No 2016-22 – September

Working Paper

CEPII

Third Country Effect of Migration: the Trade-Migration Nexus Revisited

Erik Figueiredo, Luiz Renato Lima & Gianluca Orefice

Highlights

- This paper quantifies the trade-creating effect of international migrants and identifies a new channel through which immigrants coming from third high (tariff) protected countries can affect bilateral specific import flows (third country effect).
- According to our preferred estimations, a ten percent increase in the flow of bilateral migrants increases bilateral imports by 0.45% (preference or information channel).
- A ten percent increase in the flow of migrants coming from third high (tariff) protected countries increases bilateral imports by 0.27% (third country effect).

Abstract

This paper proposes a new channel through which migrants can affect the import demand of the host country. In migrating from origin to destination country, migrants observe a change in the prices of the bundle of consumable goods. In particular, the migration decision can reflect a reduction in the price of imported goods (due to lower applied tariff) for the consumption bundle of migrants: emigration towards less (tariff) protected countries allows the consumption of products that were prohibitively protected in the origin countries of migrants. To test this channel we estimate the import demand effect of migrant groups coming from third high (tariff) protected countries. We use a theory-grounded gravity estimations and a fresh econometric techniques able to address both the zero migration flows problem and the endogeneity of migrants. Our results suggest that such a third-country immigrant effect is significant and positive.

Keywords

Trade-Migration, Third-Country Effect, Quantile Regression, Imputation.

JEL

F14, C21, C36.

Working Paper

CEPII

CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) is a French institute dedicated to producing independent, policy-oriented economic research helpful to understand the international economic environment and challenges in the areas of trade policy, competitiveness, macroeconomics, international finance and growth.

CEPII Working Paper
Contributing to research in international economics

Editorial Director:
Sébastien Jean

Production:
Laure Boivin

No ISSN: 1293-2574

CEPII
113, rue de Grenelle
75007 Paris
+33 1 53 68 55 00

www.cepii.fr
Press contact: presse@cepii.fr

RESEARCH AND EXPERTISE
ON THE WORLD ECONOMY

Third country effect of migration: the trade-migration nexus revisited[1]

Erik Figueiredo (Federal University of Paraiba)*

Luiz Renato Lima (University of Tennessee)†

Gianluca Orefice (CEPII)‡

The trade-migration nexus has been widely analyzed in the literature, often in the context of investigating the role of business and social network in shaping bilateral trade. Rauch (2011) and Wagner, Head and Ries (2002) suggest that international business and social networks - approximated by international migrants - promote trade by reducing the information cost and the diffusion of preferences. This paper empirically tests the role of immigrants in promoting international trade and proposes a new channel through which immigrants might affect the import demand in the host countries.

From a theoretical perspective, the factor content model of trade predicts substitutability between trade and migration: bilateral trade, by weakening wage inequalities across countries, is expected to reduce the incentive for bilateral migration. However, there is large empirical literature suggesting a positive correlation between migration and bilateral trade (Herarder and Saavedra 2005; Felbermayr and Toubal 2012; Felbermayr and Jung 2009). Two main arguments have been used to explain such a positive link. First, international migrants (especially if high-skilled) provide additional information on their origin country and reduce the bilateral cost of trade. This stimulates the exports of the host country towards the origin country of immigrants. In this regard, migrants help domestic firms overcome cultural barriers to trade (language, local taste of consumers, etc.) and create international business relationships (Combes, Lafourcade and Mayer 2005; Herarder and Saavedra 2005; Rauch and Trindade 2002). Second, immigrants

[1] The views expressed are purely those of the authors and may not under any circumstances be regarded as stating an official position of the institutions they are affiliated to.

*Department of Economics, Jardim Cidade Universitaria, Joao Pessoa - PB, Brazil, ZIP Code: 58.051-900 (eafigueiredo@gmail.com)

†Department of Economics, 527A, Stokely Management Center, Knoxville, Tensessee, 37996 (llima@utk.edu)

‡113 rue de Grenelle - 75007 Paris (France) (gianluca.orefice@cepii.fr)

have preferences for the consumption of goods/varieties produced in their own country of origin (home bias in consumption basket). The home bias in consumption choice increases the import demand for goods and varieties coming from the origin country of immigrants (Gould 1994).

The existing literature strongly supports the positive effect of immigrants on the imports of the host country. Using a gravity-type equation Gould (1994) and Head and Ries (1998) find a positive effect of migrant groups in USA and Canada on bilateral trade between the USA or Canada and the home countries of migrants. Girma and Yu (2002) investigate the link between immigration and trade using UK data. They find that immigration from non-Commonwealth countries has a significant export-enhancing effect. Moreover, they find a pro-imports effect of immigration from the non-Commonwealth countries. Rauch and Trindade (2002) show that the presence of Chinese groups of migrants highly stimulated the bilateral trade between China and the host countries. Peri and Requena-Silvente (2010) adopt an instrumental variable strategy on trade transactions data for Spanish provinces over the period 1995-2008, and isolate the export creation effect of new immigrants. In Peri and Requena-Silvente (2010) there is also a first attempt in differentiating the pro-trade effect of migration across different types of goods (i.e. homogeneous, moderately differentiated and differentiated goods). They find that the pro-trade effect is almost entirely due to an increase in the extensive margin and that the effect is somewhat stronger for differentiated goods. Recently, Felbermayr and Toubal (2012) propose an identification strategy able to disentangle the information and the home-country preference channels in the migration-trade nexus, finding that the preference channel is as important as the information channel in stimulating bilateral trade.

The contribution of this paper to the existing literature is twofold. First, we go deeper in estimating the preference effect of immigrants on import demand by using the quantile imputation method as a solution to the presence of many zeros in migration flows (which would lead to missing values after the log transformation). We show that the proposed solutions adopted by the literature[2] such as simply omitting observations from the analysis or adding one to the explanatory variable, bring a substantial bias in the estimations. Our strategy is able to solve this problem and it is also robust to several identifications issues of the gravity models. Our method also addresses the endogeneity of immigrants localization across destinations using a shift-share instrument *à la* Card (2001). Second, we propose a new channel through which migrants might affect the import demand of the host country. In migrating from origin to destination country,

[2]For example, Dunlevy (2006), Briant, Combes and Lafourcade (2014), Giovannetti and Lanati (2014), among others.

migrants observe a change in both the number of available varieties and the price of the bundle of consumable goods: imported varieties that were prohibitively protected in their origin country became available for consumption in the host country because produced locally at cheaper prices or simply imported by the host country at a lower cost (i.e. tariff).[3] To test this channel, we estimate the import demand effect of migrant groups coming from *third* high (tariff) protected countries. We refer to this last channel as "third country effect" in trade-migration nexus. We show that migration decision mirrors a reduction in the applied tariff for the consumption bundle of migrants: emigration towards less (tariff) protected countries allows the consumption of products that were prohibitively protected in the origin countries of migrants. The identification of this channel improves the comprehension of the effect of immigrants on the local import flows. Business and social networks (preference channel and information cost) represent only a part of the overall effect of immigrants on trade: also a more general import demand effect ("third country effect") of immigrants in the host country characterizes the trade-creating effect of migration.

The identification of the "third country effect" has also interesting consequences for the welfare of migrants in the destination country. By relying on a standard gravity model with CES utility function and elasticity of substitution larger than one (see section 2), welfare gain occurs when the price of goods decreases, substitution among goods is allowed, and the variety of goods available for consumption increases. In our framework, when a migrant arrives in a less tariff protected country, all these three conditions hold and the migrant experiences consistent welfare gain.

We test the previous channel using migration data in 19 OECD countries in years 1996, 1999, 2002 and 2005. Three main results qualify our paper. First, we do find a positive effect of migrants on the bilateral import flows between host and origin country of immigrants (confirming the validity of the business and social network effect of immigrants). By instrumenting the flow of migrants we can also conclude on the causal relationship between immigration and trade flows. Second, we find that immigrants coming from high (tariff) protected countries increase the import demand of the host country ("third country effect"). Finally, we show that the previous effect is magnified for high-quality products.

