The impact of immigration on regional exports: An application of the generalized propensity score

Guadalupe Serrano

Department of Economic Analysis, Faculty of Economics, University of Valencia, Avda dels Tarongers, s/n, Campus dels Tarongers, E-46022, Valencia, Spain

tel. 963828247 fax. 963828249

email: guadalupe.serrano@uv.es

Francisco Requena

Department of Applied Economics, Faculty of Economics, University of Valencia, Avda dels Tarongers, s/n, Campus dels Tarongers, E-46022, Valencia, Spain

tel. 963828851 fax. 963828249

email: francisco.requena@uv.es

Abstract: The migration-trade link has been studied extensively since the mid nineties, finding a positive impact through different channels. Based on the generalized propensity score (GPS) methodology, we estimate a dose-response function, depicting a non-linear impact of immigration on exports using regional data for Spain and Italy. For both countries the elasticity of province exports to immigration from a given nationality is always positive. However, it is magnitude varies with the level of immigrants: increasing with less than 100 immigrants; decreasing between 100 and 1500; increasing again with more than 1500. In contrast to previous studies that use country-level data, we do not find any exhaustion point in the effectiveness of the immigration networks on regional exports

Keywords: Immigration, exports, generalized propensity score, dose-response function, Spanish provinces, Italian provinces

JEL Classification: C210, F140, F220

Acknowledgements: We acknowledge financial support from Spanish Ministry of Science and Innovation (project ECO 2011-27619/ECON) and Generalitat Valenciana (project Prometeo/2009/098).

Introduction

The migration-trade link has been studied extensively in the economic literature since the mid 1990s. In their meta-analysis of the impact of immigration on trade, Genc et al (2011) uses 48 studies (300 estimates) and Lin (2011) uses 24 studies (184 estimates). The results confirm that an increase in the number of immigrants by 10 percent increases the volume of exports and imports by about 1-2 percent. Moreover, that impact is stronger for goods whose trade is likely to involve informational problems; and, in turn, that impact is stronger for trade with countries that are different from the reference country on a number of dimensions (different level of development, lack of common language and colonial ties); and that impact is stronger when the partner country is characterized by institutional problems.

With a few exceptions previous empirical studies assume that the relationship between migration and trade is linear. Gould (1994) analyses the impact of immigration on trade between the United States and 47 trading partners that were also sources of US immigrants for the period 1970-1986. He finds immigrants increased moderately both imports and exports. More interestingly he attempts to identify the level of immigration associated with the positive effect on trade. To do this, he estimated a specific functional form for the effect of immigrants on transaction costs which is decreasing at a decreasing rate and found that the effect of immigrants on exports is exhausted at a quite small level (around 12000 immigrants). More recently, Egger et al (2012) find a non-linear relationship between migration stocks in 2000 and bilateral trade flows in 2007 using country-level data. In particular, the elasticity of trade to migration is high when the number of immigrants in the OECD countries ranges between 200 and 4000. When the migration stocks exceed the last number, imports to OECD countries from the host countries of immigrants will not increase anymore. Both papers suggest that migrants possess economic, cultural and institutional knowledge about both the home and the host markets, so they are able to mediate economic exchange between those markets. Trade will increase while business networks are set up. But once the “bridge” has been constructed, the need for additional migrants to stimulate trade declines and even could vanish.

In this paper we investigate whether there is a causal effect of the level of immigrants (stock) on exports from the host province to the country of origin of the immigrants using data from Spanish and Italian provinces in year 2007. The hypothesis of a positive causal effect of immigration on exports is tested using the generalized propensity score (GPS) methodology recently developed by Hirano and Imbens (2004). The GPS methodology has a number of advantages compared to other econometric techniques. Firstly, the GPS method allows for continuous treatment, that is, we are able to determine the causal relationship between exports (the outcome) and immigrants (the treatment) at each value of immigration. Secondly, the GPS method enables us to identify the entire function of exports over all possible levels of immigration, so we can check if there is a nonlinear relationship between trade and immigration, as suggested by the seminal paper of Gould (1994). Thirdly, the GPS methodology allows us to analyze the level of immigration at which exports are maximized or whether the immigration-trade link exhibits turning points or discontinuities.

We argue that sub-national data provides greater precision in identifying ethnic networks and in estimating their impact on trade. Herander and Saavedra (2005) disentangle the impact of both the in-state and out-state stocks of immigrants of 36 countries on US state exports between 1993 and 1996. Since the impact of in-state immigrants is greater than that of out-state immigrants, they conclude that network links are about proximity. Using much smaller geographic units (provinces rather than states), Artal et al (2012) find for Spain. Italy and Portugal that the migration-trade link is clearly in-province: exports from a province to a country do not receive any stimuli from immigrants from this country living outside of the province. Thus, the trade-promoting effects of immigrant networks are greatest locally and any study about the functional form of the migration-trade link should take into account how immigrants are distributed within a country.[1]

Our paper is closely related to Egger et al (2012). They also use the GPS methodology and find that the migration-trade links is exhausted once a developed country reaches 4000 immigrants from a particular country. We want to examine if their results are still valid when we use sub-national data: 4000 immigrants of a particular nationality can be very concentrated or very disperse within a country. We find that after balancing the sample on a rich set of country and province covariates, the estimated dose–response function depicts a non-linear impact of immigration on exports for Spain and Italy. For both countries the elasticity of province exports to immigration from a given nationality is always positive. However, it is magnitude varies with the level of immigrants: increasing with less than 100 immigrants; decreasing between 100 and 1500; increasing again with more than 1500.

Our findings based on the estimation of a dose-response function confirm previous studies that find a positive impact of immigrants on exports using sub-national data. Moreover, the positive enhancing-trade effect of immigration networks is not exhausted for any level of immigrants in Spain and Italy, a result that does confirm the findings of Egger et al using country-level data.

The rest of this paper is structured as follows. Section 2 presents the empirical specification and the data. Section 3 presents the results of the GPS and Section 4 presents the results of the dose-response function. Section 5 concludes.

2. Theory and model specification

Two main channels have been described in the literature to explain how immigrants can enhance trade: the information/search cost channel and the transaction cost channel. Migrants can serve as information providers and trade intermediaries because they have a deep knowledge of their home country's

[1] Other papers that have taken advantage of the existence of trade and migration data collected at sub-national levels include: Canadian provinces (Wagner et al., 2002); US states (Bandyopadhyay et al., 2008); French departments (Briant et al., 2009); Spain provinces (Peri and Requena, 2010); and Italian provinces (Bratti et al, 2011).

opportunities and potential markets, access to distribution channels, contacts and familiarity to local customs, law and business practices. For certain type of products, especially differentiated ones, ethnic networks help to reduce the importance of informal barriers to international trade (Gould, 1994; Rauch and Trindade, 2002). These networks created by immigrants can be individual-specific or non individual-specific, depending on the mechanism through which they can reduce transaction costs and leading to different effects on trade (see Gould, 1994; Girma and Yu, 2002). Indeed, individual immigrants' business connections or personal contacts with the home country will lower transaction cost, but it could also be the case that transaction costs are lowered because immigrants bring to host country additional knowledge about foreign markets and about different social institutions. This second effect will be higher the more home and host countries are different and the less information is available on migrants' home countries.

However, the observed relationship between increased migrants of certain nationality and increased bilateral trade could be simply a result of selection bias. It is easy to imagine that provinces that export more to certain country also attract more immigrants of that country.

In this paper we investigate whether the trade-enhancing effect of the networks created by immigrants is the same for any level of immigrants: Is immigration a truly causal factor of trade? If so, is the trade-migration link linear? Does migration affect trade only after certain stock level or does migration lack of impact on trade after above certain threshold? If it is expected increasing trade above what it would be in the absence of migration, it is also expected that the need for additional migrants decreases once the network is working. An experimental approach to this problem would involve randomly assigning different number of immigrants of the same nationality in host destinations with similar characteristics and then observing which host destinations export more towards the country of origin of the immigrants. Unfortunately, such an experiment is practically impossible.

