On the pro-trade effects of immigrants

Massimiliano Bratti[*]   Luca De Benedictis[†]   Gianluca Santoni[‡]

Abstract

In this paper we investigate the causal effect of immigration on trade flows, using Italian panel data covering very small geographical units (NUTS-3). Exploiting the very favorable setup offered by Italy's features — the very high number of countries of origin of immigrants ('super-diversity'), the high heterogeneity of social and economic characteristics of Italian provinces coupled with the absence of cultural (e.g. language) or historical (colonial ties) attractors for immigration —, controlling for a wide set of fixed effects, and dealing with the possible distortions generated by the inappropriate choice of the areal unit, we find a positive association between immigrants' stocks and both export and import flows, in line with the past literature. However, using instrumental variables based on migration enclaves, we show that immigrants have a positive and significant effect on imports only.

Keywords: Migration, Trade, Gravity model, Super-diversity, MAUP, Transplanted-home bias effect, Business and social networks effects.
JEL Classification: F10, F14, F22, R10.

[*]DEMM – University of Milan, CHILD (Turin), IZA (Bonn) and Ld'A (Milan) . E-mail: massimiliano.bratti@unimi.it

[†]Corresponding Author. DIEF - University of Macerata. E-mail: debene@unimc.it

[‡]Fondazione Manlio Masi. DE - University of Rome "Tor Vergata". E-mail: gianluca.santoni@uniroma2.it

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.[1] At the same time, more than 20% of the value of world production was sent from the country where it was produced to a foreign country.[2] This paper is about the link between these two 'facts,' and looks at the causal effect of immigrants on international trade flows.

The paper focuses on the Italian case during the 2000s. Italy offers some interesting peculiarities on many grounds. First, it 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, 2010). Second, 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. The very fine geographical disaggregation of the Italian longitudinal data on trade and migration allows us to adopt a reliable identification strategy for the effect of immigration on international trade. Third, the characteristics of the Italian data in terms of the large heterogeneity of the countries of origin of immigrants make the Italian case potentially relevant also for other countries as well. For all these

---

[1] In the OECD countries this share rise to 8.9% (Jean-Christophe Dumont and Widmaier, 2010). 31.4 million of immigrants were living in the US; 7.8 million in Germany; 5.6 million in France; 5.3 million in Canada; 2 million in Italy. Several non-OECD countries also had very large foreign-born populations. 11 million live in Russia; 6 million in India; 1.8 million in Israel. In relative terms, high shares of immigrants were recorded in several OECD countries in 2000 (in Luxembourg 37% of the population was foreign-born; in Australia 27%), but also among non-OECD countries (Singapore: 23%, Estonia: 22%, Belize: 21%, and Latvia: 21%). The highest share of immigrants in OECD countries was for Israel, which reported a 40% of immigrants in the population. Some countries, however, have a very low share of foreign-born in their population (below 1%), such as Indonesia, Sri Lanka, Cuba, Colombia, Laos, Peru, Mongolia, Bulgaria and Thailand. Among OECD countries, the lowest shares of immigrants are observed in Mexico (0.4%) and Japan (1.1%).

[2] In spite of the recent 2009 slump, the average 1990-2010 grow rate in international trade has been more than 5% on a yearly basis, substantially higher than the 3.5% average World GDP growth rate. The world average exports-GDP ratio (imports-GDP ratio) was 27.9% (28%) in 2010 according to the World Bank DataBank.

three peculiarities of the Italian case, we think that the results of our analysis can contribute to inform the policy debate on the interplay between immigration and international trade at a more general level.

To be more specific, the Italian immigration-and-trade case is characterized by what we may call a *bi-dimensional heterogeneity.* 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, Capo 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) established 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's wall and the entering of Poland, first, and Romania, afterwards, in the European Union. Italy is now 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 to 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 very well.

The second dimension of the Italian heterogeneity we are going to deal with comes from the profound diversity in the socio-economic characteristics of Italian regions, ranging from a rich and industrialized North-west, very well connected to the core of Europe, to a largely poor and underdeveloped South. This geographical divide is routed in the evolution of social and economic historical events in Italy. Its persistence is remarkable. Regional differences are evident in income per capita, unemployment rates, sectoral specialization, firms clustering, educational levels, crime, and — of course — migration flows. This profound spatial heterogeneity provides an ideal setting to investigate the effect of immigrants on geographical entities which experience very different levels of social and economic development.

In our empirical analysis, we turn to our advantage the minimal partic-

ipation of Italy to colonialism and the Italian bi-dimensional heterogeneity mentioned before. As emphasized by Briant *et al.* (2009), in *country*-level analyses of the effect of immigrants on trade flows (see Wagner *et al.* (2002) and Peri and Requena-Silvente (2010) for a review of the issue) 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 might be spoiled by the reverse causality inherent to the fact that immigrants generally migrate 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 US (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. 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 (Head *et al.* (2010), Helliwell (1999) and De Benedictis and Taglioni (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. Moreover, in spite of the 2000-2009 time-span which we consider is relatively short, the geographical coverage of our data is extremely accurate allowing us to consider 189 countries of origin of immigrants. This is about two and half times the number of countries considered by Peri and Requena-Silvente (2010) and more than five times the ones considered by Herander and Saavedra (2005), for instance. To the best of our knowledge, our dataset has 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 elasticities of trade to immigration.

The fine geographical detail of our data can be advantageous from an empirical stand point. In line with some recent contributions (Wagner *et al.*, 2002, Dunlevy, 2006, Bandyopadhyay *et al.*, 2008, Briant *et al.*, 2009, Peri and Requena-Silvente, 2010), we will test the relationship between trade and immigration over fine spatial units: using both regional and provincial Italian data. 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 US states (Dunlevy, 2006) and more than 100 times smaller than Canadian provinces Wagner *et al.* (2002). Since the size effect of the Modifiable Areal Unit Problem (MAUP) might be important, especially at large scales, (see Briant *et al.* (2010) on the issue) it is of relevance to run the analysis at the tiniest areal unit available, and to allow for a comparative analysis of the results obtained using different spatial units. This is even more important in our specific case, as the most popular explanation for the pro-trade effects of immigration (see section 2) is based on *network effects* and interactions and knowledge flows between natives and immigrants (Rauch, 1999, 2001), which may for instance reduce the fixed costs of exporting (or importing) for domestic firms. These interactions are likely to depend on the distance between individuals and on their location within a country (Herander and Saavedra, 2005), and accordingly the proper geographical unit to assess their existence are small areas such as NUTS-3, rather than countries or larger regions (NUTS-2). In this paper, we use both regional data (NUTS-2) — 20 regions of an average size of 14000 square km —, which have been more often used in past literature, and provincial data (NUTS-3), namely 107 Italian provinces of an average size of 2800 square km, less that half the average size of the 96 French *départements* in Briant *et al.* (2009).[3]

To further guarantee the minimal possible spurious correlation, we include a wide range of fixed effects. Country-year fixed effects allows to control for the common determinants of trade and immigration at the national level. At the same time, the variability in trade and immigration at the provincial level, after including *trading-pair* fixed effects to account for push and pull factors at the home-country and host-region levels, allows us to precisely isolate the pro-trade effect of immigrants. However, to address any potentially remaining bias, we also rely on an instrumental-variables (IVs, hereafter) ap-

---

[3] To be more precise, the mean area of Italian provinces is 2,816 square km with a coefficient of variation at 0.17, almost 57 times tinier than American states (162,176 km2, 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 much smaller and more regular size also 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 at 0.33 (when Corsica and overseas French regions are excluded), whereas the related figures for Spanish provinces are 10,118 km2 with a standard deviation at 0.47 (excluding Ceuta and Melilla).

proach and identify the causal effect of immigrants on exports and imports of Italian provinces using a widely employed instrument based on historical immigration enclaves.