[3]Here we simply assume a Dixit-Stiglitz love for variety utility function. So, when the migrant moves from highly protected country (where the consumption of some varieties is prohibitive and the Inada conditions produce an infinite increase in the marginal utility from their consumption) to a less protected destination country, he/she will love consuming a wider range of varieties (imports).

The rest of the paper is organized as follows. In the next section we sketch the theoretical framework underlying estimations. Then, in section 3 we describe our dataset and econometric strategy. In section 4 we present our results and the last section concludes the paper.

1. Theoretical framework

This section presents the theoretical framework behind the gravity-style estimations reported in the rest of the paper. We base our model on the monopolistic competition model for trade as developed by Combes, Lafourcade and Mayer (2005). We slightly modify this model in order to account for the (new) "third country effect" in trade-migration nexus: a new channel capturing the preferences of migrants for those products being prohibitively protected by tariffs in their origin country.

The utility of a representative consumer residing in country i depends on total consumption of good s produced in any country j (y_{ijs}) with constant elasticity of substitution (CES) across varieties, σ.[4] To account for the preference channel, the consumption of a specific variety is weighted by a parameter a_{ij}, which indicates the preference of a specific consumer with respect to varieties produced in country j and imported in i (see Combes, Lafourcade and Mayer 2005). If we consider for simplicity a single sector, we can omit product subscript s, and the imports of country i from country j, y_{ij}, is:

$$y_{ij} = y_i P_i^{\sigma} n_j p_j^{-\sigma} a_{ij}^{\sigma-1} (1 + \delta_{ij})^{-\sigma}. \tag{1}$$

This is the standard import demand function obtained with the CES monopolistic competition model (standard micro-foundation of a gravity model). y_i is the total consumption in country i, n_j is the total number of firms in country j, P_i is the price index in country i, p_j the mill price in j, and δ_{ij} is the iceberg trade cost. Bilateral imports positively depend on export capacity of country j (size and international competitiveness) and on the price index in country i. Importantly, the level of imports of country i from j is positively affected by the preference parameter a_{ij}.[5] Being interested in the preference channel, this is a crucial parameter deserving deeper explanations. We follow Combes, Lafourcade and Mayer (2005) and assume that a_{ij} can be represented as

[4] In this paper a variety is defined as a specific product-origin country combination.

[5] The price index and the mill price are not observable so, in the empirical estimations we will capture them by including country-by-year fixed effects.

follows:

$$a_{ij} = m_{ij}^{\alpha} \exp(e_{ij}), \tag{2}$$

where, m_{ij} is the flow of immigrants residing in i and coming from j and e_{ij} is the random component. The model assumes that $\alpha > 0$, thus, through the preference channel, immigrants can increase trade between i and j. Indeed, previous empirical works have used the observed distribution of international migrants to identify the preference parameter (see Combes, Lafourcade and Mayer 2005). This is the traditional structure for home bias in consumption choice. Migrants bring their preference for home products when they migrate to destination countries. So far we simply followed existing models in the literature.

Our contribution relies on a further channel through which immigrants residing in country i can boost bilateral imports y_{ij}. In migrating from origin to destination country, migrants observe a change in the prices of the bundle of consumable goods: imported goods, prohibitively protected by tariffs in their origin country, became accessible for consumption in the host country because produced locally and/or simply because imported at a lower tariff. Specifically, we want to show whether the flow of immigrants residing in i, and coming from countries having high tariff protection towards country j, can stimulates trade between i and j. We refer to this new channel in the trade-migration nexus literature as the "third country effect", as migrants coming from third country $(z \neq j)$ can boost bilateral imports y_{ij}. A similar channel has been explored by Firsin (2016) in the context of "proximity" measures based on geographical factors. The third country effect based on import tariff has not yet been explored (to the best of our knowledge).

To formalize this hypothesis, let us consider a pair of countries, i and j, with country i applying tariff τ_{ij} on a specific product coming from j. We can define a set of countries $(k \neq j)$ whose applied tariff to country j is higher than the that applied by country i to j: $\Omega_k = \{(k) : \text{tariff}_{kj} \geq \text{tariff}_{ij}\}$, for $k = 0, 1, 2, ..., K$. For each importer country i, we can compute the total amount of immigrants coming from third countries (k) in Ω_k as: $sm_i = \sum_{k=0}^{K} m_{ik}$. Thus, sm_i will represent the number of immigrants living in i but coming from countries imposing higher tariffs levels than i on products coming from j. According to our hypotheses, immigrants coming from countries in Ω_k can increase the imports of i from j; so we modified (2) to account for this new channel:

$$a_{ij} = sm_i^{\gamma} m_{ij}^{\alpha} \exp(e_{ij}). \tag{3}$$

The new parameter, γ, measures the "third country effect" in the trade-migration nexus. If our argument is valid, it is expected to be positive. For instance, considering the US imports from Germany, sm_i could represent the total amount of immigrants in the US coming from Latin American countries (where tariffs on German imports are, on average, higher than US tariffs on products made in Germany).

In what follows, we estimate in turn the standard preference channel in the trade-migration nexus (with a proper Instrumental Variable approach to assess the causal relationship), and then we move to the estimation of the third country effect as highlighted in the theoretical framework reported above.

2. Econometric issues and data

This section provides a preliminary discussion of the econometric issues in estimating the preference and third country effect in the trade-migration nexus. By combining equations (1) and (2) the standard preference channel can be represented (and then estimated) by a simple stochastic version of a gravity model:

$$y_{ijs,t} = \exp\left(\beta_1 \ln m_{ij,t} + \beta_2 \ln x_{ijs,t}\right) \eta_{ijs,t}, \tag{4}$$

where the subscript s for product has been reintroduced and $y_{ijs,t}$ represents the imports of country i from country j for product s at time t; and $m_{ij,t}$ is the flow of immigrants in i coming from j at time t. The vector $x_{ijs,t}$ is the vector of covariates traditionally associated with the bilateral trade flow $y_{ijs,t}$ in a gravity model. The vector $x_{ijs,t}$ thus includes: bilateral tariff, distance, contiguity, and country-year fixed effects capturing the multilateral resistance terms (as in Anderson and van Wincoop 2003). Finally, $\eta_{ijs,t}$ is a non-negative random variable such that $E\left(\eta_{ijt,s} | m_{ij,t}, x_{ijs,t}\right) = 1$. Model (4) is nonlinear in parameters and therefore cannot be estimated with ordinary least squares (OLS), but can be estimated by the Pseudo Poisson Maximum Likeliood estimator proposed by Santos Silva and Tenreyro (2006). Another identification problem emerges if we take the log of immigrants. Indeed, when several observations in $m_{ij,t}$ are equal to zero, the log-transformation leads to lots of missing observations. The literature has proposed two solutions to this type of problem. The first consists of omitting the missing observations. But as shown in Abrevaya and Donald (2011), simply omitting the observations from the analysis brings a substantial bias in the estimations. The second solution, widely adopted by most of the studies, is to consider $\ln(m_{ij,t} + 1)$ instead of $\ln(m_{ij,t})$ in equation (4). However, this last solution also induces a bias in the migration elasticity. In other words, from equation (4) β_1 is an approximation to $\frac{m}{y}\frac{dy}{dm}$, which is the migration elasticity. However, if one considers $\ln(m_{ij,t}+1)$, then we obtain $\beta_1^* = \frac{(m+1)}{y}\frac{dy}{dm} > \beta_1$. Thus, the solution largely employed in the literature will ultimately bias the migration elasticity upwards.