However we can perform an analysis that attempts to overcome this problem by balancing the treatment and control groups (those country-province pairs with less immigrants and those with more immigrants) to a host of country- and province characteristics, including geographic, demographic and economic information. The technique of utilizing the generalized propensity score to recover unbiased estimates of the trade-enhancing effect of immigrants. The effect of increased number of immigrants living in a host province on province´s exports to the country of origin of the immigrants is then estimated using a dose–response function, which provides a predicted outcome for every level of the treatment variable, conditional on a balanced distribution of known covariates.

This study follows from the existing migration-trade literature, but the primary contribution of the paper is to estimate the effect of migration on trade using an estimated dose–response function which is based on a balanced distribution of characteristics of the host provinces and sending countries. This approach has been recently implemented Egger et al (2012), which study the effect of immigrants on bilateral imports between 27 OECD host countries and a large number of trading partners in 2005. They find that stocks of less than 200 immigrants or more than 4000 immigrants of a specific nationality have no positive impact on bilateral imports. In our paper we examine the impact of immigrants on exports using sub-national data for two countries in 2007: 50 Spanish provinces and 103 Italian provinces. If we find a positive relationship between immigration and trade, this will provide additional support for the causal effect of migration on trade. If the relationship is not always positive for any level of immigrants, this will suggest that the trade enhancing effect of the immigration networks only works for certain level of immigrants. If the positive relationship depends on a certain number of immigrants of immigrants that get the network to work, that number will be defined at province-level rather than country-level.

2.1. Model specification and estimation

The approach in this section follows the standard counterfactual approach described by Rubin (1974). We utilize specific techniques described by Imbens (2000) and Hirano and Imbens (2004). For a sample of units indexed by i=1, … ,N there are a set of potential outcomes $Y_i(t)$ for a given level of treatment $t \in T$. The goal of the analysis is to estimate the average dose–response function (DRF), $\mu(t) = E[Y_i(t)]$ for all t. The observed variables for each unit *i* are a vector of covariates X_i, the level of treatment received T_i, and the observed outcome for the level of treatment actually received $Y_i = Y_i(T_i)$.

The primary assumption that needs to be made for this approach to work is that the information contained in the covariates X_i is sufficient to make the outcome independent of the level of treatment: $Y_i(t) \perp T_i | X_i \; \forall t \in T$ (*weak unconfoundedness).* Given this assumption, the generalized propensity score (GPS) is the conditional density of the treatment given the covariates: $r(t, x) = \int_{T|X}(t|X = x)$. If estimated correctly, the GPS has the property that the treatment groups will be independent of the covariates (loosely speaking, $X \perp Y(t)\{T = t\}|r(t, x)$). In combination with the *weak unconfoundedness* assumption, this *balancing property* also implies that the assignment to treatment is weakly unconfounded *given the GPS* (see Hirano and Imbens, 2004). Then, we can evaluate the GPS at a given treatment level by considering the conditional density of the respective treatment level t. Hence, each and every number of immigrants in a country-province pair i translates into a unique propensity score. This last result allows the estimation of the average DRF by using the GPS to remove selection bias.

This *balancing property* can be verified empirically in order to check that the GPS has been estimated properly. To check this assumption, the values of covariates at each level of the GPS and each level of treatment are compared with one another (a procedure we describe in more detail later in the paper). If the GPS has been estimated correctly, then the differences in means between these variables at various levels of the treatment variable (immigration stock) will not be statistically significant. In short, after controlling for the propensity to have more immigrants, the various treatment groups should not differ on the basis of other covariates.

Following Hirano and Imbens (2004), we utilize a flexible parametric specification of the GPS. First, the conditional level of the treatment (immigration stock) is modeled using the normal distribution:

(1) $\ln T_i | X_i \approx N(\beta_0 + \beta_1' X_i, \sigma^2)$

After estimating $(\beta_0, \beta_1, \sigma^2)$ by ordinary least squares, the GPS is calculated as:[2]

(2) $\widehat{R_\iota} = \frac{1}{\sqrt{2\pi\widehat{\sigma^2}}} exp\left(\frac{1}{2\widehat{\sigma^2}}\left(lnT_i - \widehat{\beta_0} - X'\widehat{\beta_1}\right)^2\right)$

In a second stage, we estimate the dose-response function using the estimated GPS by following two steps. The first step involves estimating the conditional expectation of Yi given Ti and $\widehat{R_\iota}$. Following Hinaro and Imbens (2004), we implement a partial mean approach by assuming a (flexible) parametric form for the regression function of Yi on Ti and $\widehat{R_\iota}$.

(3) $E[Y_i | T_i, \widehat{R_\iota}] = h(T_i, \widehat{R_\iota}; \alpha)$

The second step involves averaging this conditional expectation over $\widehat{R_\iota^t}$ to get the value of the dose-response function at t. Intuitively, we need to integrate over $\widehat{R_\iota^t} = \hat{r}(t, X_i)$ in the second step because the potential outcomes at t are independent of T conditional on $\widehat{R_\iota}$. With these estimates in hand, the dose–response function can be modeled over the entire range of treatment using the following:

(4) $E[\widehat{Y_\iota(t)}] = \frac{1}{N}\sum_{i=1}^{N} h(T_i, \widehat{R_\iota^t}; \hat{\alpha})$

The dose–response function estimates the average impact of each level of the treatment T for every observed level of the GPS values given the covariates. If the assumptions posited above hold, the effect of this is to remove bias from comparisons in the treatment status by balancing on the covariates.

3. DATA, GPS ESTIMATION AND DOSE-RESPONSE FUNCTION

To attempt to estimate the impact of migration on exports, we begin by presenting the data and reporting estimates from a standard OLS regression for each host country separately. Next, we discuss the properties of the estimated

[2] As Hirano and Imbens (2004) point out, there are other model specifications that can be used to estimate the GPS. OLS is the best estimator in this case as the dependent variable, while not normally distributed, is continuous, and the properties of OLS are well-known and the estimates easy to replicate.

generalized propensity score, particularly how well it balances each sample. We end this section by reporting the results of the dose–response function for each sample.

3.1. Data

We employ three sets of variables in this paper. First, we choose average bilateral exports from a province to a particular country in 2007 as the outcome of interest.[3] Second, we use bilateral stocks of foreigners residing in provinces of Italy and Spain at the beginning of 2006. Third, we use a large number of observed pre-treatment "push" and "pull" characteristics of bilateral migration as elements of X. Most of the variables at province level (population, GDP, unemployment rate, agriculture and non-public services share in value added) steam from ISTAT (Italian National Institute of Statistics) for Italian provinces and from INE (Spanish National Institute of Statistics) for Spanish provinces over the period 2002-2006. For the characteristics of the country of origin of immigrants we resort to the World Bank's WDI database for data on GDP, population, unemployment and income distribution (GINI). We follow Head and Mayer (2000) to construct the geodesic distance variable between each Italian and Spanish province and each foreign country. We calculate a weighted average of the great circle distance (in kilometers) from the capital of each province to the five most important cities of each partner country, in which the weights are the respective populations of the latter.[4] Bilateral culture distance measures based on language, education and industrialization differences are obtained from Dow and Karunaratna (2006).[5] Political freedom index is

[3] Trade flows are obtained from Agencia Tributaria - Aduanas (http://www.agenciatributaria.es) for Spain and from ISTAT (www.coeweb.istat.it) for Italy and cover the period 1993-2008.

[4] The great circle distance between i's and j's cities is calculated as follows. First we transform the latitude φ and the longitude λ into radians (x π /360). Second, the formula used to calculate the distance between the pair of cities is $\Delta_{ij} \equiv \lambda_j - \lambda_i$, $d_{ij} = \arccos\lfloor \sin\varphi_i \sin\varphi_j + \cos\varphi_i \cos\varphi_j \cos\Delta_{ij} \rfloor z$, with z= 6367 for km. Third, we calculate the population-weighted average distance between the capital of the province i and the cities of the foreign countries "cou" using the formula $D_{i,cou} = \sum_{j\in cou} w_j d_{ij}$, $w_j = pop_j / pop_{cou}$.