We think that our analysis has at least four merits. First, the risk of a spurious correlation between trade and immigration is minimized due to the very fine geographical scale of our analysis. Second, the extensive country coverage of our dataset ensures that any sample selection bias stemming from the specific choice of the countries entering the analysis has been avoided. Third, to further rule out the possibility of an endogeneity bias that could inflate our coefficient of interest, we control for omitted common determinants including time-varying country-specific and *trading-pair* fixed effects in the regressions, and as in Briant *et al.* (2009), Peri and Requena-Silvente (2010) we make use of an IVs approach, where stocks the geographical distribution of immigrants' residence permits in 1995 (seven years before the lower limit of our time-span) and immigrants' flows at the nationwide level serve to compute an instrument (the imputed stock of immigrants). 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* (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 exports and imports trade flows.

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 the Appendix) and describes Italian super-diversity in immigration. 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 partly depart from the existing literature on the effect of immigrants on bilateral trade, allowing for geographic spillovers; in subsection 4.3 we describe the strategy used to tackle the endogeneity issue and report the causal effect of immigrants on trade resulting from two-stage least squares (2SLS) estimates. Section 5 concludes.

2 The pro-trade effects of immigrants

The international trade literature based on the estimate of a gravity equation (De Benedictis and Taglioni, 2011) — 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, Briant *et al.*, 2009, Peri and Requena-Silvente, 2010, Coughlin and Wall, 2011), for different samples, periods and estimation techniques have generally confirmed a strong effect of immigrants on trade.

In figure 1 we summarized the results of a sample of relevant contributions to the literature in terms of estimated elasticity of trade (imports, in black, and exports, in white) to immigrants. The vertical lines indicate the simple meta-modal elasticity derived from Genc *et al.* (2011), which is 0.12 for exports and 0.15 for imports. As it is evident the estimates show a high degree of variability. Before 2002 the literature was dominated by cross-country studies. Wagner *et al.* (2002) set the standard in the subsequent literature, 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 information on trade and immigration at the sub-national level (for Canada, in their 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 the one of exports, and that both are positive and generally significantly different from zero. Why?

The mechanisms at the basis of the common explanations of what usually drives 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 have to 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 orga-

Figure 1: Summary of the literature. Estimated elasticity of trade to immigrants: Imports (black) Exports (white)
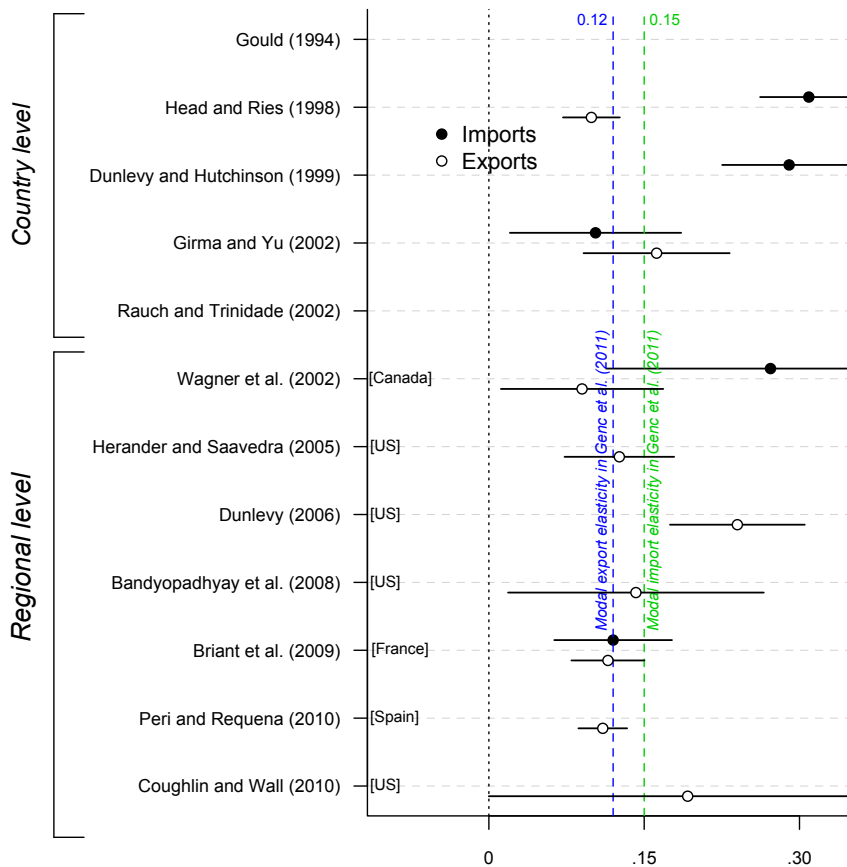
Note: The figure plots data obtained from several contributions to the literature on the effect of immigration on trade. Black dots indicate the elasticity of imports to immigrants, white dots the one of exports, segments indicate 95% confidence intervals. The horizontal axis is trimmed for visual purposes. The complete list of papers is included in the references.

nized markets in matching international demand and supply. Several studies have explored the role of ethnic networks in international trade since Rauch (2001).[4]

A further point associated with this first explanation is also 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) We may call this explanation the *business and social network effect* of immigrants on trade.

The second, less explored explanation is that immigrants are characterized by different habits in consumption with respect to natives, and they may slowly modify their original home-biased demand after settling in the host-country. 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 the one of exports, this was interpreted by deductive reasoning as supporting the idea that there should be something that makes the two elasticities to be different, and this 'something' was attributed to a persistent difference in tastes between immigrants and natives.

Recently, some more clean evidence of the relevance of the transplanted home-bias effect has been put forward by Bronnenberg *et al.* (2010), Atkin (2010) and Mazzolari and Neumark (2010). Bronnenberg *et al.* (2010) 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

---

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

9

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' than 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 the one consumed in the home-country is also found in Mazzolari and Neumark (2010), 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 will 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

3.1   Italy as super-diversity

Our data are obtained from mainly two publicly available datasets from the Italian National Statistical Institute (ISTAT). Trade flow data refer to the value of imports and exports of 107 Italian provinces (NUTS3) concerning 210 country trading partners around the world, over the period 2002-2009.[5]

---

[5] More precisely, we will consider 103 provinces until 2006 and 107 afterwards. The number of Italian provinces has 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 will not consider these latter changes in the number of Italian provinces.