The problem of missing observations of $\ln(m_i)$ can be solved by using the so called imputation techniques as, for instance, the GMM imputation (IM-GMM) method developed by Abrevaya and Donald (2011). In order to apply the IM-GMM method, one needs to consider the linear version of (4), which is obtained by taking logarithms of both sides of the equation, that is:

$$\ln y_{ijs,t} = \beta_1 \ln m_{ij,t} + \beta_2 \ln x_{ijs,t} + \ln \eta_{ijs,t}, \tag{5}$$

where $\ln y_{ijs,t}$ is now defined on the real line $\mathbb{R}$ and $y_{ijs,t}$ is assumed to be positive as it is the case for the database used in this paper. However, the empirical model based on (5) is subject to the log-linearization bias caused by the presence of heteroskedasticity in $\ln \eta_{ijs,t}$ as pointed out by Santos Silva and Tenreyro (2006), which is a direct consequence of the Jensen's inequality. To address this problem, Figueiredo, Lima and Schaur (2016) proposed using quantile regression to estimate model (5). The idea is that, unlike the mean function, the quantile function is not subject to the Jensen's inequatily because quantiles are invariant to monotone transformations. In other words, if $h(\cdot)$ is a nondecreasing function on $\mathbb{R}$, then for any random variable Y, $Q_\tau(h(Y)) = h(Q_\tau(Y))$,where $Q_\tau(\cdot)$ is the $\tau - th$ quantile function. Based on this property, they show that identification of the exponential model (4) leads to identification of the log-linear model (5) and vice-versa, even under the presence of heteroskedasticity and without assuming any knowledge about the distribution function of $\eta_{ijs,t}$.

The PPML estimator proposed by Santos Silva and Tenreyro (2006) is able to address the problem of heteroskedasticity in the error term, but does not solve the bias due to the presence of zeros in bilateral migration data (as explanatory variable).Thus, in this paper we eliminate the bias caused by missing observations of $m_{ij,t}$ and log-linearization of (4) by using quantile regression with imputation methods. In particular, we apply the quantile imputation method (IM-QR) proposed by Wei, Ma and Carroll (2012) to estimate the following quantile model:

$$Q_\tau(\ln y_{ijs,t}|\ln m_{ij,t}, \ln x_{ijs,t}) = \ln m_{ij,t}\beta_1(\tau) + \beta_2(\tau) \ln x_{ijs,t}. \tag{6}$$

where, $\ln x_{ijs,t}$ is always observed and contains the intercept, but the (log of) migration $\ln m_{ij,t}$

may be missing. We assume that n is the total sample size, and n_1 of these n observations are complete, while the remaining n_0 of them are missing. The main goal is to estimate the regression parameter $\beta(\tau) = (\beta_1(\tau), \beta_2^T(\tau))^T$. The estimating algorithm is completely described in the appendix of this paper. We estimated $\beta(\tau)$ for the three most representative quartiles, i.e., $\tau = 0.25; 0.50; 0.75$ so that we can capture the effect of migration on low trade volume ($\tau = 0.25$), median ($\tau = 0.50$) and high trade volumes ($\tau = 0.75$).

2.1. Data

In our analysis we consider a panel data set with 177 countries of origin and 19 OECD destination countries for the years of 1996, 1999, 2002 and 2005.[6] Both, origin and destination countries, are summarized in Table 1. The bilateral trade flows are represented by the values of nominal imports in thousand dollars disaggregated at HS 4-digit level from BACI database (CEPII) and are always positive in this database.[7]

Table 1 about here

Migration data are from the International Migration Database (IMD) gathered by the OECD. We consider the total number of "inflows of foreign population by nationality", resident in one of the 19 OECD destination countries and born in one of the 177 countries of origin. The migration data has 17% of observations equal to zero.

Tariff data are from World Integrated Trade Solution (WITS), span from 1995 to 2005, and refer to the import tariff applied by OECD countries with respect to all 177 origin countries at HS4 level. Other control variables are from the standard sources. Data on distances, common border, language and colonial link are from CEPII.[8] Table 2 summarizes the descriptive statistics.

Table 2 about here

[6] We use repeated cross section approach in order to keep a manageable dataset considered the need for huge set of fixed effects in the estimation.
[7] We dropped agricultural sectors.
[8] Available here `http://www.cepii.fr/CEPII/fr/bdd_modele/presentation.asp?id=6`.

3. Estimation results

3.1. The business and social networks effect in trade-migration nexus

By combining equations (1) and (2) and taking the log-transformation, we can derive a fixed-effects specification fully consistent with the theoretical framework (as done by Combes, Lafourcade and Mayer 2005). All the unobservable country-year specific factors affecting the import demand (such as P_i and p_j) are captured by two sets of fixed effects: country i-year and country j-year fixed effects. Then, we can apply a standard gravity specification as in Anderson and van Wincoop (2003) with multilateral resistance terms identified by fixed effects. Bilateral trade costs δ_{ij} are approximated by country-pair specific variables affecting bilateral trade (as tariffs, distance, common language, border and colonial relationship).[9] Thus, our baseline empirical equation is thus the following:

$$\ln y_{ijs,t} = \beta_1 \ln m_{ij,t} + \beta_2 \ln x_{ijs,t} + \theta_{st} + \theta_{it} + \theta_{jt} + \ln \eta_{ijs,t}. \tag{7}$$

As discussed above, the dependent variable is the import value of country i from j in sector s and time t; m_{ijt} is the flow of immigrants coming from j but residing in i. The set of control variables x_{ijst} includes: (i) bilateral distance, (ii) common language dummy, (iii) common border dummy and (iv) the tariff applied by country i on imports from j in sector s. In equation (7) we include a set of sector-by-year fixed effects (θ_{st}) controlling for any sector specific shock (i.e. technology and/or productivity shocks). We also include country-by-year fixed effect (θ_{it} and θ_{jt}). These two sets of fixed effects control for several country specific factors affecting bilateral trade: (a) market capacity and supply capacity of country i and j, (b) multilateral resistance term (Anderson and van Wincoop 2003); (c) price index and mill price (as indicated in equation 1) and; (d) for country specific business cycle.[10]

We estimate equation (7) by using three estimators. We start by using the Poisson Pseudo Maximum Likelihood estimator as suggested by Santos Silva and Tenreyro (2006). The PPML estimator is applied to a non-linear version of the baseline model (7), as reported in equation (4), in which we use $\ln(m_{ijs} + 1)$ to keep country-pair with zero migrants' flow. As discussed

[9] However, in a final robustness check we also run a specification with country-pair fixed effects capturing all the bilateral specific factors affecting trade.

[10] We are aware that the first-best strategy to control for the multilateral resistance term in country pair-sector specific gravity equation would be including country-sector-year fixed effects. However this has not been possible for computational reasons.

above, this solution might produce biased estimations. So we estimate equation (7) using two estimators based on the GMM imputation (IM-GMM) method proposed by Abrevaya and Donald (2011), and the quantile imputation regression (IM-QR) suggested by Wei, Ma and Carroll (2012). Contrary to PPML, these estimators use imputed data instead of adding one to m_{ijs}. The IM-QR considered a grid of $K_n = 19$ points, i.e., $\tau = 0.05, 0.10, 0.15, ..., 0.90, 0.95$, $M = 10$ repetitions (see appendix), and compute standard errors by accounting for cluster errors by importer and exporter countries (Parente and Santos Silva 2016).

The main reason for using the quantile imputation regression (IM-QR) is that if the identification condition of the exponential model (4) is valid, $E\left(\eta_{ijs,t} | \ln m_{ij,t}, \ln x_{ijs,t}\right) = 1$, then: a) the PPML estimator will suffer from bias from the transformation in the migration variable; b) the IM-GMM will suffer from the log-linearization bias but will fix the bias caused by using $\ln(m_{ij,t} + 1)$ and; c) the IM-QR will avoid biases from both sources. However, it is important to notice that our quantile estimator is not identifying the average effect of m and x on y, $E\left[y|m,x\right]$, but rather the effect of m and x on the quantiles (entire distribution) of y, $Q_\tau(y|m,x)$.[11]

The further advantage of the IM-QR method is that we can capture the (potential) asymmetric effect of migration across quantiles in the import flows distribution. We can test whether migration has a small effect on low trade volumes but a large effect on big trade volumes (or vice-versa). Finally, likewise the mean effect $E\left[y|m,x\right]$, the median effect $Q_{\tau=0.50}(y|m,x)$ can also be used to represent the most likely effect of migration on trade. For these reasons, the quantile method with imputed observations is much more general than the methods that are only based on the mean effect $E\left[y|m,x\right]$.