[5] Phychic distances can be downloaded from www.mbs.edu/home/dow/research/public/psydist.html

obtained from Freedom in the World (FIW) and Human right index is obtained from Amnesty International.[6]

Altogether, our study covers 50 Spanish provinces and 103 Italian province of residence as well 87 and 112 countries of origin of immigrants for Italy and Spain, respectively (see table A.1 in the Appendix for a list). The first columns of table 1 (SPAIN) and table 2 (ITALY) provide the mean, standard deviation, minimum and maximum for all variables (outcome, treatment and covariates) we use in our study. Variables in levels refer to year 2000 (except trade flows that refer to 200 and migration stocks that refer to 2006), while variables in growth rates are calculated for the period 2000-2005.

3.2. GPS estimation

Our first step in this analysis is to estimate the conditional distribution of immigration given the covariates (*weak unconfoundedness*). This is estimated via standard Ordinary Least Squares. The estimated coefficients of the GPS model and their estimated standard errors are reported in Columns 5 and 6 in table 1 (SPAIN) and table 2 (ITALY). Notice that we are not interested in the interpretation and statistical significance of the individual effects of the covariates in table 2 but in getting a powerful GPS. The adjusted R-square is 0.66 for Spain and 0.54 for Italy, indicating a priori that the GPS for Spain is slightly more powerful than for Italy. [7]

With the GPS in hand, the next step is test whether it does in fact have the property of making treatment groups independent of the covariates. To test this, we first compared the means for every covariate in the analysis for different levels of immigration. The entire sample of country-province pairs are distributed in quartile intervals according to the immigration stocks. Each interval contains 1006 observations in the case of Spain and 2125 observations

[6] Data are publicly available in www.freedomhouse.org and www.amnesty.org.
[7] We have explored alternative specifications for Spain and for Italy, separately. We included cubic terms of population and GDP or new variables (e.g. income per capita rather than GDP, a dummy if the province was in the coast). We report the preferred specification for Spain based in the value of R-square and in the assessment of the balancing property. We use the same specification for Italy because alternative specifications allowed us to reach high R-square (up to 0.75) but the assessment of the balancing property did not improve with respect to the one reported in the paper.

in the case of Italy. T-statistics for the test of equality of means are reported in the first columns of table 3 (SPAIN) and table 4 (ITALY) for each variable in each interval with respect to the other observations outside the interval. As the t-statistics of 7.45 for the *population country* variable in table 3 (SPAIN) shows, the average population size of the country of origin in country-province pairs with low number of immigrants is much lower than country-province pairs with larger number of immigrants. We repeated this analysis for the four intervals of immigrants for every covariate in the analysis. As tables 3 (SPAIN) and 4 (ITALY) show, there are large differences in the treatment intervals in terms of the covariates. For the 30×4 = 120 possible differences, 94 and 105 have t-statistics whose absolute value exceeds 1.96 in the Spanish sample and the Italian sample, respectively.

In the treatment literature, it is well known that methods that adjust for pre-treatment observable variables are likely to work poorly if there is not enough overlap in the distribution of covariates by treatment level. In that literature, it is common to gauge the overlap by looking at the distribution of the propensity score across treatment levels, sometimes restricting estimation to the common support region. In the case of continuous treatments it is considerably more difficult to gauge this condition since there is a continuum of treatment levels by definition and consequently multiple parameters of interest, each of them requiring a potentially different support condition.

In this paper we undertake two exercises to assess the balancing of covariates. The first exercise follows Dehejia and Wahba (2002) and consists of using histograms to check visually the extent of overlap in the supports of different levels of the treatment. For that purpose, we divide the GPS values into intervals and, for each interval, we compute the value of the GPS for each country-province pair at the median level of the treatment for the interval. Subsequently, we compute the value of the GPS at the same median level of the treatment for all country-province pair s that are not part of the interval in question. Finally, we compare the supports of the values of the GPS for these two groups (pairs in the interval in question and the rest) by superimposing their histograms.

Next we keep only control country-province pairs in other intervals than in the interval of reference if they share a common GPS support with treated pairs in the interval of reference. Since this is done for each of the four intervals, we ensure that each country-province pair within a certain interval lies within the range of observable characteristics of each other interval.

This exercise is repeated for each quartile in turn, resulting in four plots for each of our samples. These plots are shown in Figure 1 for the full Spanish sample and Figure 2 for the full Italian sample. These figures show that the overlap in the support of the estimated GPS across quartiles is very good in general. All observations in grey that lie outside the range of red bars are dropped. The exclusion of 1372 and 1131 country-province pairs for the Spanish and the Italian sample, respectively, ensures comparability of the ones left in each of the samples.

The second exercise follows Hirano and Imbens (2004) and consists of dividing the levels of the treatment into several intervals (quartile in our case). Then, within those intervals, we stratify country-province pairs into several groups of the GPS evaluated at the median value of the treatment of the corresponding interval (in our case we choose 6 groups). Within each quartile of this GPS range, we then compare the average covariate values for those in that range of the GPS with those outside of that range of the GPS, across each treatment level. These average differences in covariate values within each of the six GPS groups across treatment levels are then combined into a single figure, weighted by the number of respondents at each level of the GPS. The t-test for differences of means reported in tables 3 and 4 are based on this difference.

The differences in the treatment levels after balancing on the GPS appear on the right-hand side of table 3 for the Spanish sample and in table 4 for the Italian sample. When using only comparable units with a common probability support in GPS-space we are left with 2653 and 7372 country-province pairs in the Spanish and Italian samples, respectively. Regarding the balancing property, the last row in table 3 and 4 show that the median and average t-statistic drop drastically for all the intervals.

In the case of Spain, as the table 3 shows, there are few differences left among treatment intervals after balancing, with only one covariate (a dummy for border) whose t-statistic exceeds 1.96. Thus, the GPS as estimated has the desired property of balancing the Spanish sample. In the case of Italy, as table 4 shows, there are some significant differences left among treatment intervals after balancing, all of them corresponding to characteristics of the importing countries. Following the GPS literature (Mattei and Bia, 2008; Flores et al ,2012), if the balancing is rejected we should either re-estimate the GPS equation (1) including new covariates, or change the number of treatment intervals, or change the number of GPS groups. We have tried different specifications, increase the number of intervals (up to 8) and increase the number of groups (up to 10). For presentation purposes, we opted for maintaining the same specification, number of treatment intervals and GPS groups for Spain and Italy because we were not able to achieve the desired property of balancing for the Italian sample and, at the end, we realized that the estimates of the dose-response function and its derivative in the Italian sample did not change significantly under different specification and assessments of the balancing property.

3.3. Estimated dose–response function

The next step is to run a regression with the level of bilateral exports as the dependent variable and the conditional distribution of immigrants given the covariates on the right hand side (Eq. 3). We adopt a polynomial parameterization of the immigration stock (either in level or log transformed), the GPS (either in level or log transformed), and its interactive terms. The preferred specification based on the election of the variables in levels or log-form as well as the order of the polynomial terms was based on the Akaike Information Criterion (AIC) for each sample separately. The corresponding results for the preferred dose-response function are summarized in Table 5. The estimates from this regression do not have a direct interpretation, but are instead utilized in the calculation of the dose–response function (Hirano and Imbens, 2004).

The dose–response function is then estimated for every level of immigrants, as described in equation (4). The estimate of the level of exports in the dose–response function is an estimate of what would have happened to provinces´ exports to a particular country at each actual treatment level had they in fact been assigned to a different treatment level. So for instance, the estimated level of bilateral exports in country-province pairs with "*certain number*" immigrants in the dose-response graphic shows what would have happened on average to country-province pairs that had *any* level of immigrants. This is superior to standard estimates, which only show what the effect of the treatment is on country-province pairs at that specific level of treatment only.

The predicted level of exports at every level of treatment (immigration), given the covariates, is shown in figure 3.[8] The same graph displays the density distribution of immigrants by country-province pairs to illustrate that most provinces have less than 100 immigrants of certain nationality (57,2% in the Spanish sample and 72,1% in the Italian sample), while the percentage of provinces with more than 1500 immigrants of a certain country is small (12.1% in Spain and 4,8% in Italy). The dose-response functions for Spain and Italy exhibit very similar patterns. Quite clearly, there is a positive impact of immigration on exports at the country-province level. However, the level of exports varies in a non-linear way with immigration. Regional exports increase rapidly in the range from 1 to 100 immigrants, then decline in the range from 100 to 1500 and finally increase again.