The data are measured such that exports and imports are associated with the province of shipment, i.e. the province where the custom transaction was registered.[6] The original values, in Euros, have been converted in US 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. 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 immigrants by country of origin (home-country) and province or region of destination (in the host-country, which is in our case Italy), defining immigrants as residents born abroad with a foreign nationality.

Table 1 shows the top 20 countries of origin of immigrants in 2009. The top five countries by the number of foreign-born population are Romania, Albania, Morocco, China and Ukraine, accounting for about 50 percent of the total foreign 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); which experienced also the highest growth rate over the period. The change in the ranking between 2002 and 2009 is reported in figure 2 which shows some 'big movers'. Moldova and Ukraine, for instance, gain 32 and 23 positions, respectively, while Senegal looses 9 positions. An interesting feature of the immigration pattern in Italy is the uneven distribution of immigrants across Italian provinces. Figure 3 shows the map of Italy where provinces are colored according to the share of foreign-born population in the total population, with 'darker' provinces having a higher share of immigrants. While in 2002 none of the 103 provinces registered a share higher than 10 percent, in 2009, 23 provinces had over 10 percent of foreign born residents, mainly in the Center and the North of the country. The picture also reveals some spatial clustering for immigrants: provinces richer in immigrants are more likely to be close to each other. We will address the potential consequences of this issue for our analysis in Section 4.2.

---

[6] The information of Extra-EU transactions are based on the *Documento Amministrativo Unico* (DAU) which is registered for each commercial transaction, for the intra-EU exchanges the custom system has been replaced, since 1993, by the Intrastat standard.

Table 1: Immigrants by country of origin

Ranking in 2009	Country of origin	Number of immigrants in 2009	% of total immigrants in 2009	Annual growth rate, 2002/2009 (%)	Ranking in 2002
(1)	Romania	887763	20.96	40.45	(3)
(2)	Albania	466684	11.02	11.76	(1)
(3)	Morocco	431529	10.19	10.51	(2)
(4)	China	188352	4.45	15.51	(4)
(5)	Ukraine	174129	4.11	68.99	(28)
(6)	Philippines	123584	2.92	9.67	(5)
(7)	India	105863	2.50	16.99	(9)
(8)	Poland	105608	2.49	20.04	(15)
(9)	Moldova Rep.	105600	2.49	60.20	(40)
(10)	Tunisia	103678	2.45	8.33	(6)
(11)	Macedonia	92847	2.19	16.25	(12)
(12)	Peru	87747	2.07	14.60	(10)
(13)	Ecuador	85940	2.03	32.67	(25)
(14)	Egypt	82064	1.94	13.82	(13)
(15)	Sri Lanka	75343	1.78	11.99	(11)
(16)	Bangladesh	73965	1.75	20.27	(20)
(17)	Senegal	72618	1.71	10.24	(8)
(18)	Pakistan	64859	1.53	16.72	(18)
(19)	Serbia	57877	1.37	1.19	(7)
(20)	Nigeria	48674	1.15	12.97	(19)
	Top 20 countries	3,434,724	81.1	20.66	
	TOTAL	4,223,154	100	14.9	

Source: ISTAT

12

Figure 2: Ranking of immigrants by country of origin
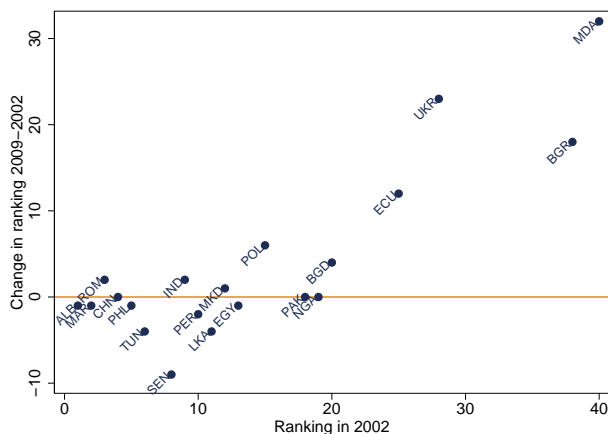
Even if the distribution across provinces of foreign residents reveals a relative concentration in Northern Italy, the number of provinces with zero immigrants from a particular country of origin is rather small. This is an instance of what we called Italian 'super-diversity'. Table 2 reports the mean number of nationalities registered in each province at the beginning and at the end of the period under study. Looking at the summary statistics, the mean value of nationalities found in a province is around 111 in year 2002, and about 125 in 2009. Provinces with less coverage of nationalities are concentrated, as we may expect looking at Figure 3, in Southern provinces.

If we look at the distribution of immigrants from the perspective of each ethnic community, immigrants from the same country of origin locate, 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 immigrant community is just 1), others are spread all over Italy (the observed maximum always hits the theoretical one). Focusing on the twenty most representative nationalities we register huge differences in their geographical distribution. The most concentrated groups are from Egypt, Ecuador and the Philippines. The distribution exhibit in 2009 a coefficient of variation[7] from 40% to 80% higher than the

---

[7] The coefficient of variation refers to the distribution of the province's share of the

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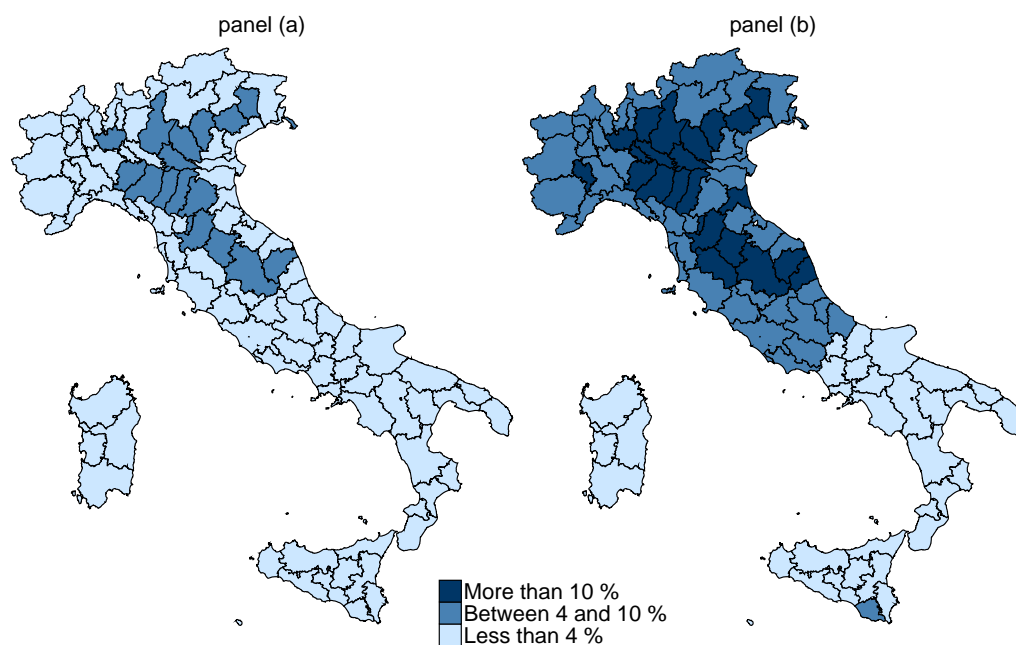
panel (a)   panel (b)

More than 10 %
Between 4 and 10 %
Less than 4 %

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

	Mean	Std. Dev	Min	P25	Median	P75	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: The total number of Italian provinces is 103 (107 from 2006) while the total number of foreign nationalities is 189.