3.1.1. Baseline results

Table 3 shows baseline regression results. As expected (and coherently with the existing literature) the presence of migrants from country j stimulates the imports of i coming from j. This result is robust across all the estimators we use. Regarding the effect of immigration on trade, our initial expectations were that PPML, IM-GMM and IM-QR estimators would be significantly different from each other.[12] As shown in the section 3, the transformation in the migration varaible, $\ln(m_{ij,t} + 1)$, bias the migration elasticity upwards. In fact, PPML shows the highest value compared to IM-GMM and IM-QR at the median. On the other hand, the difference between IM-GMM and IM-QR at the median may represent the log-linearization bias.

[11] Note that by the equivariance property $Q_\tau(y|\ln m, \ln x)$ can be directly recovered from $Q_\tau(\ln y|\ln m, \ln x)$.

[12] Emphasizing that our quantile model does not identify the average effect of $m_{ij,t}$ on $y_{ijs,t}$.

Using our preferred estimator, IM-QR at the median, we find that a 10% increase in the flow of migrants coming from j increases the imports of i from j by the 0.9% - which is equivalent to a 0.75% tariff reduction. The IM-QR estimator however reveals a significant asymmetric effect: the presence of migrants has stronger effect the larger is the import volume from country j. While for the 25th percentile of the import distribution, a 10% increase in the flow of immigrants leads to a 0.4% increase in imports; for the 75th percentile of the import distribution the effect of a ten percent increase in immigration leads to a 1.3% increase in the imports (i.e. equivalent to a 1% tariff reduction). This is a novel result in the literature because the effect of migration on trade has always been assumed a homogeneous over the distribution of trade flows.

Control variables have the expected sign. Bilateral distance has a negative and significant coefficient with magnitude in line with the existing literature (see Head and Mayer 2014). Common border and language have positive effects on imports, while the applied tariff of i on imports from j has a strong negative effect with magnitude in line with the existing literature (see Berthou and Fontagné 2016). Surprisingly, we find a negative coefficient for common colonial dummy. Such puzzling effect can be caused by the endogeneity of the migration variable. The presence of one endogenous variable across regressors might provide biased estimated coefficients. Indeed, after instrumenting the bilateral migration flows in the next section - i.e. when we use exogenous variation in the bilateral flow of migrants - the coefficient for common colonial dummy turns out positive and in line with the results of the existing literature.

Table 3 about here

3.1.2. Instrumental variables approach

In this context, there are two potential sources of endogeneity for the flow of migrants (see Combes, Lafourcade and Mayer 2005; Peri and Requena-Silvente 2010). First, unobserved productivity shocks in country i might simultaneously affect its import demand and the incentive of people in migrating to country i. This omitted variable concern is strongly reduced by the inclusion of country-by-year fixed effects described above. Second, the large availability of imported products from j, might push migrants (from j) to settle in a specific country i (reverse causality). Both the omitted variable and the reverse causality problem are addressed here by an instrumental variable strategy. We apply an IV control function approach to the quantile regression by following the procedure developed by Chernozhukov and Hansen (2008).[13] The instrumental variable we use here was introduced by Card (2001) and it is largely used by the

[13] See also Lee (2007).

migration and labor market literature. In particular, we instrument the changes in the bilateral flow of migrants across destinations with an imputed bilateral immigrants flow, $z_{ij,t}$, computed as follows:

$$z_{ij,t} = \frac{\text{Immigrants}_{ij,t=1993}}{\sum_j \text{Immigrants}_{ij,t=1993}} \times \sum_j \text{Immigrants}_{ij,t}.$$

Where $\text{Immigrants}_{ij,t=1993}$ is the flow of immigrants in country i coming from j in the year 1993. While rarely applied to the trade and migration literature (exception is Peri and Requena-Silvente 2010), this instrumental variable is common in the literature analyzing the labor market effect of immigrants (Card 2001, Ottaviano and Peri 2006, Card 2009). Basically, we use the distribution of immigrants by origins in 1993 for a given destination i, and attribute to each migration group (in each destination) the net growth of immigrants in country i. The underlying intuition is that immigrants tend to settle where other persons of the same nationality already live. This makes the instrument a good predictor for the actual bilateral migration flows. Moreover, being based on the distribution of immigrants across origins as of 1993, the constructed flows are not affected by any origin-destination-specific demand shocks. The first stage estimation of our IV strategy will be the following:

$$\ln m_{ij,t} = \alpha_1 \ln(z_{ij,t}) + \alpha_2 x_{ij,t} + v_{ij,t}, \tag{8}$$

with $\widehat{v_{ij,t}}$ being the control function of our IV approach (i.e. the residual value obtained from a quantile estimator of (8)). In the second step we will proceed with the quantile estimation of equation (7). Notice that in moving from IM-QR to IV estimations we lose an important share of observations due to missing values in bilateral migration flows in 1993 (see descriptive evidence about bilateral flow of migrants in Table 4).[14]

Table 4 about here

Results for the first stage estimation are reported in Table (5). As expected the imputed flow of immigrants (our instrumental variable) is positively and highly significant in explaining the actual flow of migrants (endogenous variable) with coefficient of 0.900 when $\tau = 0.5$.

[14] Our method allows us to recover the missing observations of $\log(m_{ij,t})$ caused by the presence of migration flows ($m_{ij,t}$) equal to zero, but it does not allow us to recover missing observations of migration flows because it would require an imputation method valid to non-linear quantile regression models which is out of the scope of this paper.

Our instrumental variable can thus be considered relevant. The validity of the instrument (i.e. orthogonality with respect to bilateral imports) cannot be tested with an overidentification test (Sargan test) because our model is exactly identified. However, our instrumental variable uses the distribution of immigrants by origin country in 1993 to allocate the total inflow of immigrants in each destination at time t. This approach is based on the fact that, because of information and preferences, new immigrants tend to move to the same destination country where previous immigrants from the same country already live. This is because they know about opportunities in those locations from the network of immigrants, and because they enjoy the amenities of living with their co-nationals. These reasons to co-locate are driven by preference and information, not by demand shocks. For these reasons our instrument can be considered exogenous with respect to imports of country i from j. In details, our exclusion restriction is that the distribution of immigrants across destinations for each country of origin in 1993 is uncorrelated with the distribution of changes in labor demand shocks (which can be eventually correlated with import demand shock) in the destination from 1996 - conditional on country-year and sector-year fixed effects.

Table 5 about here

Second stage results reported in Table (6) confirm our baseline results: the presence of migrants coming from origin j in destination i stimulates the imports of country i from j. Control variables have (again) the expected sign with colonial relationship dummy turning to a "plausible" positive coefficient.

Table 6 about here

3.2. The third country effect in trade-migration nexus

3.3. The third country effect in trade-migration nexus

So far we use new econometric approach (IV on IM-QR) to confirm what existing literature already found: a positive effect of migrants on bilateral imports. Now we add to the existing literature by testing what we called (in section 2) the "third country effect" in trade-migration nexus. Is it possible that immigrants coming from a third country ($k \neq j$) have positive effects on bilateral imports (y_{ij})? This is the question addressed in this section. The economic rationale is the following. Migrating from country j to i implies (also) a sudden change in the price of consumable goods for immigrants. When the migrant arrives in a less protected destination country he/she immediately copes with a new (wider) set of varieties for consumption (varieties that were prohibitively protected at origin countries become available for consumption). Therefore, assuming Dixit-Stiglitz love for variety utility function, he/she will consume a wider set of varieties (imported goods). For example, Argentine, Brazilian and Chilean immigrants in USA may boost USA imports from Germany for those German products prohibitively protected in such South American countries.