The shape of the dose-response function and its derivative can be seen in more detail figure 4 (SPAIN) and 5 (ITALY) after we split the treatment level below and above the median (361 and 513, respectively, for Spain and Italy). The values of the derivative of the dose-response function are always positive, which indicates that any increase or decrease in the dose-response function is statistically different from zero. However, the magnitude of the change varies significantly. The value of the derivative of the dose-response function varies is

[8] We calculated the confidence intervals for the DRF and its derivative using bootstrapping. The intervals were very small so we did not report them in the graphs. It is important to point out that the derivative of the DRF was statistically different from zero at confidence level of 1% for any level of immigrants.

0,05 for Spain and 0,15 for Italy in country-province pairs with a less than 10 immigrants and 0,01 for country-province pairs with around 100 immigrants. Beyond 100 immigrants the derivative is close to zero, but statistically remains positive.

4. CONCLUSIONS

In this paper we examine the impact of immigrants on regional exports for Spain and Italy for the year 2007 using a relatively novel approach in this literature: the dose–response function shows the estimated impact of a given level of immigrants on the level of bilateral exports conditional on the generalized propensity score averaged over every country-province pair in the sample. This is in effect a way of answering the counterfactual question of what would have happened to a given province-country pair had they received a different level of immigrants.

Our results show that any level of immigrants living in a province stimulates exports from this province to the country of origin of immigrants. Only in the range of immigrants between 100 and 1500, there is a decline in the magnitude of impact, although the impact is still positive.

Our conclusions using sub-national data are different from to those of Egger et al (2012) using country-level data: there is no need for a critical mass of immigrants to make a trade-enhancing effect "networks" operative within a province in Spain and Italy; there is no exhaustion point in the effectiveness of the immigration networks on regional exports.

REFERENCES

Bandyopadhyay S., Coughlin C.C. and Wall H.J. (2008), "Ethnic Networks and State Exports", Review of International Economics, 16, 1, 199-213.

Bratti, M., De Benedictis, L. and Santoni, G. (2011), "On the Pro-trade Effects of Immigrants", mimeo, http://works.bepress.com/luca_de_benedictis/21

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.

Dehejia, Rajeev H. and Wahba, Sadek (2002), "Propensity Score-Matching Methods for Nonexperimental Causal Studies." Review of Economics and Statistics, 84(1), 151-161.

Dow, D. and Karunaratna, A. (2006), "Developing a Multidimensional Instrument to Measure Psychic Distance Stimuli", Journal of International Business Studies, 37(5), 575-577.

Egger, P., Ehlrich, M. and Nelson, D. (2012) "Migration and Trade", World Economy, doi: 10.1111/j.1467-9701.2011.01429.x

Flores, C. A., Flores-Lagunes, A., Gonzalez, A. and Neumann, T.C. (2012), "Estimating the effects of length of exposure to a training program: The case of Job Corps". The Review of Economics and Statistics, 3, 94 (1), 153-171

Genc, M., Gheasi, M. Nijkam, P. and pot, J. (2011), "The Impact of Immigration on International Trade: A Meta-Analysis", IZA Discussion Papers, number 6145

Gould D. M. (1994), "Immigrant Links to the Home Country: Empirical Implications for United States Bilateral Trade Flows", Review of Economics and Statistics, 76, 2, 302-16.

Hirano K. and G. Imbens (2004), "The Propensity Score with Continuous Treatment", in A. Gelman and X.-L. Meng (eds.), Applied Bayesian Modelling and Causal Inference from Missing Data Perspectives,(New York: Wiley),73-84.

Imbens, G. (2000). The role of the propensity score in estimating dose–response functions. Biometrika, 87, 706–710.

Lin, F. (2011) The pro-trade impacts of immigrants: a meta-analysis of network effects. Journal of Chinese Economic and Foreign Trade Studies 4(1): 17-27.

Mattei A., and Bia M. (2008) " A STATA Package for the Estimation of the Dose-Response Function through Adjustment for the Generalized Propensity Score". The STATA Journal , 8(3), 354-373.

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.

Wagner D., Head K. and Ries J. (2002), "Immigration and the Trade of Provinces", Scottish Journal of Political Economy, 49 (5), 507-525.

Table 1.SPAIN.

Summary Statistics of Covariates and Their Coefficients in the GPS Model.

		50 SPANISH PROVINCES				50 SPANISH PROVINCES		
		Summary descriptives				GPS model		
		(1)	(2)	(3)	(4)	(5)		(6)
		Mean	Std. Dev.	Min	Max	GPS coeff		s.e.
Outcome	ln(exports prov-cou). N=3778	7,43	2,99	0	15,78			
Treatment	number immigrants prov-cou. N=4025	1077	5653	1	148906			
Covariates	ln(population cou)	9,80	1,48	6,46	14,09	1,575	***	0,202
	ln(population cou)2	98,14	30,24	41,80	198,57	0,002		0,011
	ln(GDP cou)	11,68	1,91	7,42	16,44	-0,947	***	0,137
	ln(GDP cou)2	139,97	44,99	55,01	270,14	-0,004		0,007
	growth population cou	1,00	1,03	-1,54	3,54	-0,283	***	0,034
	growth GDP cou	14,86	5,50	2,23	35,36	0,105	***	0,006
	ln(physical distance cou)	8,26	0,80	5,41	9,90	-1,319	***	0,045
	distance language cou	-0,56	1,62	-3,87	0,53	-0,690	***	0,030
	education distance cou	1,39	0,69	0,10	2,79	-0,268	***	0,084
	industrialisation distance cou	0,42	0,68	-1,26	1,45	-1,326	***	0,112
	political freedom index cou	2,80	1,70	1,00	7,00	-0,057	**	0,025
	human right index cou	2,42	0,95	1,00	4,92	-0,030		0,048
	Gini cou	39,27	9,54	24,70	60,00	0,027	***	0,006
	unemployment rate cou	9,98	6,60	0,70	36,10	0,010	**	0,004
	border cou	0,04	0,19	0,00	1,00	1,390	***	0,178
	member EUEFTA cou	0,27	0,44	0,00	1,00	0,593	***	0,072
	dummy =1 if imports 1995 > 0 prov-cou	0,71	0,45	0,00	1,00	0,361	***	0,069
	ln(1+number of emigrants in cou)	2,29	2,33	0,00	10,57	0,221	***	0,017
	ln(population prov)	6,44	0,84	4,52	8,71	-5,876	***	1,365
	ln(population prov)2	42,17	11,16	20,46	75,83	0,490	***	0,106
	ln(GDP prov)	9,53	0,89	7,63	12,14	7,418	***	1,758
	ln(GDP prov)2	91,59	17,35	58,22	147,29	-0,368	***	0,092
	growth population prov	1,25	1,11	-0,46	3,71	0,381	***	0,029
	growth GDP prov	7,30	0,82	4,98	9,16	-0,078	**	0,035
	unemployment rate prov	7,77	2,73	2,16	13,67	-0,011		0,015
	unemployment rate change prov	-5,09	3,49	-14,56	1,15	0,045	***	0,010
	share agriculture in value added prov	5,41	3,92	0,17	17,98	-0,001		0,010
	share services in value added prov	64,25	6,89	53,60	83,24	0,036	***	0,005
	growth in agriculture value added prov	0,04	0,14	-0,23	0,56	0,028		0,199
	growth in services value added prov	0,30	0,21	-0,05	1,01	0,250	*	0,133
	Constant					-13,300	**	4,589
	R-squared (GPS model)					0,661		
	Observations (full sample)					4025		

Note: All the covariates expressed in levels refer to year 2000 and those expressed in growth rates refer to period 2000-2005. In the GPS estimation, the dependent variable is ln(number immigrants prov-cou) in 2006. Symbols ***, **, * stand for significance level at the 1%, 5% and 10%, respectively.