14

median value. On average, in the first five provinces of residence are located around 70% of immigrants. The degree of concentration is relatively high with respect to Albania, Morocco and Tunisia, the most evenly distributed nationalities, as in this case the first five locations encompass 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, but despite that the associated coefficient of variation is around one third the value for Egypt.

Table 3: R^2 for models including different 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: 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 preserve the zeros in the estimation).

In addition to the specific information contained in Table 2, we can observe that a significant portion of the variation comes from the $province \times country$ dimension of the data. It is therefore worth exploring this characteristic in a more explicit way, in relation with the variables we are going to focus on in the analysis. Therefore, in Table 3 we report the R^2 of different linear regressions having the dependent variable (ln)exports, (ln)imports and (ln)immigrants regressed on a different set of dummies. Including the dummy $province$, $country$ and $time$ separately, already allows to take into account between 73% and 82% of the variability in the data; if instead we

---

total number of foreign residents by nationality.

Table 4: Within-country, within-province, within-year correlations

	Exports	Imports	Distance	Immigrants
Exports	1.0000			
Imports	0.1248	1.0000		
Distance	-0.0206	-0.0641	1.0000	
Immigrants	0.0753	0.1166	-0.0747	1.0000

Note: This table shows the correlations between the residuals of the different variables from a regression controlling for province, country and year fixed effects. In all cases the dependent variable is $ln(1 + x)$, were x are exports, imports or immigrants (to preserve the zeros in the estimation). All correlations are significant at 1% level.

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 categorical effect, the R^2 decreases by 5%. The inclusion of the full set of fixed effects captures almost 90% of the variation in both exports and imports flows and over 98% of immigration stocks, our main variable of interest, when using *province × country* fixed effects in panel (a). By contrast, the R^2 is between 73% and 84% in the case *region × 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 × country* dummies, or even more demanding specifications also including time-varying versions country and province fixed effects (as reported in Table 3), clashes with the very little variation left in the data. In the empirical analysis that follows we therefore opt for a *region × country* specification of the *trading-pair* dummies.

Another fundamental variable in the analysis is distance, that we measure as the geodesic distance between the centroid of the Italian local unit area and the capital city of the country of origin of immigrants. The within-country, within-province, and within-time correlations between exports, imports, immigration and distance (all in natural logs) are included in Table 4. Formally, what it is reported in Table 4 is the correlation between the residuals of the regression of each dependent variable (exports, imports, immigrants, distance) on country-specific, province-specific and year-specific dummies. As expected from the literature on trade costs (Anderson and van Wincoop, 2004) and the gravity equation (De Benedictis and Taglioni, 2011), distance is negatively correlated with both exports and imports, the

correlation being stronger for imports, and with immigration. Exports and imports are positively correlated, while, consistently with our priors derived from the literature described in section 2, immigration is significantly and positively correlated with both exports and imports, but the correlation is stronger with the latter. We will see if this link survives when estimating the gravity equations, and if it is indicative of a causal effect.

Table 5: Summary Statistics (2002-2009)

	Mean	Std. Dev	Min	P25	Median	P75	Max
Strictly positive exports 102957/157718							
Exports	20985.01	109576.50	0	68.41	716.63	6805.21	4968820
Distance	5385.64	3912.44	73.86	1841.07	4583.35	8288.85	19029.86
Immigrants	207.56	1389.90	0	0	7	48	139821
All exports 157718							
Exports	13698.84	89094.68	0	0	58.32	2036.22	4968820
Distance	6471.09	4334.02	73.86	2945.17	5879.97	8921.47	19029.86
Immigrants	136.44	1127.37	0	0	1	18	139821
Strictly positive imports 74259/ 157718							
Imports	29018.07	243106.7	0	76.17	863.10	7737.83	19228701
Distance	5100.91	3980.44	73.86	1515.41	4277.60	8559.80	19029.86
Immigrants	280.12	1628.79	0	2	14	83	139821
All imports 157718							
Imports	13662.69	167440.40	0	0	0	657.71	19228701
Distance	6471.09	4334.02	73.86	2945.17	5879.97	8921.47	19029.86
Immigrants	136.44	1127.37	0	0	1	18	139821

Note: Exports and imports are in thousands of euros, as reported on custom data, immigrants are number of foreign-born Italian residents. Distance is the average number of kilometers between provinces' centroids and foreign capital cities.

Table 5 reports the summary statistics for trade flows, as well as the average immigrant stock and distance of imports and exports flows by province. The average distance of trade flows is quite high, over 6,400 km, but if we restrict the analysis to the strictly positive flows the average distance shrinks to 5,100 km for imports and 5,385 km for exports. When considering province-time observations with strictly positive exports or imports,

the average number of foreign-born residents is 208 and 280, respectively, but with a significant variation across provinces and nationalities as shown by the standard deviation. The median value of immigrants associated to a bilateral trade between a country and an Italian province is 7 in the case of (strictly positive) exports, and 14 in the case of (strictly positive) imports. Interestingly enough, these figures fall when including province-time observations with zero exports or imports, which is another sign that immigrants and trade are positively associated.

To further explore the preliminary evidence of the correlation between trade and immigration flows, we report in Figure 4 the kernel density of the log value of imports and exports for Italian provinces characterized by a positive value of immigrants against those provinces without any foreign-born population. As we can see, the two distributions are quite different: provinces with a foreign-born population present on their territories tend to trade more internationally.

4 Empirical results

This section represents the core of our empirical analysis. 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 partly depart from the existing literature by estimating an 'augmented' gravity equation, which allows for geographic spillovers, retaining the exogeneity assumption. Last but not least, in subsection 4.3 we allow for potential endogeneity of immigration, and report two-stage least squares (2SLS) estimates.

4.1 Ordinary least squares

We use the following Anderson and van Wincoop (2003) specification[8] as our *benchmark* model for the gravity equation:

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

---

[8] 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) on the formulation of a theory-based gravity equation.