The third country effect of immigration on trade described above can be represented by combining equations (1) and (3). In order to be coherent with the theoretical framework described above, for each product s we have to define the set of third countries ($k \neq j$) whose import tariff imposed to j is higher than the tariff imposed by i on j. Specifically: $\Omega_{ks,t} = \{(k) : \text{tariff}_{kjs,t} \geq \text{tariff}_{ijs,t}\}$, for $k = 0, 1, 2, ..., K$. Having the set of countries k with higher tariff protection than i (on j's exports), we can compute the total amount of immigrants coming from such countries ($\Omega_{ks,t}$) and residing in i as: $sm_i = \sum_{k \subset \Omega_{ks,t}} m_{ik}$.

Therefore, we augment the empirical specification in (7) by including the amount of immigrants coming from third countries ($k \subset \Omega_{ks,t}$). So we consider the following econometric specification:

$$\ln y_{ijs,t} = \beta_1 \ln m_{ij,t} + \beta_2 \ln sm_{it} + \beta_3 \ln x_{ijs,t} + \theta_{st} + \theta_{it} + \theta_{jt} + \ln \eta_{ijs,t}, \tag{9}$$

where, sm_{it} represents the number of immigrants living in i from countries belonging to the set $\Omega_{ks,t}$.[15] Other variables have the same meaning as before. Therefore, β_2 will represent the "third country effect" of immigrants on import flows from j to i.

We adopted the IV procedure proposed in the previous subsection (with $m_{ij,t}$ instrumented by imputed immigration flows as described in the previous section). Estimation results for equation (9) are reported in Tables 7 and 8 (first and second steps). In general, the home bias effect of immigration as well as the control variables have a similar behavior to those obtained in the previous subsection. Interestingly, we find a positive and significant coefficient for β_2. This means that the number of immigrants residing in i and coming from third countries ($k \subset \Omega_{ks,t}$) stimulate the bilateral import demand y_{ij}. The set of third countries ($k \subset \Omega_{ks,t}$) is product specific and includes countries with higher tariff protection than i on products coming from j. So, we may conclude that immigrants stimulate the import demand for products being highly

[15]Notice that the amount of migrants coming from high protected countries (sm_{it}) is sector specific, so we can keep country-year fixed effects in our estimations.

protected by tariffs in their origin country. This result is robust across all quantiles, the parameter related to third country effect of immigration is always positive and significant.

Table 7 about here

Table 8 about here

3.3.1. Placebo test

A potential concern is that other factors might drive the third country effect of immigrants. So, in this section we conduct a placebo test (i.e. falsification test) aiming at excluding such concern. Instead of keeping immigrants coming from countries with high protection level ($k \subset \Omega_{ks,t}$), we keep its complement - i.e. immigrants coming from countries having lower import tariff than i on $j's$ exports in sector s ($\Omega^{-}_{ks,t} = \{(k) : \text{tariff}_{kjs,t} < \text{tariff}_{ijs,t}\}$, for $k = 0, 1, 2, ..., K$).

The idea is that immigrants coming from less protected countries do not face any reduction in the price of consumption bundle, so they should not affect the import demand at destination (null effect on bilateral imports y_{ijst}). Results reported in tables (9) and (10) support our story: immigrants coming from less protected countries do not affect import demand at destination. This result reinforces our argument in favor of a "third country effect" of migration.

Table 9 about here

Table 10 about here

3.3.2. Results by quality of products

If our economic intuition in correct, the channel highlighted above should be stronger for high quality products. Indeed, high quality imported products are often prohibitively protected in the country of origin of migrants (i.e. developing countries). In Figure 1 we show that the average tariff applied by migrants' countries of origin (developing countries) on high quality products (top-20% ladder) is significantly higher than the average tariff applied on low-quality products (bottom 20% ladder). So, in migrating from high to low protected country, immigrants will find high-quality products cheaper than in their origin country. We test this hypotheses by using the Khandelwal (2010) data on the quality ladders of products and replicate equation (9) by increasing product quality levels. Results reported in Table (12) support this intuition. The third country effect of immigration gets bigger the higher is the quality level of the products.

Figure 1 about here

Table 11 about here

Table 12 about here

For products belonging to the bottom 20% of the quality distribution (*low quality products*), the standard business and social network channel is definitely more important than the *third country effect*, i.e. $\beta_1 > \beta_2$; while for high-quality products (on average more protected by tariffs in developing countries - origin of immigrants) the third country effect gets bigger and is almost equal to the standard business and social network effect.

3.3.3. Robustness using country pair fixed effects

In our baseline estimations, country-pair specific factors affecting bilateral import demand have been approximated by geographical variables (as distance, common border, language and colonial relationship). This is a common practice in gravity estimations for trade. However, it might be the case that some (unobserved) country pair specific variables affecting both bilateral imports and migration flows are omitted. The omission of such variables is expected to introduce a bias in the estimations reported above. In order to address such concern, in this section we report estimation results for equation (9) also including country-pair fixed effects.

The inclusion of a further set of fixed effect can be computationally intensive because we need to estimate a large amount of coefficients on country-pair (and country-year) fixed effects. In order to increase computational efficiency, we employ the method proposed by Canay (2011) which eliminates the country-pair fixed effects beforehand, making the implementation of the estimator computationally feasible regardless of the number of country pair fixed effects in the analysis.[16] Results for this robustness check are reported in Tables 13 and 14 and confirm what we have discussed above.

Table 13 about here

Table 14 about here

[16]The results should be analyzed with caution due to the presence of the incidental parameter problem. For details, see Canay (2011).

4. Concluding Remarks

This paper quantifies the trade-creating effect of international migrants and identifies a new channel through which immigrants coming from third high (tariff) protected countries can affect bilateral specific import flows: the *third country effect*. In migrating from a high to low tariff protected country, migrants experience both an increase in the number of available varieties and a reduction in the price for imported products: goods being highly protected in their country of origin become cheaper in their destination country because less protected by tariffs. For this reason, migrants coming from highly protected countries might boost bilateral imports. We find overwhelming evidence of this channel. According to our preferred estimations, a ten percent increase in the flow of bilateral migrants increases bilateral imports by 0.45% (*preference or information channel*). Similarly, a ten percent increase in the flow of migrants coming from third high (tariff) protected countries increases bilateral imports by 0.27% (*third country effect*). These effects might appear negligible, but they are not. Indeed, in tariff equivalent terms, the preference channel corresponds to a 0.70% tariff reduction while the third country effect corresponds to a 0.43% tariff reduction.

We also contribute to the existing literature by estimating the standard business and social networks effect in the trade-migration nexus by adopting fresh econometric techniques able to address both the problem of zeros in migration flows and endogeneity of immigrants' settlement (IV on IM-QR). Furthermore, we adapt such econometric technique to an instrumental variable approach. By using the shift-share instrument for migration (as in Card 2001), we solve the endogeneity problem in migration settlement and provide a causal interpretation for the impact of migration on bilateral trade.

5. References

Anderson, James and Eric van Wincoop, Gravity with gravitas: a solution to the border puzzle, *American Economic Review*, 93 (2003), 170-192.

Abrevaya, J. and Donald, S. (2011). *A GMM approach for dealing with missing data on regressors and instruments*, mimeo.

Berthou, A. and Fontagné, L. (2016). Variable trade costs, composition effects, and the intensive margin of trade, *The World Economy*, 39(1), 54-71.

Briant, A., Combes, P. and Lafourcade, M. (2014). Product complexity, quality of institutions and the pro-trade effect of immigrants. *The World Economy* 37(1), 63-85.

Canay, I. (2011). A simple approach to quantile regression for panel data, *The Econometrics Journal* 14(3), 368-386.

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

Card, D. (2009). Immigration and inequality, *American Economic Review*, 99(2), 1-21.

Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: a robust inference approach, *Journal of Econometrics*, 142(1), 379-398.

Combes, P., Lafourcade, P. and Mayer, T. (2005). The trade-creating effects of business and social networks: evidence from France, *Journal of International Economics*, 66(2), 1-29.

Dunlevy, J. (2006). The influence of corruption and language on the protrade effect of immigrants: evidence from the american states, *The Review of Economics and Statistics*, 88(1),182-186.

Felbermayr, G. and Jung, B. (2009). The pro-trade effect of the brain drain: sorting out confounding factors, *Economics Letters*, 104(2), 72-75.

Felbermayr, G. and Toubal, F. (2012). Revisiting the trade-migration nexus: Evidence from new OECD data. *World Development*, 40(5), 928-937.

Figueiredo, E., Lima, L. and Schaur, G. (2016). The effect of the Euro on the bilateral trade distribution, *Empirical Economics*, 50(1), 17-29.

Firsin, O. (2016). Do immigrants promote exports to third party countries? On the role of geographic and linguistic proximity, *Midwest International Trade Meetings*, Spring 2016.

Giovannetti, G. and Lanati, M. (2014). *Do high-skill immigrants trigger high quality trade?*, mimeo.

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

Gould, D. (1994). Immigrant links to the home country: empirical implications for US bilateral trade flows, *The Review of Economics and Statistics*, 76(2), 302-316

Head, K. and Ries, J. (1998). Immigration and trade creation: econometric evidence from Canada, *Canadian Journal of Economics*, 31(1), 47-62.

Head, K. and Mayer, T. (2014). Gravity equations: workhorse, toolkit, and cookbook, chapter to appear in the Handbook of International Economics Vol. 4, eds. Gopinath, Helpman, and Rogoff, Elsevier.

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

Khandelwal, A. (2010). The long and short (of) quality ladders, *Review of Economic Studies*, 77(4), 1450-1476.

Lee, S. (2007). Endogeneity in quantile regression models: a control function approach, *Journal of Econometrics*, 141(2), 1131-1158.

Ottaviano, G. and Peri, G. (2006). The economic value of cultural diversity: evidence from U.S. cities. *Journal of Economic Geography*, 6(1), 9-44.

Parente, P. and Santos Silva, J. (2016). Quantile regression with clustered data, *Journal of Econometric Methods*, 5(1), 1-15.

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

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

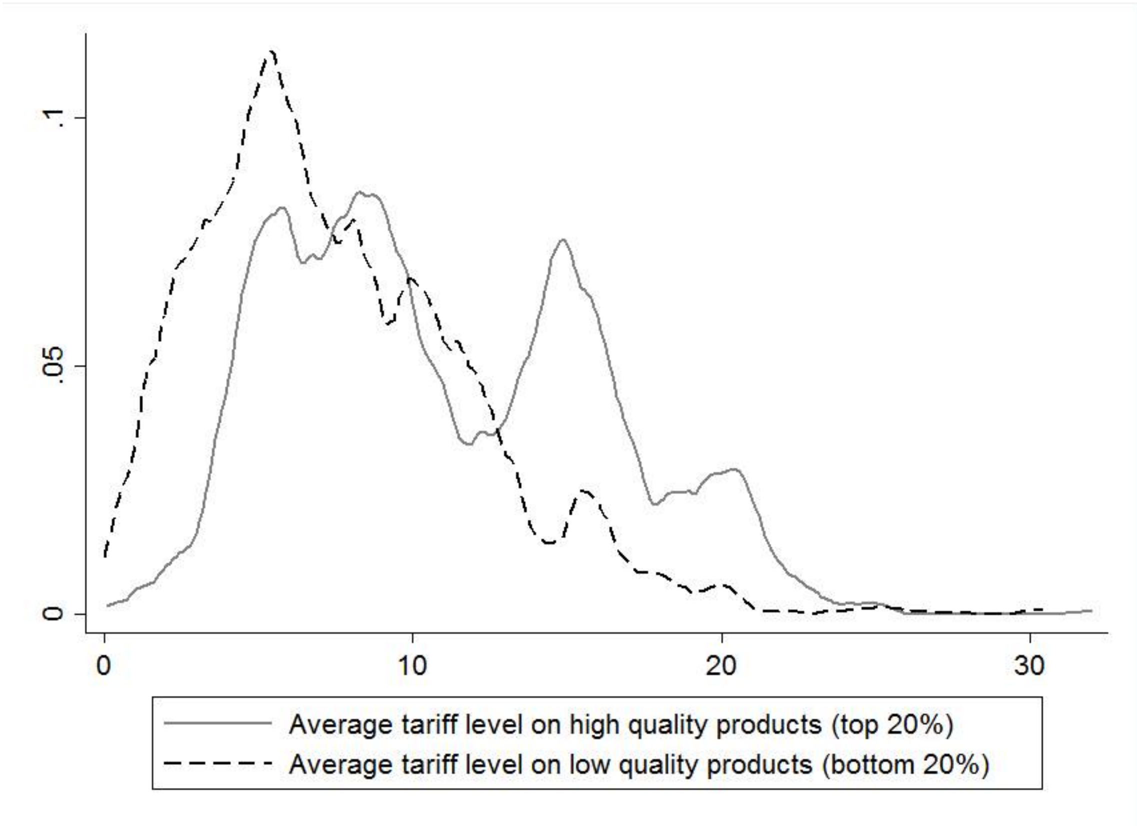
Rauch, J. and Trindade, V. (2002). Ethnic Chinese networks in international trade, *The Review of Economics and Statistics*, 84(1), 116-130.

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

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

Wei, Y., Ma, Y. and Carroll, R. (2012). Multiple imputation in quantile regression, *Biometrika*, 99(2), 423-438.

Figure 1 – Average tariff levels imposed by the country of origin of immigrants by quality ladder.



Source: Authors' calculation on WITS data. Quality ladder data are from Khandelwal (2010).

Table 1 – List of countries

Afghanistan	Denmark*	Laos	Saint Vicent
Albania	Djibouti	Latvia	Samoa
Algeria	Dominica	Lebanon	San Marino
Angola	Dominica	Liberia	Sao Tome and Principe
Antigua	Ecuador	Libya	Saudi Arabia
Argentina	Egypt	Lithuania	Senegal
Armenia	El Salvador	Macedonia	Seychelles
Australia*	Equatorial Guinea	Madagascar	Sierra Leone
Austria*	Eritrea	Malawi	Slovakia
Azerbaijan	Estonia	Malaysia	Slovenia
Bahamas	Ethiopia	Maldives	Solomon
Bahrain	Fiji	Mali	Somalia
Banglade	Finland*	Malta	South Africa
Barbados	France*	Marshall Islands	Spain*
Belarus	Gabon	Mauritania	Sri Lanka
Belgium*	Gambia	Mauritius	Sudan
Belize	Georgia	Mexico	Suriname
Benin	Germany*	Micronesia	Sweden*
Bhutan	Ghana	Moldova	Switzerland*
Bolivia	Greece*	Mongolia	Syria
Bosnia and Herzegovina	Grenada	Morocco	Tajikistan
Brazil	Guatemala	Mozambique	Tanzania
Brunei	Guinea	Myanmar	Thailand
Bulgaria	Guinea-Bissau	Nepal	Togo
Burkina	Guyana	Netherlands*	Tonga
Burundi	Haiti	New Zealand*	Trinidad and Tobago
Cambodia	Honduras	Nicaragua	Tunisia
Cameroon	Hungary	Niger	Turkey
Canada*	Iceland*	Nigeria	Turkmenistan
Cape Verde	India	Norway*	Tuvalu
Central African Republic	Indonesia	Oman	Uganda
Chad	Iran	Pakistan	Ukraine
Chile*	Iraq	Palau	United Kingdom*
China	Ireland*	Panama	United States*
China, Hong Kong SAR	Israel	Papua New Guinea	Uruguay
China, Macao SAR	Italy*	Paraguay	Uzbekistan
Colombia	Jamaica	Peru	Vanuatu
Comoros	Japan*	Philippinas	Venezuela
Congo, Rep. of the	Jordan	Poland	Vietnam
Costa Rica	Kazakhst	Portugal*	Yemen
Cote d'Ivoire	Kenya	Qatar	Zambia
Croatia	Kiribati	Russia	Zimbabwe
Cuba	Korea	Rwanda	
Cyprus	Kuwait	Saint Kitts and Nevis	
Czech Republic	Kyrgyzst	Saint Lucia	

Note: (*) both origin and destination country.