Table 2.ITALY.

Summary Statistics of Covariates and Their Coefficients in the GPS Model.

		103 ITALIAN PROVINCES Summary descriptives				103 ITALIAN PROVINCES GPS model	
		(1)	(2)	(3)	(4)	(5)	(6)
		Mean	Std. Dev.	Min	Max	GPS coeff	s.e.
Outcome	ln(exports prov-cou). N=8004	7,92	2,77	0	15,39		
Treatment	number immigrants prov-cou. N=8503	319	1559	1	62020		
Covariates	ln(population cou)	9,89	1,43	7,20	14,09	1,266 ***	0,133
	ln(population cou)2	99,94	29,52	51,86	198,57	-0,024 ***	0,007
	ln(GDP cou)	11,66	1,85	7,42	16,44	-0,267 **	0,105
	ln(GDP cou)2	139,44	43,88	55,01	270,14	0,003	0,005
	growth population cou	1,02	1,03	-1,54	3,54	-0,852 ***	0,024
	growth GDP cou	14,66	4,94	2,23	28,97	0,030 ***	0,004
	ln(physical distance cou)	8,07	1,01	4,31	9,85	-0,622 ***	0,027
	distance language cou	0,14	0,34	-0,74	0,53	-0,147 **	0,068
	education distance cou	1,06	0,75	-0,42	2,53	0,667 ***	0,055
	industrialisation distance cou	0,72	0,67	-0,96	1,75	0,296 ***	0,083
	political freedom index cou	2,83	1,64	1,00	7,00	0,008	0,018
	human right index cou	2,47	0,96	1,00	4,92	0,086 ***	0,036
	Gini cou	38,76	9,06	24,70	58,65	-0,010 ***	0,004
	unemployment rate cou	9,85	6,53	0,70	36,10	0,011 ***	0,004
	border cou	0,08	0,28	0,00	1,00	0,851 ***	0,067
	member EUEFTA cou	0,24	0,45	0,00	1,00	0,001	0,044
	dummy =1 if imports 1995 > 0 prov-cou	0,92	0,26	0,00	1,00	0,406 ***	0,067
	ln(1+number of emigrants in cou)	2,52	2,44	0,00	13,35	0,196 ***	0,009
	ln(population prov)	6,09	0,72	4,49	8,30	0,858	0,674
	ln(population prov)2	37,59	9,20	20,15	68,85	-0,087 *	0,052
	ln(GDP prov)	9,30	0,78	7,45	11,90	-1,462 *	0,794
	ln(GDP prov)2	87,08	14,96	55,54	141,71	0,137 ***	0,040
	growth population prov	0,58	1,17	-9,83	1,95	0,138 ***	0,017
	growth GDP prov	3,63	0,95	1,50	7,23	-0,046 *	0,028
	unemployment rate prov	9,63	7,58	1,71	30,53	-0,058 ***	0,010
	unemployment rate change prov	-3,06	4,30	-17,83	1,49	-0,054 ***	0,012
	share agriculture in value added prov	3,62	2,26	0,27	11,10	-0,013	0,009
	share services in value added prov	68,17	7,56	52,70	86,75	0,014 ***	0,003
	growth in agriculture value added prov	-0,02	0,03	-0,11	0,08	0,333	0,620
	growth in services value added prov	0,20	0,11	0,01	0,69	0,393 ***	0,145
	Constant					-2,150 **	2,314
	R-squared (GPS model)					0,540	
	Observations (full sample)					8503	

Note: See table 1.

Table 3. SPAIN.

Differences in the treatment levels before and after balancing on the GPS: t-stats for equality of means

Covariates	Prior to balancing on the GPS Interval Q1	Interval Q2	Interval Q3	Interval Q4	After balancing on the GPS Interval Q1	Interval Q2	Interval Q3	Interval Q4
ln(population cou)	**7,45**	**6,80**	**-3,01**	**-11,33**	1,45	-1,67	-0,23	0,77
ln(population cou)2	**7,37**	**6,44**	**-3,20**	**-10,68**	1,35	-1,69	-0,11	0,71
ln(GDP cou)	**7,99**	**4,84**	**-4,05**	**-8,79**	-0,01	-0,18	0,82	-0,68
ln(GDP cou)2	**8,32**	**4,80**	**-4,61**	**-8,52**	-0,13	-0,17	0,98	-0,78
growth population cou	**-7,03**	-1,61	**4,11**	**4,50**	0,42	-0,14	-0,65	0,30
growth GDP cou	**2,54**	**4,11**	-1,72	**-4,93**	0,52	-1,40	0,15	1,10
ln(physical distance cou)	**-6,76**	-1,20	1,79	**6,15**	1,50	-1,21	-0,74	0,82
distance language cou	**-12,62**	**-3,28**	**5,39**	**10,41**	0,69	-0,35	-0,14	-0,51
education distance cou	**-3,24**	-0,18	**3,01**	0,41	1,41	-0,97	-1,49	1,51
industrialisation distance cou	**-3,58**	0,68	**3,03**	-0,13	1,65	-1,65	-1,45	1,02
political freedom index cou	**-4,05**	0,56	**2,35**	1,13	0,96	-0,96	-0,82	1,10
human right index cou	0,63	**4,28**	1,60	**-6,52**	1,60	-1,41	-1,31	1,72
Gini cou	**7,29**	**5,55**	**-2,32**	**-10,59**	1,35	-1,31	-0,81	1,53
unemployment rate cou	1,49	1,88	-1,45	-1,92	-0,08	-0,76	0,81	0,05
border cou	**7,25**	**7,25**	0,86	**-15,73**	-0,94	-1,52	-1,43	**2,60**
member EUEFTA cou	**6,64**	0,67	-1,38	**-5,91**	-0,45	0,50	0,10	0,06
dummy =1 if imports 1995 > 0 prov-cou	**16,58**	**1,98**	**-3,90**	**-14,49**	-0,16	1,53	-0,61	-0,96
ln(number of emigrants prov-cou)	**20,87**	**8,90**	**-7,76**	**-22,17**	-1,88	0,54	1,75	-0,45
ln(population prov)	**20,29**	**3,04**	**-2,17**	**-21,31**	-0,15	1,68	-0,22	-1,51
ln(population prov)2	**19,99**	**3,73**	**-2,10**	**-21,87**	-0,20	1,63	-0,15	-1,49
ln(GDP prov)	**21,06**	**3,10**	**-1,96**	**-22,39**	-0,02	1,80	-0,49	-1,42
ln(GDP prov)2	**20,70**	**3,61**	-1,89	**-22,72**	-0,07	1,77	-0,44	-1,41
growth population prov	**13,85**	**3,30**	**-2,97**	**-14,19**	0,84	0,39	-0,54	-0,78
growth GDP prov	**5,77**	0,19	-1,17	**-4,78**	0,38	0,32	-0,33	-0,23
unemployment rate prov	**2,39**	-0,85	-1,40	-0,14	-0,20	0,04	0,46	-0,55
unemployment rate change prov	**5,72**	**2,41**	0,68	**-7,85**	0,31	0,59	-0,77	-0,10
share agriculture in value added prov	**-17,61**	0,07	1,39	**16,00**	-0,07	-1,81	0,71	1,18
share services in value added prov	**13,61**	**3,25**	**-2,75**	**-14,12**	-0,70	0,78	0,40	-1,00
growth in agriculture value added prov	**-8,90**	0,37	1,84	**6,64**	0,41	-0,79	-0,16	0,86
growth in services value added prov	0,90	-1,67	0,10	0,67	-0,09	-0,03	0,12	0,32
Observations	1006	1006	1006	1007	568	830	736	519
Median abs(t-value)	**7,27**	**3,19**	**2,24**	**8,65**	0,48	1,09	0,63	0,80
Mean abs(t-value)	**9,29**	**3,28**	**2,69**	**9,98**	0,72	1,01	0,68	0,93

Note: t-values reported in bold face indicate significance at the 5% level. The four intervals of approximately the same size are generated according to the distribution of migration stocks. Observations which do not satisfy the common support condition are excluded from the respective intervals. In order to account for the GPS values we split up the GPS values evaluated at the median treatment intensity of the respective interval into six groups of approximately same size according to the GPS distribution.