Figure 4: Kernel distribution of trade flows for provinces with and without immigrants
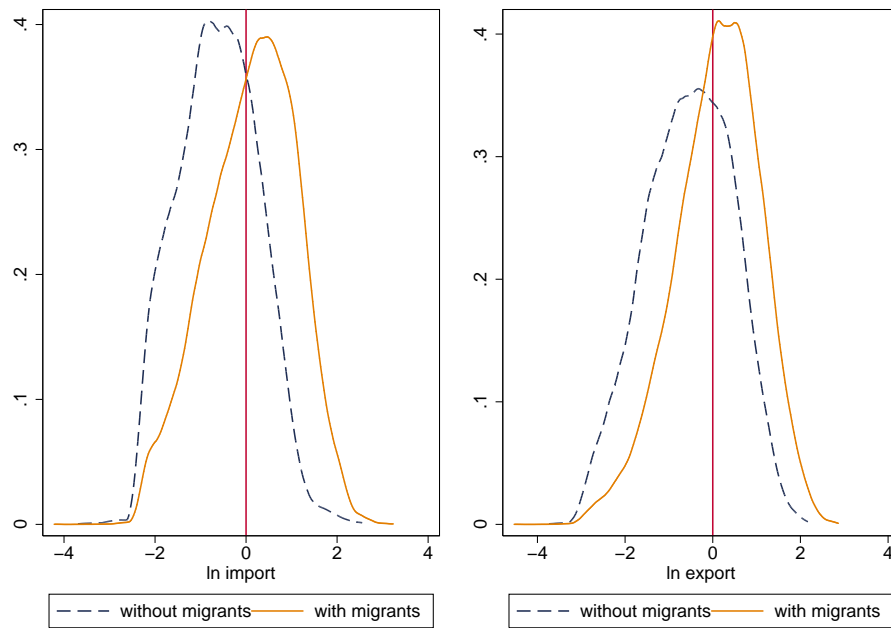
Note: The figure reports the standardized values - mean 0 and standard deviation 1 - of both imports and exports.

where i, r, j and t are the subscripts for the Italian province (NUTS3), the region where the province is located (NUTS2), the foreign country, and time, respectively. $\boldsymbol{\delta}_{rj}$ 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 $dist_{ij}$, which is the great-circle distance between province i and country j. Due to the time lag of the independent variables, trade flows cover the 2003-2009 period and immigration stocks the 2002-2008 period. ϵ_{ijt} is a classical error term. Note that GDPs and the immigrant stock are lagged one year so as they are predetermined with respect to trade.

In order to use a log-log expression 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), and add one to both bilateral trade flows and the stock of immigrants.[9] Moreover, in spite of the suggestion by Silva and Tenreyro (2006) in favor of a general use of a Poisson pseudo-maximum-likelihood (PPML) estimator, in our case, 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), and the log-linear specification turns out to be the more reliable alternative among the many possible different choices.[10] It is worth

---

[9] Authors treated zero trade observations in different ways. Just to take a few examples, Bandyopadhyay *et al.* (2008) restrict the analysis to the observations with positive trade, while as we said Dunlevy (2006), Peri and Requena-Silvente (2010), Coughlin and Wall (2011) add one to zero trade observations. Although the log-log expression of the gravity model has been recently subject to some criticism by Silva and Tenreyro (2006), the debate on which should be 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 very specific point of the gravity literature).

[10] To make an account of our attempts, 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 Silva and Tenreyro. After rescaling the dependent variable and centering all the continuos variables, the estimator apparently achieved convergence but did not return the standard errors presumably due to the high singular-

noting that we are not including in the gravity equation *province × country* and *province × year* fixed effects because, from what we reported in Table 3, they would absorb almost all variation in both trade and immigration, leaving virtually nothing else to be explained. By contrast, we use larger geographical units, that is administrative areas corresponding to NUTS2, to define both *region × country* fixed effects and *region × year* home areal units time-varying fixed effects. This enables us to still 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.[11] A similar approach is used by Wagner *et al.* (2002), which, in the case of Canadian provinces, included fixed effects at the national level. In particular, they 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).

It is important to stress that specification (1) is to the best of our knowledge the most comprehensive one 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). Many subsequent contributions to the literature just included *countries × time* (i.e. *trading-pair*) fixed effects (Dunlevy, 2006, Wagner *et al.*, 2002, Briant *et al.*, 2009). Some recent papers use *trading-pair* and *year* fixed effects (Bandyopadhyay *et al.*, 2008) or *trading-pair* and *countries × year* fixed effects (Peri and Requena-Silvente, 2010, Coughlin and Wall, 2011) but *region × year* fixed effects are generally omitted. Yet, the inclusion of the complete set of fixed effects is likely to attenuate the potential endogeneity of migration flows. Indeed, *trading-pair* fixed effects are likely to account for factors that may spur trade between an Italian province and a specific country such as cultural proximity or historical ties (e.g., past Italian emigration from a given province towards a certain country,) *country × 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 af-

---

ity of the variance matrix. This problem is likely to be caused by the existence of many dummy variables (especially *trading-pair* fixed effects) which take value one for a very low proportion of the estimation sample. The same happened when using a a Zero-Inflated Poisson Model or Zero-Inflated Negative Binomial Model.

[11] In particular, we are exploiting between-province within-region variation over-time.

fect both trade and migration flows towards many Italian provinces, and *region × 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 × country* level.

Table 6: Gravity equation for exports (OLS) — province level

	(1)	(2)	(3)	(4)	(5)
$\ln(Y_{it-1}Y_{jt-1})$	1.638***	2.108***	2.101***	2.174***	2.167***
	(0.011)	(0.027)	(0.027)	(0.029)	(0.029)
$\ln(1+IMM_{ijt-1})$	0.506***	0.084***	0.083***	0.059***	0.058***
	(0.013)	(0.019)	(0.019)	(0.020)	(0.020)
EU_{jt}, $EFTA_{jt}$	0.108	-0.282***	-0.280***		
	(0.072)	(0.053)	(0.052)		
$\ln(dist_{ij})$	-1.270***	-1.832***	-1.829***	-1.876***	-1.873***
	(0.032)	(0.386)	(0.384)	(0.389)	(0.388)
$contiguity_{ij}$	-1.105***	0.116	0.117	0.141	0.142
	(0.411)	(0.306)	(0.305)	(0.318)	(0.316)
year FE	yes	yes			
region × year FE			yes		yes
country × year FE				yes	yes
trading-pair FE		yes	yes	yes	yes
N. observations	135,586	135,586	135,586	135,586	135,586
R-squared	0.61	0.79	0.79	0.80	0.80
N. clusters	20,009	20,009	20,009	20,009	20,009

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

Note. The dependent variable is $\ln(1+EXPORT_{ijt-1})$, i.e. export flows of province i from country j at time t. *Trading-pair* fixed effects (FE) are defined at the *region × country* level. Export flows cover the period 2003-2009. Standard errors are clustered at the province by (importer or exporter) country level.

Before commenting on the results obtained with the *benchmark* specification (1), we report specifications including fewer fixed effects to stress the importance of controlling for unobserved heterogeneity. In these specifications

Table 7: Gravity equation for imports (OLS) — province level

	(1)	(2)	(3)	(4)	(5)
$\ln(Y_{it-1}Y_{jt-1})$	1.649***	2.017***	2.012***	2.105***	2.100***
	(0.013)	(0.034)	(0.034)	(0.036)	(0.036)
$\ln(1 + IMM_{ijt-1})$	0.796***	0.363***	0.362***	0.344***	0.344***
	(0.016)	(0.026)	(0.026)	(0.027)	(0.027)
EU_{jt}, $EFTA_{jt}$	1.929***	-0.367***	-0.366***		
	(0.088)	(0.086)	(0.086)		
$\ln(dist_{ij})$	-0.581***	-2.837***	-2.834***	-2.899***	-2.897***
	(0.039)	(0.592)	(0.592)	(0.595)	(0.595)
$contiguity_{ij}$	0.236	-0.301	-0.301	-0.284	-0.283
	(0.555)	(0.339)	(0.338)	(0.357)	(0.356)
$year$ FE	yes	yes			
$region \times year$ FE			yes		yes
$country \times year$ FE				yes	yes
$trading\text{-}pair$ FE		yes	yes	yes	yes
N. observations	135,586	135,586	135,586	135,586	135,586
R-squared	0.63	0.76	0.76	0.77	0.77
N. clusters	20,009	20,009	20,009	20,009	20,009