Table 2 – Descriptive statistics

Variables	Mean	SD	Min	Max
Immigration	5,289.104	16,780.13	0	218,822
Nominal imports	8,848.832	177,964.1	0.0008	3.66e+07
Tariff	3.433	5.579	0	293.33
Ln distance	8.453	1.027	5.322	9.826
Contiguity	0.073	0.260	0	1
Language	0.187	0.390	0	1
Colonial	0.056	0.230	0	1

Table 3 – Pro-trade effects of immigration: PPLM, IM-GMM and IM-QR

			IM-QR		
Variables	PPML	IM-GMM	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
ln Immigration	0.104^a	0.096^a	0.043^a	0.087^a	0.133^a
	(0.012)	(0.013)	(0.002)	(0.002)	(0.002)
ln [(Tariff/100)+1]	-2.370^a	-1.458^a	-1.063^a	-1.193^a	-0.949^a
	(0.378)	(0.086)	(0.077)	(0.079)	(0.082)
ln Distance	-0.318^a	-0.768^a	-0.986^a	-0.888^a	-0.761^a
	(0.027)	(0.039)	(0.005)	(0.004)	(0.004)
Contiguity	1.410^a	1.001^a	0.797^a	0.867^a	0.873^a
	(0.065)	(0.034)	(0.019)	(0.016)	(0.015)
Language	0.366^a	0.437^a	0.484^a	0.330^a	0.182^a
	(0.050)	(0.023)	(0.011)	(0.030)	(0.010)
Colonial	-0.699^a	-0.234^a	-0.000	-0.098^a	-0.291^a
	(0.058)	(0.024)	(0.018)	(0.014)	(0.014)
Country-by-year FE	yes	yes	yes	yes	yes
Sector-by-year FE	yes	yes	yes	yes	yes
Observations	554,251	554,251	554,251	554,251	554,251

Notes: Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 4 – Descriptive statistics: immigration data by year

Variables	Sample	Mean	SD	Min	Max
1993	938	6,949.054	16,371.480	75	141,587
1996	4,081	1,636.261	6,743.530	0	163,556
1999	4,415	1,165.150	5,248.153	0	147,402
2002	4,617	1,597.601	7,242.482	0	218,822
2005	4,626	1,403.378	6,973.104	0	161,445

Table 5 – Instrumental variable: step 1*

Variables	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
ln z	0.892^a	0.900^a	0.834^a
	(0.002)	(0.001)	(0.090)
ln [(Tariff/100)+1]	-0.381^a	-0.169^a	-0.026
	(0.063)	(0.028)	(0.017)
ln Distance	-0.099^a	-0.156^a	-0.047^a
	(0.006)	(0.002)	(0.001)
Contiguity	0.228^a	0.127^a	0.473^a
	(0.012)	(0.003)	(0.061)
Language	-0.164^b	-0.167^a	-0.002
	(0.088)	(0.002)	(0.004)
Colonial	0.302^a	0.046^a	-0.286^a
	(0.007)	(0.003)	(0.003)
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	169,070	169,070	169,070

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 6 – Instrumental variable: step 2

	No IV: IM-QR			IV: IM-QR		
Variables	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
ln Immigration*	0.063^a	0.079^a	0.106^a	0.016^b	0.041^a	0.066^a
	(0.002)	(0.001)	(0.006)	(0.007)	(0.005)	(0.005)
ln [(Tariff/100)+1]	-2.796^a	-1.399^a	-0.898^a	-2.815^a	-1.477^a	-1.043^a
	(0.188)	(0.163)	(0.150)	(0.189)	(0.157)	(0.151)
ln Distance	-0.985^a	$-.884^a$	-0.740^a	-0.992^a	-0.892^a	-0.738^a
	(0.011)	(0.009)	(0.009)	(0.011)	(0.009)	(0.009)
Contiguity	1.125^a	1.211^a	1.182^a	1.151^a	1.229^a	1.237^a
	(0.030)	(0.024)	(0.024)	(0.031)	(0.023)	(0.025)
Language	0.101^a	0.020	-0.048^a	0.119^a	0.030^b	-0.031^b
	(0.017)	(0.015)	(0.016)	(0.018)	(0.015)	(0.016)
Colonial	0.484^a	0.331^a	0.029	0.522^a	0.332^a	0.017
	(0.025)	(0.019)	(0.019)	(0.025)	(0.018)	(0.019)
$\hat{v}$	–	–	–	-0.153^a	-0.140^a	-0.136^a
				(0.013)	(0.012)	(0.012)
Country-by-year FE	yes	yes	yes	yes	yes	yes
Sector-by-year FE	yes	yes	yes	yes	yes	yes
Observations	169,070	169,070	169,070	169,070	169,070	169,070

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 7 – Third country effect in trade-migration nexus: IV model - step 1*

Variables	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
ln z	0.893^a	0.906^a	0.843^a
	(0.003)	(0.001)	(0.001)
ln of total Immigration* from $\Omega_{ks,t}$	-0.000	-0.004^a	-0.007^a
	(0.000)	(0.000)	(0.000)
ln [(Tariff/100)+1]	-0.399^a	-0.216^a	-0.153^a
	(0.008)	(0.028)	(0.024)
ln Distance	-0.099^a	-0.148^a	-0.061^a
	(0.009)	(0.002)	(0.002)
Contiguity	0.227^a	0.142^a	0.425^a
	(0.013)	(0.004)	(0.006)
Language	-0.164^a	-0.194^a	-0.010^a
	(0.008)	(0.003)	(0.001)
Colonial	0.300^a	0.060^a	0.256^a
	(0.008)	(0.003)	(0.002)
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	169,070	169,070	169,070

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 8 – Third country effect in trade-migration nexus: IV model - step 2

	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Main variables			
ln Immigration*	0.028^{a}	0.045^{a}	0.064^{a}
	(0.007)	(0.006)	(0.007)
ln of total Immigration* from $\Omega_{ks,t}$	0.036^{a}	0.027^{a}	0.022^{a}
	(0.001)	(0.001)	(0.001)
Control variables			
ln [(Tariff/100)+1]	-1.758^{a}	-0.628^{a}	-0.415^{a}
	(0.201)	(0.174)	(0.161)
ln Distance	-1.004^{a}	-0.901^{a}	-0.739^{a}
	(0.011)	(0.009)	(0.009)
Contiguity	1.115^{a}	1.194^{a}	1.224^{a}
	(0.030)	(0.024)	(0.025)
Language	0.125^{a}	0.032^{b}	-0.027
	(0.018)	(0.015)	(0.016)
Colonial	0.527^{a}	0.333^{a}	0.007
	(0.024)	(0.019)	(0.019)
$\hat{v}$	-0.144^{a}	-0.137^{a}	-0.135^{a}
	(0.013)	(0.011)	(0.013)
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	169,070	169,070	169,070

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 9 – Placebo test: IV model - step 1*

Variables	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
ln z	0.896[a]	0.912[a]	0.864[a]
	(0.004)	(0.002)	(0.006)
ln of total Immigration* outside $\Omega_{ks,t}$	-0.002[a]	-0.001	-0.003[a]
	(0.000)	(0.001)	(0.000)
ln [(Tariff/100)+1]	-0.421[a]	-0.378[a]	-0.222[a]
	(0.009)	(0.005)	(0.008)
ln Distance	-0.094[a]	-0.132[a]	-0.156[a]
	(0.008)	(0.007)	(0.007)
Contiguity	0.225[a]	0.176[a]	0.346[a]
	(0.005)	(0.012)	(0.021)
Language	-0.128[a]	-0.199[a]	-0.294[a]
	(0.009)	(0.007)	(0.011)
Colonial	0.324[a]	0.065[a]	0.234[a]
	(0.008)	(0.005)	(0.010)
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	169,070	169,070	169,070

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. ([a]), ([b]) and ([c]) denote statistical significance at 1%, 5% and 10%, respectively.