Table 4. ITALY.

Differences in the treatment levels before and after balancing on the GPS: t-stats for equality of means

	Prior to balancing on the GPS				After balancing on the GPS			
Covariates	Interval Q1	Interval Q2	Interval Q3	Interval Q4	Interval Q1	Interval Q2	Interval Q3	Interval Q4
ln(population cou)	**21,02**	**9,21**	**-7,62**	**-22,72**	-1,23	-1,62	0,95	**2,34**
ln(population cou)2	**20,69**	**9,37**	**-6,94**	**-23,29**	-1,16	-1,55	0,81	**2,29**
ln(GDP cou)	**25,52**	1,21	**-14,73**	**-11,34**	**-4,07**	**2,09**	**4,03**	-1,43
ln(GDP cou)2	**25,07**	1,76	**-14,41**	**-11,80**	**-3,98**	1,94	**3,95**	-1,32
growth population cou	**-18,46**	-0,86	**4,68**	**14,45**	1,25	-1,82	-0,61	-0,77
growth GDP cou	**6,70**	**2,66**	**2,34**	**-11,78**	0,65	0,39	-2,17	1,63
ln(physical distance cou)	**-24,17**	**-3,92**	**9,80**	**17,87**	**3,07**	-1,17	**-2,66**	1,82
distance language cou	-1,84	1,32	**6,56**	**-6,04**	1,05	0,22	**-1,26**	0,77
education distance cou	**-10,00**	**8,16**	**10,22**	**-8,38**	**5,54**	**-3,32**	**-5,07**	**2,32**
industrialisation distance cou	**-8,94**	**7,00**	**10,75**	**-8,80**	**4,63**	**-3,08**	**-4,17**	**3,21**
political freedom index cou	**-2,14**	**7,45**	**5,88**	**-11,23**	**3,37**	**-2,47**	**-2,80**	**2,20**
human right index cou	**5,71**	**11,05**	-0,19	**-16,74**	**2,20**	**-3,55**	-1,71	**2,82**
Gini cou	**-14,69**	**4,83**	**5,74**	**3,96**	**3,02**	**-3,36**	-1,50	**2,09**
unemployment rate cou	**5,09**	**8,73**	0,93	**-14,88**	**2,32**	-1,94	-1,21	1,06
border cou	**9,54**	**3,80**	**-2,72**	**-10,63**	-0,40	0,55	0,96	-1,18
member EUEFTA cou	**10,58**	**-4,00**	**-9,06**	**2,50**	**-4,24**	**2,23**	**4,01**	**-2,24**
dummy =1 if imports 1995 > 0 prov-cou	**19,89**	1,38	**-8,35**	**-12,63**	-1,94	0,32	1,48	0,70
ln(number of emigrants prov-cou)	**18,07**	1,05	**-13,79**	**-5,14**	**-4,17**	1,29	**4,96**	**-2,45**
ln(population prov)	**15,51**	**7,51**	-1,83	**-21,50**	-0,52	-0,06	0,82	-0,92
ln(population prov)2	**15,34**	**7,69**	-1,56	**-21,81**	-0,52	-0,08	0,84	-0,89
ln(GDP prov)	**18,49**	**9,22**	**-2,71**	**-25,51**	-0,66	-0,22	1,09	-1,02
ln(GDP prov)2	**18,32**	**9,36**	**-2,41**	**-25,81**	-0,67	-0,22	1,10	-0,98
growth population prov	**10,13**	**3,27**	**-3,05**	**-10,35**	-0,28	-0,10	0,51	0,10
growth GDP prov	-1,60	-0,78	-0,02	**2,39**	0,00	-0,18	-0,11	0,38
unemployment rate prov	**-9,00**	**-5,39**	**3,20**	**11,22**	0,03	0,58	-0,48	0,02
unemployment rate change prov	**6,47**	**4,63**	**-2,28**	**-8,84**	0,20	-0,58	0,32	0,07
share agriculture in value added prov	**-11,80**	**-6,29**	**3,39**	**14,78**	0,30	0,44	-0,97	0,72
share services in value added prov	**-3,41**	-1,41	**1,97**	**2,84**	-0,38	0,28	-0,08	-0,05
growth in agriculture value added prov	**-5,54**	**-2,93**	1,94	**6,53**	-0,02	0,22	-0,39	0,41
growth in services value added prov	-1,48	**-2,09**	0,44	**3,12**	-0,30	0,31	-0,28	0,26
Observations	2125	2125	2125	2128	1914	2013	1962	1483
Median abs(t-value)	**12,64**	**4,73**	**5,21**	**11,56**	1,24	1,23	1,24	1,25
Mean abs(t-value)	**13,19**	**5,22**	**5,84**	**13,14**	1,97	1,34	1,91	1,42

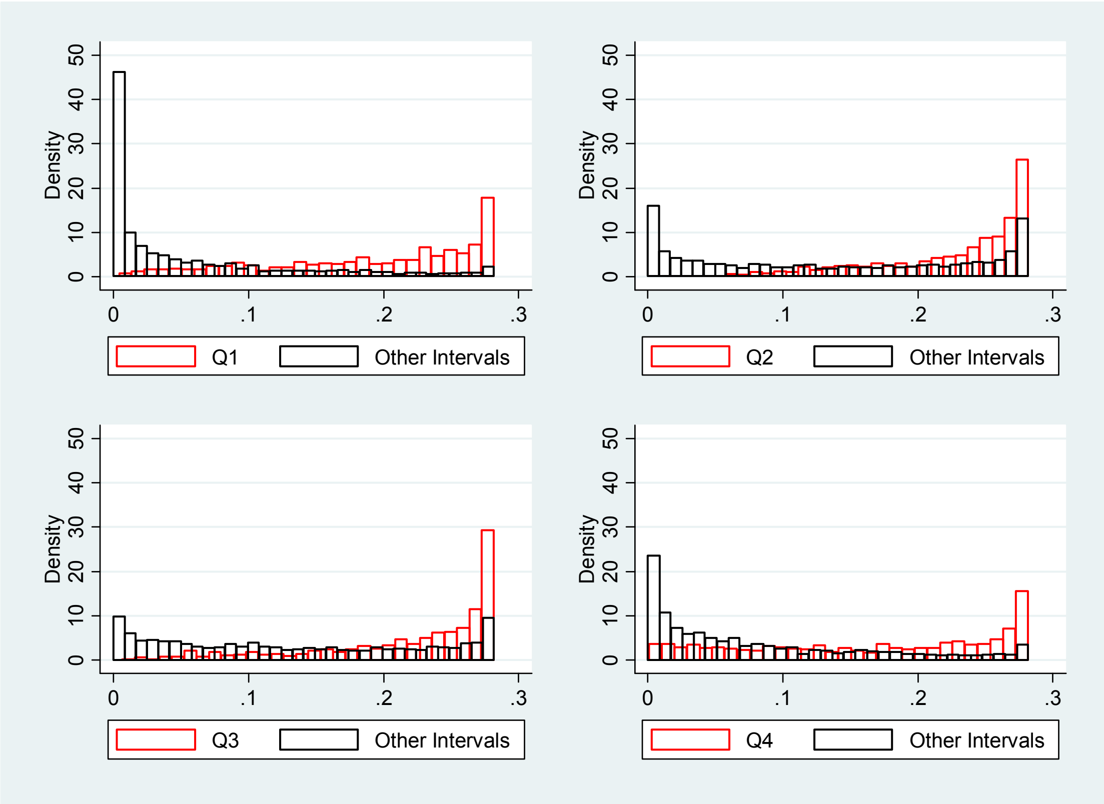
Note: See table 3

Table 5. SPAIN & ITALY.
Estimated Parameters of the Conditional Distribution of log of exports given log of Immigration Stocks and the GPS.

