History*, 59, 4, 1043–1062.

Eaton J. and Tamura A. (1994) Bilateralism and regionalism in Japanese and U.S. trade and direct foreign investment patterns, *Journal of the Japanese and International Economies*, 8, 478–510.

Epstein G.S. and Gang I.N. (2004) Ethnic networks and international trade, *IZA Discussion Paper*, 1232.

Felbermayr G., Jung B. and Toubal F. (2010) Ethnic networks, information, and international trade: Revisiting the evidence, *Annales d'Économie et de Statistique*, 97/98, 41–70.

Genc M., Gheasi M., Nijkamp P. and Poot J. (2012) The impact of immigration on international trade: a meta-analysis, in: *Migration Impact Assessmemt*, Nijkamp P., Poot J. and Sahin M., eds., Edward Elgar Publishing Limited, Northhampton (MA), New Horizons in Regional Science, chapter 9, 301–337.

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

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, 302–16.

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, 1, 47–62.

Helliwell J.F. (1999) Language and trade, in: *Exploring the Economics of Language*, Breton A., ed., Canadian Heritage, Canada.

Helpman E., Melitz M. and Rubinstein Y. (2008) Estimating trade flows: Trading partners and trading volumes, *The Quarterly Journal of Economics*, 123, 2, 441–487.

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

ISTAT (2011) La popolazione straniera residente in Italia, Technical report, Statistiche in breve - Istituto Nazionale di Statistica, Rome, Italy.

Jargowsky P.A. (2005) The ecological fallacy, in: *The Encyclopedia of Social Measurement*, Kempf-Leonard K., ed., Academic Press, San Diego, California, volume 1, 715–722.

Liu X. (2009) GATT/WTO promotes trade strongly: Sample selection and model specification, *Review of International Economics*, 17, 428–446.

Mazzolari F. and Neumark D. (2012) The effects of immigration on product diversity, *Journal of Population Economics*, 25, 3, 1107–1137.

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.

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

Rauch J.E. (2001) Business and social networks in international trade, *Journal of Economic Literature*, 39, 4, 1177–1203.

Rauch J.E. and Trinidate V. (2002) Ethnic Chinese networks in international trade, *Review of Economics and Statistics*, 84, 1, 116–130.

Samuels M.L. (1993) Simpson's paradox and related phenomena, *Journal of the American Statistical Association*, 88, 421, 81–88.

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

Santos Silva J.M.C. and Tenreyro S. (2011) Poisson: Some convergence issues, *Stata Journal*, 207–212.

Staiger D. and Stock J. (1997) Instrumental variables regression with weak instruments, *Econometrica*, 65, 3, 557–586.

Vertovec S. (2006) The emergence of super-diversity in Britain, *Oxford University Centre on Migration, Policy and Society working papers*, , 25, http://www.compas.ox.ac.uk/fileadmin/files/pdfs/Stevendf.

Vertovec S. (2007) Super-diversity and its implications, *Ethnic and Racial Studies*, 30, 6, 1024–1054.

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

White R. (2007) Immigrant-trade links, transplanted home bias and network effects, *Applied Economics*, 39, 7, 839–852.

White R. and Tedesse B. (2007) Immigration policy, cultural pluralism and trade: Evidence from the White Australia policy, *Pacific Economic Review*, 142, 2, 489–509.