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

Note. The dependent variable is $\ln(1 + IMPORT_{ijt-1})$, i.e. import flows of province i from country j at time t. *Trading-pair* fixed effects (FE) are defined at the $region \times country$ level. Import flows cover the period 2003-2009. Standard errors are clustered at the province by (importer or exporter) country level.

we also included a dummy for EU's or EFTA's participating countries, and a dummy for contiguity between province i and country j, whenever they are not absorbed by the fixed effects included (or not dropped due to collinearity).[12] In column (1) of Table 6, we report the results of a specification only including time dummies, *year*: the estimated elasticity of exports with respect to immigration is very high, at 0.503.[13] Meaning that a 1% increase in the stock of immigrants settled in a specific Italian province rises exports from that province to the country of origin of immigrants by 0.5%. 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 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), which include *trading-pair* and *country* $\times$ *year* fixed effects but exclude *region* $\times$ *year* fixed effects. In this case the drop in the elasticity is smaller than in column (3). Shifting to our *benchmark* specification has practically no additional effect on the estimated elasticity in column (5), which becomes 0.058. Hence, it seems that the combination of *trading-pair* and *country* $\times$ *year* fixed effects is likely to capture 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 all specifications the coefficient on immigration is statistically significant at least at the 1% level. When we consider imports, we find a similar fall in the estimated elasticities by progressively adopting more complex specifications in terms of fixed effects. The elasticity of imports with respect to immigrants is 0.798 in column (1), falls to 0.486 when adding *trading-pair* fixed effects in column (2). Also in this case, comparison of column (2) with the following columns shows that, like for exports, most unobserved heterogeneity is captured by the *trading-pair* fixed effects, and the estimated elasticity turns out to be little sensitive to the inclusion of the remaining fixed effects. In the *benchmark* model in column (5) the elasticity of imports with respect to immigration is 0.344. Also in this case, like for

---

[12] We included these regressors to take into account possible nonlinearities in distance and in analogy to Peri and Requena-Silvente (2010), which is the closest contribution to ours.

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

exports, the coefficient on immigration is always statistically significant at least at the 1% level.

Thus, as predicted by the economic theory and found in most of the previous literature (section 2), our estimate of the elasticity of imports with respect to immigrant stocks is much larger than the one 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.

As we said, a unique feature of our dataset is that it gives us the opportunity to assess the impact of the size of the areal units chosen as the base for the analysis of the trade elasticities to immigration on the magnitude of these elasticities. In particular, both trade and immigrant stocks are available at a very disaggregated geographical level in our data, i.e. Italian provinces. For the sake of comparison, we also report here the estimates of a gravity equation using region-level data. In this case the size of the unit of analysis is similar to the one adopted in Peri and Requena-Silvente (2010).

We use the following specification for the gravity equation:[14]

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

where r, j and t are the subscripts for the the Italian regions (NUTS2), the foreign country, and time, respectively. $\boldsymbol{\delta}_{ij}$ are *region × country* (*trading-pair*) fixed effects, $\boldsymbol{\theta}_{jt}$ are *country × year* fixed effects, and $\boldsymbol{\phi}_{rt}$ *region × 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 from country j located in region r, Y_{rt-1}, Y_{jt-1} region and foreign countries GDPs at time $t-1$, and ϵ_{rjt} is a classical error term. As the main aim of the region-level analysis is to provide a term of comparison to assess the existence of a potential MAUP bias, 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. In these specifications we also include a dummy for EU's or EFTA's participating countries, and a dummy for contiguity between region r and country j, whenever they are not absorbed by the fixed effects included. The OLS estimates are reported in tables 8 and 9. As it is

---

[14] Note that because *trading-pair* fixed effects are now defined at the *region × country* level ($\boldsymbol{\delta}_{rj}$), distance ($\ln(dist_{rj})$) drops from the regression.

clear from both tables, the initial large positive elasticities for exports and imports estimated in column (1) tend to fall adding fixed effects . In the last column where *trading-pair*, *region × year* and *country × year* fixed effects are included, these elasticities even turn out to be negative, although they are never statistically significant.

Table 8: Gravity equation for exports (OLS) — regional level

	(1)	(2)	(3)	(4)	(5)
$\ln(Y_{rt-1}Y_{jt-1})$	1.528***	0.800***	0.853***	-0.332	0.909**
	(0.023)	(0.145)	(0.144)	(0.833)	(0.460)
$\ln(1 + IMM_{rjt-1})$	0.359***	0.028	0.016	-0.029	-0.042
	(0.022)	(0.064)	(0.065)	(0.078)	(0.078)
EU_{jt}, $EFTA_{jt}$	0.851***	-0.039	-0.041		
	(0.136)	(0.067)	(0.069)		
$\ln(dist_{rj})$	-1.125***				
	(0.054)				
$contiguity_{rj}$	-1.226***				
	(0.214)				
year FE	yes	yes			
region × year FE			yes		yes
country × year FE				yes	yes
trading-pair FE		yes	yes	yes	yes
N. observations	25,853	25,853	25,853	25,853	25,853
R squared	0.66	0.89	0.90	0.90	0.91
N. clusters	3,740	3,740	3,740	3,740	3,740

*, **, *** statistically significant at the 10, 5 and 1 percent level, respectively.
Note. The dependent variable is $\ln(1 + EXPORT_{rjt-1})$, i.e. export flows of region r from country j at time t. *Trading-pair* fixed effects (FE) are defined at the *region × country* level. Export flows cover the period 2003-2009. Standard errors are clustered at the region by (importer or exporter) country level.

It is interesting to compare the *benchmark* estimates in the province-level analysis (column (5) of tables 6 and 7) and the region-level analysis (column (5) of tables 8 and 9). As it is clear from the comparison, the use of a finer geographic unit of analysis increases the estimated elasticity, which changes sign and becomes statistically significant. One potential explanation is that considering regions instead of provinces may produce an *aggregation bias*. Think of two regions A and B which are equally populated both in terms of

Table 9: Gravity equation for imports (OLS) — regional level

	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(Y_{rt-1}Y_{jt-1})$	1.742***	0.800***	0.694***	-0.159	0.987*	
	(0.026)	(0.145)	(0.195)	(0.955)	(0.556)	
$\ln(1+IMM_{rjt-1})$	0.636***	0.028	-0.129	-0.084	-0.096	
	(0.028)	(0.064)	(0.086)	(0.097)	(0.098)	
EU_{jt}, $EFTA_{jt}$	0.869***	-0.039	0.245			
	(0.184)	(0.067)	(0.153)			
$\ln(dist_{rj})$	-0.780***					
	(0.069)					
contiguity$_{rj}$	-0.068					
	(0.281)					
year FE	yes	yes				
$region \times year$ FE			yes		yes	
$country \times year$ FE				yes	yes	
$trading\text{-}pair$ FE		yes	yes	yes	yes	
N. observations	25,853	25,853	25,853	25,853	25,853	
R squared	0.67	0.89	0.89	0.90	0.90	
N. clusters	3,740	3,740	3,740	3,740	3,740	