	SPAIN	ITALY
ln(migr)	0.928***	1.027***
	[0.366]	[0.109]
ln(migr)^2	-0.399***	-0.581***
	[0.109]	[0.0369]
ln(migr)^3	0.0355***	0.0552***
	[0.0085]	[0.00333]
R	-8.227***	-29.82***
	[2.627]	[1.979]
R^2		24.68***
		[5.826]
R x ln(migr)	3.396***	8.252***
	[0.584]	[0.234]
Constant	6.297***	7.907***
	[0.329]	[0.172]
Observations	2516	7730
R-squared	0.033	0.281
AIC	10811.32	37940.61

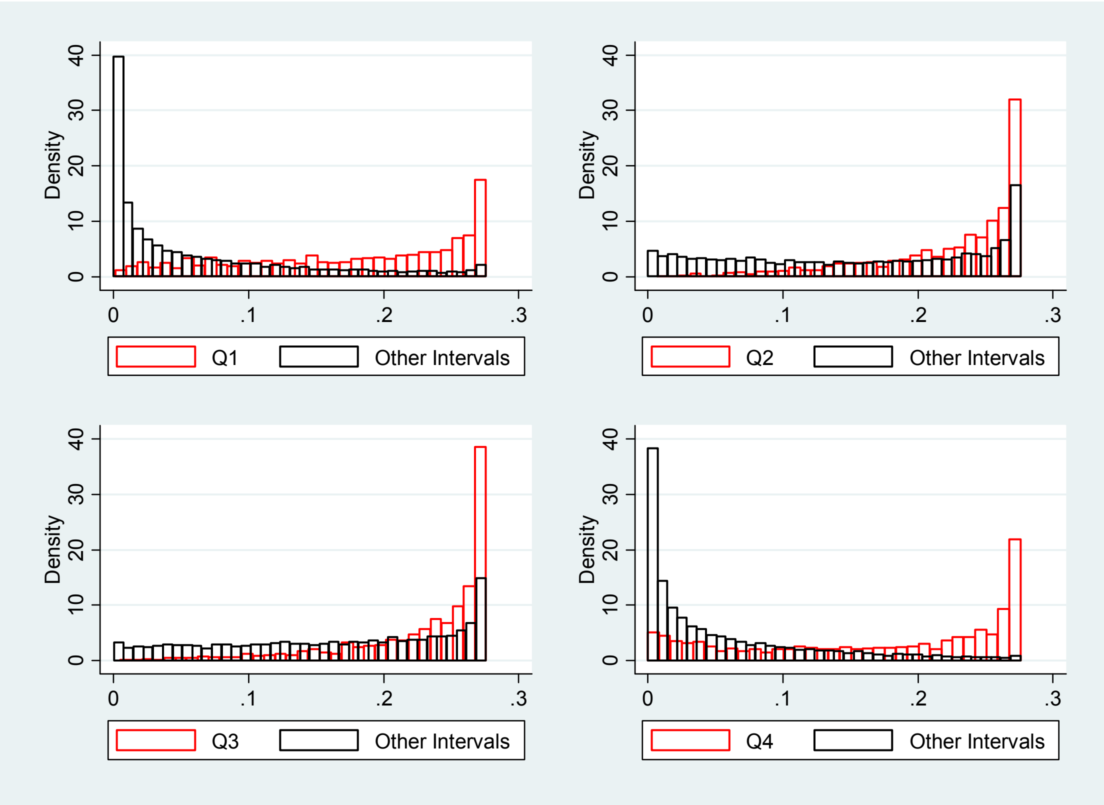
Note: ***, ** and * for significance levels at 1, 5, and 10%, respectively. $\ln(migr)$ refers to logarithm of the stock of immigrants in exporter province from importing country. R refers to generalized propensity score calculated according to equation (2) using the coefficients from the first stage regression in Table 1 for Spain and Table 2 for Italy. We estimate the standard errors of the dose-response function by bootstrapping with 1000 iterations that take into account that the second-stage estimates involve imprecision from first-stage estimates. We did not include country-province pairs with zero export flows.

Figure 1. SPAIN. GPS Support Condition



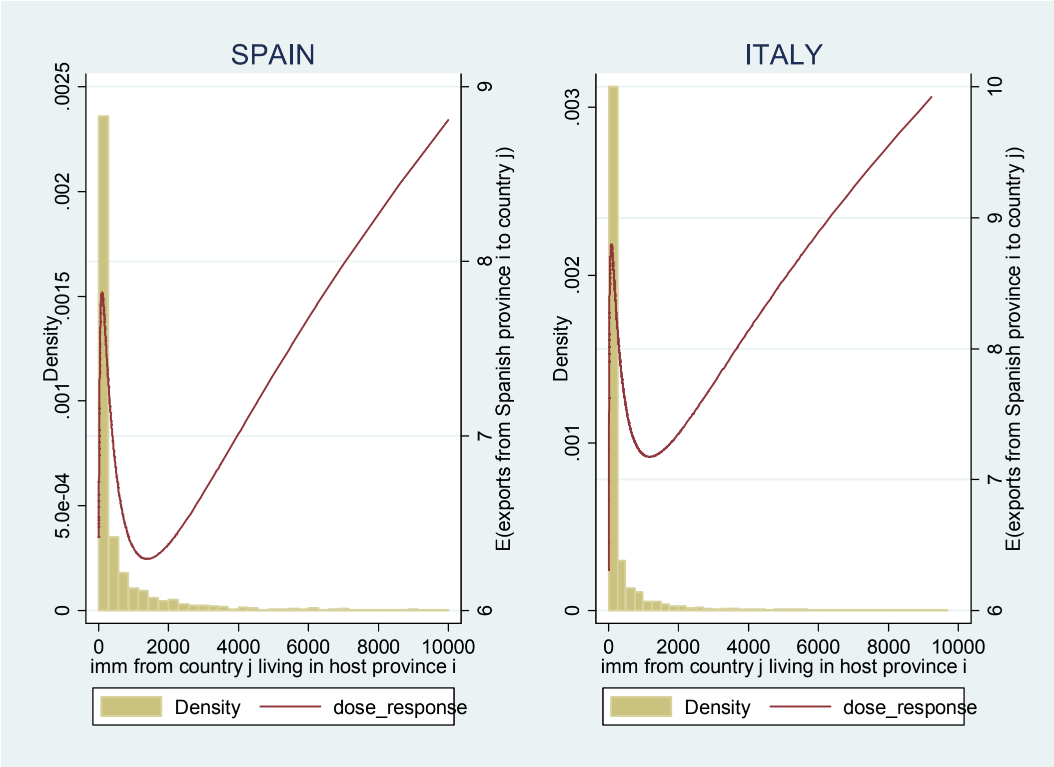
Notes: There are four intervals of equal size, each of them having 25% of the distribution of migration stocks (in ascending order). Country-province pairs with relatively low migration belong to interval Q1whereas pairs with high migration belong to interval Q4. In each histogram, the generalized propensity scores are evaluated at the median migration level of the respective interval, for both the observations within that particular interval as well as for the respective control observations belonging to all other intervals.

Figure 2. ITALY. GPS Support Condition



Notes: See Figure 1

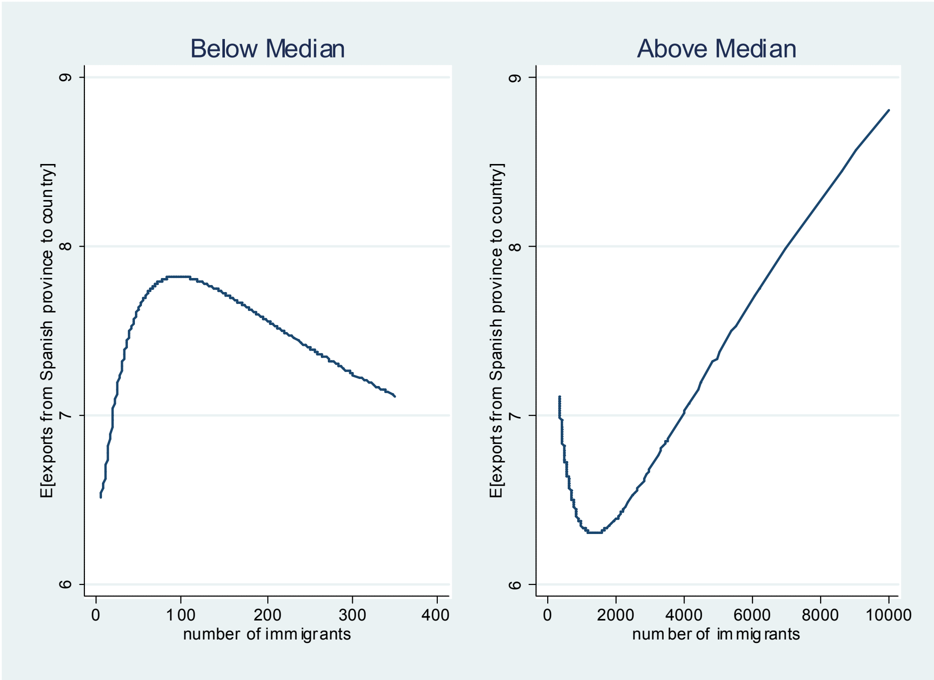
Figure 3. Distribution of immigrants stocks (density) and dose-response function



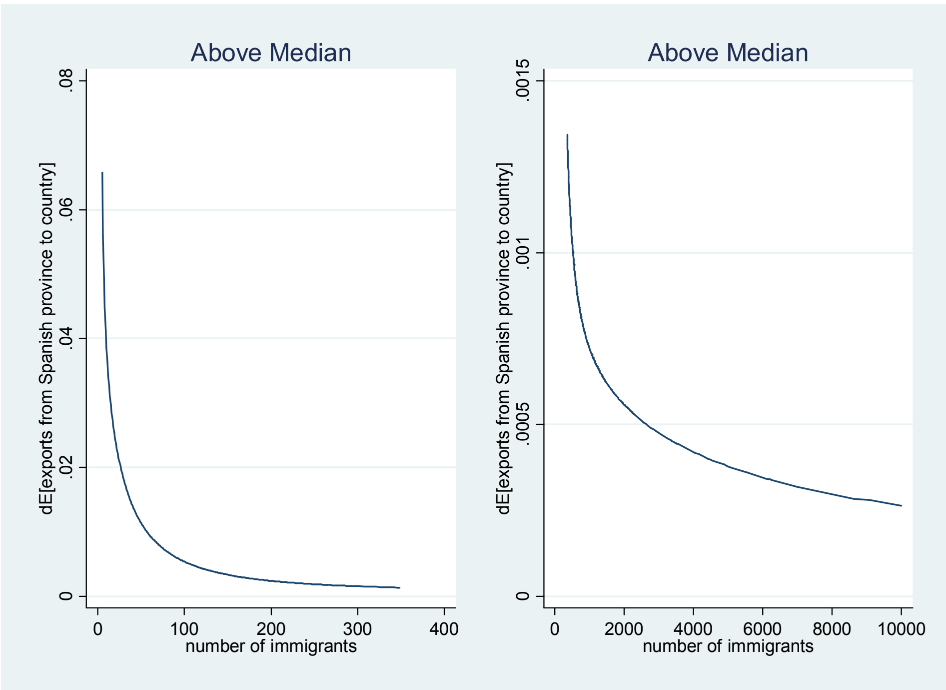
Note: The density distribution of immigration is based on the total number of country-province pairs in the sample. Density for values of immigration above 10000 have been trimmed for presentation purpose.

Figure 4. SPAIN. Dose-response function and Derivative function of dose-response function

DOSE-RESPONSE FUNCTION (SPAIN)



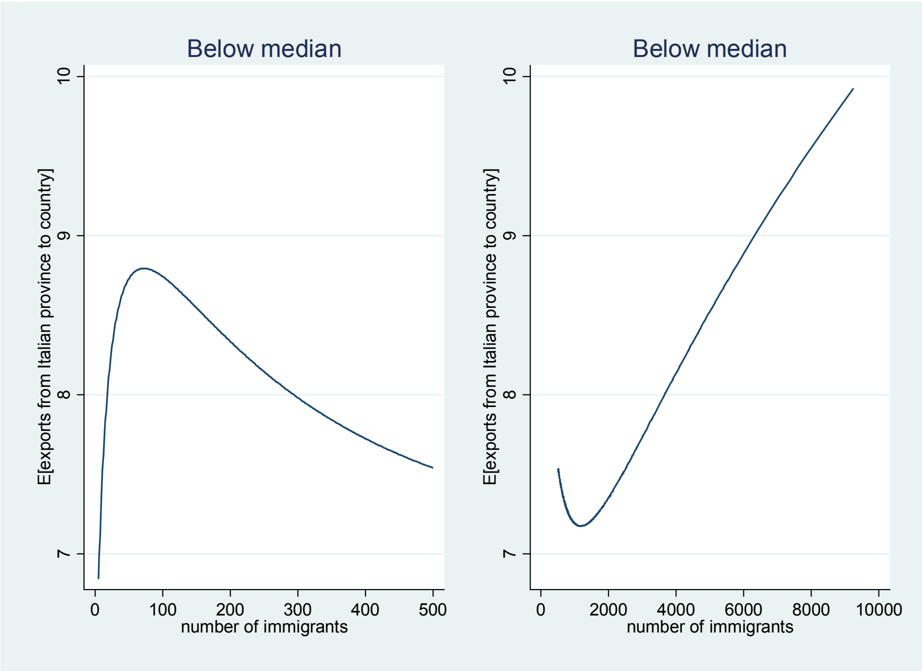
DERIVATE OF DOSE-RESPONSE FUNCTION (SPAIN)



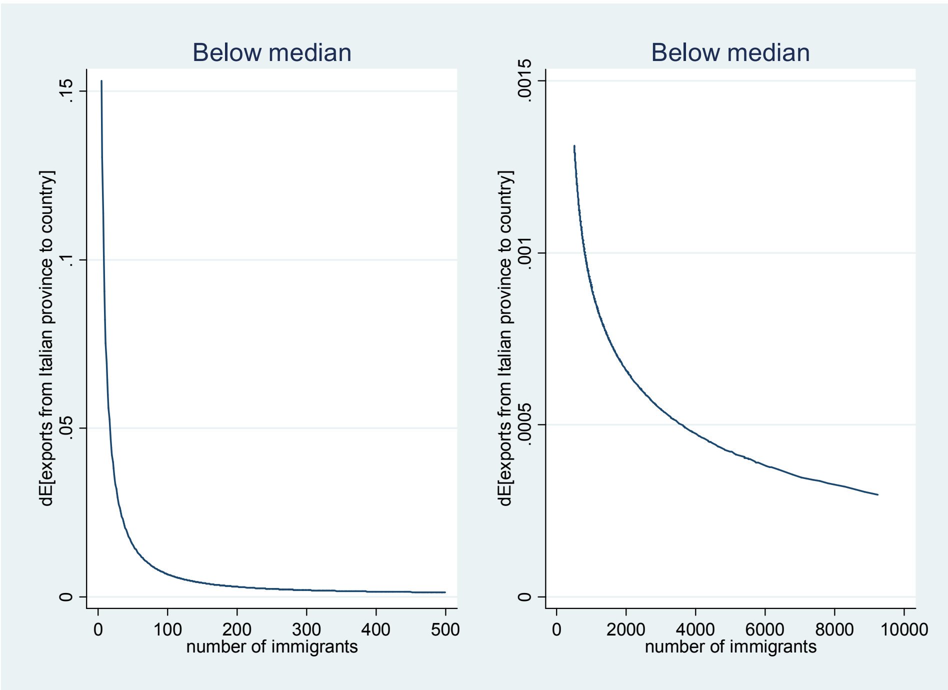
Note: The median value of the treatment is 361 based on the range of treatment values. Observations with treatment level above 10000 immigrants are trimmed for presentation purpose.

Figure 5. ITALY. Dose-response function and Derivative function of dose-response function

DOSE-RESPONSE FUNCTION (ITALY)



DERIVATE OF DOSE-RESPONSE FUNCTION (ITALY)



Note: The median value of the treatment is 513 based on the range of treatment values. Treatment level above 10000 immigrants are trimmed for presentation purpose.