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

Note. The dependent variable is $\ln(1+IMPORT_{rjt-1})$, i.e. import flows of region r from country j at time t. *Trading-pair* fixed effects (FE) are defined at the *region $\times$ country* level. Import flows cover the period 2003-2009. Standard errors are clustered at the region by (importer or exporter) country level.

natives and immigrants, with two provinces each, 1 and 2. Let us assume that the main source of the pro-trade effects of immigration are contacts between natives and immigrants. Immigrants, for instance, may provide natives with useful information on how to do business with their origin countries, or help enforce contracts (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 in province B_2. If we use provinces as the unit of analysis, the magnitude of the trade-creating effect of immigrants will be estimated correctly and be 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 will create a sort of measurement error in the migration variable. In particular, it will appear from the data as if in both regions the amount of interactions between natives and immigrants were the same (as both regions have the same number of migrants) causing an under-estimation of the elasticity of trade with respect to the stock of immigrants.

4.2 Geographic spillovers from immigration on trade

Most of the papers existing in the literature, and the specification (1), do not allow for geographic spillovers from migration on trade. However taking account of potential spillovers is important. Province i may, for instance, have a low stock of immigrants from a specific country, but it 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-rich provinces. Moreover, if there are inter-province spillovers, and the provinces' immigrant stocks are spatially correlated, then the coefficient on province's own immigration may also pick up the effect of immigrants located in other provinces, that is it may produce an *omitted variables bias*.[15]

To the best of our knowledge, ethnic spillovers on trade have been considered only by Herander and Saavedra (2005), which 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

---

[15] It may important to stress that in our case the spatial correlation concerns an independent variable.

Table 10: Gravity equations for exports and imports with geographic spillovers (OLS) — province level

	Exports (1)	Imports (2)
$\ln(Y_{it-1}Y_{jt-1})$	2.168***	2.101***
	(0.029)	(0.036)
$\ln(1 + IMM_{ijt-1})$	0.055***	0.340***
	(0.020)	(0.027)
$\ln(1 + IMMsp_{ijt-1})$	0.013**	0.015
	(0.007)	(0.009)
$\ln(dist_{ij})$	-1.911***	-2.887***
	(0.381)	(0.584)
N. observations	135586	135586
R-squared	0.80	0.77
N. clusters	20,009	20,009

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

Note. The dependent variables are $\ln(1 + EXPORT_{ijt-1})$ and $\ln(1 + IMPORT_{ijt-1})$, i.e. exports and imports 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. Standard errors clustered at the $province \times country$ level are reported in parentheses. Trade flows cover the period 2003-2009.

with the importance of proximity to immigrant networks. In our context, the idea is that not only the stock of immigrants from country j in province i but also the stock of immigrants of the same nationality in other provinces may affect province i's trade. Here, to investigate geographic spillovers we follow Andersson *et al.* (2004) and Andersson *et al.* (2009) and build a 'gravity variable' ($MIGsp_{ijt}$) by aggregating the stocks of immigrants from nationality j in the other Italian provinces (excluding province i) using the inverse distance between provinces (d_{iz}) as weights, that is

$$IMMsp_{ijt} = \sum_{z=1,z\neq i}^{I} \frac{IMM_{ijt}}{d_{iz}} \tag{3}$$

where I is the total number of provinces. Hence, unlike Herander and Saavedra (2005) we weight the stocks of immigrants with the distances between province i and all the other Italian provinces. We think that this is an intuitive way of capturing spillover effects if the latter depend on geographic proximity.

The 'augmented' gravity equation becomes

$$\ln(1 + X_{ijt}) = \boldsymbol{\delta}_{rj} + \boldsymbol{\theta}_{jt} + \boldsymbol{\phi}_{rt} + \ln(Y_{it-1}\,Y_{jt-1}) + \alpha_0\,\ln(1 + IMM_{ijt-1}) +$$
$$+ \alpha_1 \ln(1 + IMMsp_{ijt-1}) + \epsilon_{ijt}. \tag{4}$$

where another difference with respect to Herander and Saavedra (2005) — who used a cross-sectional specification— is that we include a wide set of fixed effects.

We only estimated the *benchmark* specification for both exports and imports, which is reported in Table 10. In both cases, the elasticity of imports and exports with respect to the stock of immigrants in the province is very similar to that obtained from the specification excluding the spillover variable, suggesting the absence of an omitted variables bias. The spillover variable turns out to be statistically significant only for exports, with an elasticity much smaller than the one on the province's immigration. This is broadly consistent with the findings of Herander and Saavedra (2005) of the importance of geographic proximity in ethnic networks.

4.3   Endogeneity and two-stage least squares

The OLS results in the previous subsection 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 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) which 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.[16] An attempt to address endogeneity of migrations 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 (Card, 2001), which we also follow in our paper. The presence of a community of immigrants from a given country in a certain 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 (see also Peri and Requena-Silvente, 2010). 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

---

[16] 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.

2006 and 2009.[17] 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 N_{ijt} the number of immigrants from country j located in province i at time t, and N_{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

$$\theta_{ijt} = \frac{N_{ijt}}{N_{jt}}. \tag{5}$$

After considering the lagged distribution of immigrants by nationality across provinces θ_{ij95} and having defined N_{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

$$\hat{N}_{ijt} = \theta_{ij95}N_{j0} + \theta_{ij95}\sum_{z=0}^{t}F_{jz} = \theta_{ij95}(N_{j0} + \sum_{z=0}^{t}F_{jz}) = \theta_{ij95}N_{jt} \tag{6}$$

where F_{jz} is the total net inflow of immigrants from country j in Italy at time z. The instrument is then given by the product of two terms, the first (θ_{ij95}) exhibits trading-pair variation, and the second (N_{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 like in our paper the immigrant stock is measured in logarithms. After taking logs, equation (6) becomes

$$ln(\hat{N}_{ijt}) = ln(\theta_{ij95}) + ln(N_{jt}) \tag{7}$$

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 × country* and *country × year* fixed effects, they will

---

[17] As we said, 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.

completely absorb the instrument's variation. Adding one to the stock of migrants and taking logs will make no longer possible to write the imputed stock of immigrants as a linear function of $ln(\theta_{ij95})$ and $ln(N_{jt})$, that is

$$ln(1 + \hat{N}_{ijt}) = ln(1 + \theta_{ij95}N_{jt}) \neq ln(\theta_{ij95}) + ln(N_{jt}) \qquad (8)$$

and the variation in the log of the modified imputed stock of immigrants (i.e., the imputed stock after adding one) is not completely absorbed now by *trading-pair* and *country×year* fixed effects. However, in the 2SLS estimates identification will stem only from the nonlinearity of $ln(1+\hat{N}_{ijt})$ in these fixed effects, which may be too weak in many cases (see table 3 where only few residual variation remains after controlling for *trading-pair* fixed effects at the *country × province* level).[18]

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 × country* level.

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 (6) should be 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 (N_{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 (θ_{ij95}), conditional on *trading-pair* fixed effects, *region × year* and *country × 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

---

[18] 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 × country* and *country × year* fixed effects. They add one to both trade and the immigrant stock to retain observations with zeros in their logarithmic specification, but on top of that in the 2SLS specification, which uses the imputed stock of immigrants as the excluded instrument, they omit *trading-pair* dummies from the first stage.

local opportunities for trade growth with their home countries. However, controlling for *trading-pair* (*region × country*) fixed effects help 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 China back in 1995 chose to locate in the Livorno rather than in the Grosseto province (both in Tuscany) because Livorno 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.[19] Moreover, the small size of provinces is important as it makes less likely that migration flows from a specific foreign country towards a single Italian province account for most immigration from that country towards Italy.

The results of the 2SLS estimation are reported in Table 11. 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. Column (1) shows that the export elasticity to immigrants is very low, and is not different from zero in statistical terms. Our results thus differ from Peri and Requena-Silvente (2010).[20] By contrast, the elasticity of imports to immigrants estimated with 2SLS is 0.548, highly statistically significant and larger than the OLS estimate.

Our 2SLS estimates suggest overall that the co-movement of immigrant stocks and exports may be spurious and mainly due to trading-pair *time-variant* factors having a positive correlation with both. By way of contrast, immigrants appear to have a genuine positive effect on the level of imports. This latter outcome could be due to a *business and social network effect* or to a *transplanted home-bias effect* in consumption; given the absence of an effect

---

[19] The opportunities for trade of Chinese immigrants, for instance, are likely to differ whether they locate in Germany or Italy, or whether they locate in Lombardy (Northern Italy) or Calabria (Southern Italy).

[20] However, it is important to stress that our 2SLS specification is different as we also included *trading-pair* fixed effects in the first stage.

Table 11: Gravity equations for exports and imports (2SLS) — province level

	Export (1)	Import (2)
First stage		
$\ln(1 + I\hat{M}M_{ijt-1})$	0.433***	
	(0.007)	
F-test instrument	3,871.20 [0.000]	
Second stage		
$\ln(Y_{it-1}Y_{jt-1})$	2.207***	1.947***
	(0.038)	(0.049)
$\ln(1 + IMM_{ijt-1})$	0.005	0.548***
	(0.038)	(0.053)
$\ln(dist_{ij})$	-1.938***	-2.786***
	(0.367)	(0.566)
N. observations	135,586	135,586
N. clusters	20,009	20,009

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

Note. The dependent variables are $\ln(1 + EXPORT_{ijt-1})$ and $\ln(1 + IMPORT_{ijt-1})$, i.e. export and import flows of province i from country j at time t, respectively. The regressions also control for *region × country*, *country × year* and *region × year* fixed effects. Trade flows cover the period 2003-2009. Robust standard errors are clustered at the *province × country* level and are reported in parenthesis, p-values in square brackets.

of immigrants on exports, we interpret the pro-trade effect of immigrants recorded on imports as the prevalence of the latter effect on the former.

5   Concluding remarks

This paper uses the large increase of immigrants from several countries into Italian provinces that took place in the 2000s, i.e. the Italian "*super-diversity*," to estimate the causal effect of immigrants on import and export flows. Using a panel of bilateral trade flows for 107 Italian provinces (NUTS-3) and 210 countries and corresponding data for immigrant stocks in Italian provinces by country of origin (189 countries) we find a large and robust elasticity of import flows to the stock of immigrants. Relying on an instrument built upon past immigrant enclaves in Italian provinces, we show that the in-

strumental variable estimates, including *region-country*, *country × year* and *region × year* fixed effects, indicate that a 10 percent increase in immigrant stocks leads to a 5.48 percent increase in import flows. On the other hand, the export elasticity to immigrants is very low, and is not different from zero in statistical terms.

The main explanation for the pro-trade effect of immigration given in the literature is related to the fact that immigrants reduce the fixed costs of exporting. We provide instead some new causal evidence that for the Italian case the pro-trade effect of immigrants takes place more through a *transplanted home-bias effect* in consumption (only affecting imports) than through a *business and social network effect* (affecting both exports and imports). By no means this implies neither that business and social networks are ineffective in fostering trade connections, since they play a potential role *also* in importing goods from the home-country, nor that the impact of immigrants on exports should always be insignificant. Trade elasticity to immigrants may be very well goods-specific, depending for instance on the degree of market differentiation in the host-country. Moreover, the trade-creating effects of immigrants may differ by immigrants' nationalities. When focusing on total trade and all nationalities of immigrants the estimated elasticity is an 'average' elasticity depending on the goods-mix composition of Italian trade, and on the nationalities composing Italian immigration. We show that such 'average' elasticity in Italy is positive and significant only for imports.

Appendix

Main variables' description

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 trading partners around the world, over the period 2002-2009. Data are measured such 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 US current dollars, using the nominal exchange rate from WDIs.

Immigrants. Data on foreign born residents by province (NUT-3) are taken from the demographic portal of ISTAT, and reports the stock of foreign-born residents per province at the 31 of December of each year (from 2002-2009).

GDP. Data on country Gross Domestic Product are taken from the *World Development Indicators* (WDIs), and are expressed in current US 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.

References

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

Anderson J.E. and van Wincoop E. (2004) Trade Costs, *Journal of Economic Literature*, 42, 3, 691–751.

Andersson R., Quigley J.M. and Wilhelmsson M. (2004) University decentralization as regional policy: the Swedish experiment, *Journal of Economic Geography*, 4, 4, 371–388.

Andersson R., Quigley J.M. and Wilhelmsson M. (2009) Urbanization, productivity, and innovation: Evidence from investment in higher education, *Journal of Urban Economics*, 66, 1, 2–15.

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., Dub J.P.H. and Gentzkow M. (2010) The Evolution of Brand Preferences. Evidence from Consumer Migration, mimeo.

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

Card D. (2001) Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration, *Journal of Labor Economics*, 19, 1, 22–64.

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 vs Extensive Margins, *Federal Reserve Bank of St. Louis Working Paper*, 2010-016B.

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

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

Dunlevy J.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.

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. (2011) The impact of immigration on international trade: a meta-analysis, *NORFACE MIGRATION Discussion Paper*, 2011-20, 507–525.

Girma S. and Yu Z. (2002) The Link between Immigration and Trade: Evidence from the United Kingdom, *Weltwirtschaftliches Archiv*, 138, 1, 115–130.

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.

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 (2010) La popolazione straniera residente in Italia, Technical report, Statistiche in breve - Istituto Nazionale di Statistica, Rome, Italy.

Jean-Christophe Dumont G.S. and Widmaier S. (2010) International Migrants in Developed, Emerging and Developing Countries:An Extended Profile, Technical Report 114, OECD Social, Employment and Migration Working Papers, Paris.

Mazzolari F. and Neumark D. (2010) The Effects of Immigration on Product Diversity, *mimeo*.

Peri G. and Requena-Silvente F. (2010) The trade creation effect of immigrants: evidence from the remarkablecase 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.

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

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/Steven

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.