Table 10 – Placebo test: IV model - step 2

	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Main variables			
ln Immigration*	0.026[a]	0.042[a]	0.061[a]
	(0.007)	(0.006)	(0.006)
ln of total Immigration* outside $\Omega_{ks,t}$	-0.017	-0.011	-0.008
	(0.014)	(0.010)	(0.008)
Control variables			
ln [(Tariff/100)+1]	-2.454[a]	-1.255[a]	-0.906[a]
	(0.200)	(0.159)	(0.146)
ln Distance	-0.993[a]	-0.898[a]	-0.737[a]
	(0.008)	(0.009)	(0.009)
Contiguity	1.146[a]	1.209[a]	1.232[a]
	(0.030)	(0.051)	(0.040)
Language	0.152[a]	0.054[a]	-0.005
	(0.018)	(0.015)	(0.006)
Colonial	0.506[a]	0.318[a]	-0.009
	(0.025)	(0.019)	(0.020)
$\hat{v}$	-0.153[a]	-0.151[a]	-0.151[a]
	(0.013)	(0.011)	(0.013)
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	169,070	169,070	169,070

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 11 – Third country effect in trade-migration nexus: IV model by quality, step 1 – $\tau = 0.50$**

Variables	Bottom 20%	40%-60%	Top 20%
ln z	0.904[a]	0.911[a]	0.909[a]
	(0.001)	(0.002)	(0.001)
ln of total Immigration* from $\Omega_{ks,t}$	-0.004[a]	-0.003[a]	-0.003[a]
	(0.000)	(0.000)	(0.000)
ln [(Tariff/100)+1]	-0.242[a]	-0.352[a]	-0.138[c]
	(0.065)	(0.066)	(0.075)
ln Distance	-0.155[a]	-0.123[a]	-0.147[a]
	(0.004)	(0.006)	(0.003)
Contiguity	0.149[a]	0.137[a]	0.134[a]
	(0.010)	(0.011)	(0.007)
Language	-0.197[a]	-0.182[a]	-0.196[a]
	(0.006)	(0.007)	(0.005)
Colonial	0.069[a]	0.055[a]	0.055[a]
	(0.006)	(0.007)	(0.006)
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	25,788	20,387	19,994

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. ([a]), ([b]) and ([c]) denote statistical significance at 1%, 5% and 10%, respectively.

Table 12 – Third country effect in trade-migration nexus: IV model by quality, step 2 – $\tau = 0.50$

	Bottom 20%	40%-60%	Top 20%
Main variables			
ln Immigration*	0.128^{a}	0.098^{a}	0.078^{a}
	(0.019)	(0.020)	(0.021)
ln of total Immigration* from $\Omega_{ks,t}$	0.032^{a}	0.038^{a}	0.067^{a}
	(0.003)	(0.002)	(0.004)
Control variables			
ln [(Tariff/100)+1]	-3.107^{a}	-1.904^{a}	-1.765^{a}
	(0.354)	(0.234)	(0.186)
ln Distance	-0.959^{a}	-0.945^{a}	-0.983^{a}
	(0.029)	(0.009)	(0.034)
Contiguity	0.869^{a}	0.903^{a}	1.053^{a}
	(0.085)	(0.097)	(0.090)
Language	-0.056	0.103^{a}	0.294^{a}
	(0.048)	(0.036)	(0.055)
Colonial	0.296^{a}	0.198^{a}	0.045
	(0.072)	(0.065)	(0.074)
$\hat{v}$	-0.159^{a}	-0.132^{a}	-0.053
	(0.029)	(0.011)	(0.037)
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	25,788	20,387	19,994

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 13 – Third country effect in trade-migration nexus: country-pair fixed effects

	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Main variables			
ln Immigration*	0.017^a	0.032^a	0.041^a
	(0.006)	(0.005)	(0.006)
ln of total Immigration* from $\Omega_{ks,t}$	0.019^a	0.016^a	0.014^a
	(0.003)	(0.001)	(0.001)
Control variable			
ln [(Tariff/100)+1]	-1.146^a	-0.501^a	-0.402^a
	(0.154)	(0.132)	(0.093)
Country-pair FE	yes	yes	yes
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	169,070	169,070	169,070

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

Table 14 – Third country effect in trade-migration nexus: country-pair fixed effects by quality - $\tau = 0.50$

	Bottom 20%	40%-60%	Top 20%
Main variables			
ln Immigration*	0.093^a	0.065^a	0.044^a
	(0.009)	(0.008)	(0.008)
ln of total Immigration* from $\Omega_{ks,t}$	0.021^a	0.028^a	0.047^a
	(0.002)	(0.001)	(0.001)
Control variable			
ln [(Tariff/100)+1]	-1.983^a	-1.003^a	-0.943^a
	(0.123)	(0.145)	(0.138)
Country-pair FE	yes	yes	yes
Country-by-year FE	yes	yes	yes
Sector-by-year FE	yes	yes	yes
Observations	25,788	20,387	19,994

Notes: * we considered the imputed values for immigration. Standard cluster errors by importer and exporter countries in parentheses. (^{a}), (^{b}) and (^{c}) denote statistical significance at 1%, 5% and 10%, respectively.

6. Appendix

In this appendix we describe the steps of the estimator developed by Wei, Ma and Carroll (2012), which is adapted from Wei, Ma and Carroll (2012, pp. 424-425). In what follows, we assume that $\ln x$ is always observed and contains the constant term, but the flow of migration $\ln m$ may be missing. Thus, we can write the quantile model as

$$Q_\tau(\ln y|\ln m, \ln x) = \ln m\beta_1(\tau) + \beta_2(\tau)\ln x. \tag{10}$$

We assume that n is the total sample size, and n_1 of these n observations are complete, while the remaining n_0 of them have missing. Additionally, we consider a missing at random assumption that conditional on $\ln x$. The main goal is to estimate the regression parameter $\beta(\tau) = (\beta_1(\tau), \beta_2^T(\tau))^T$. It is assumed the assumption missing at random, which means that conditional on x, the event that m is missing is independent of x and the response variable y. The method has the following steps:

Step 1: For a set of points of $\tau \in (0,1)$, perform quantile regression with the complete data only and write the resulting coefficients as $\hat{\beta}(\tau)$. That is, for a set of τ values in $(0,1)$, obtain $\hat{\beta}(\tau) = \arg\min_\beta \sum_{i,j=1}^{n_1} \rho_\tau\left(\ln y_{ij} - \ln m_{ij}\beta_1(\tau) + \beta_2(\tau)\ln x_{ij}.\right)$, where $\rho_\tau(r) = r(\tau - I(r<0))$ is an asymmetric L_1 loss function. In practice, τ is typically chosen to be evenly spread and sufficiently dense grid on $(0,1)$.

Step 2: Impute the missing values based on $f(\ln m|\ln y, \ln x) \propto f(\ln y|\ln m, \ln x)f(\ln m|\ln x)$, so they can be uniquely determined from the two densities $f(\ln y|\ln m, \ln x)$ and $f(\ln m|\ln x)$.

Step 2a: In this step, we estimate the conditional density $f(\ln y|\ln m, \ln x)$. Given the assumption that the model (10) is correctly specified we can write the conditional density $f(\ln y|\ln m, \ln x)$ as a function of the quantile coefficient process, that is, $f(\ln y|\ln m, \ln x) = f\{\ln y|\ln m, \ln x, \beta_0(\tau))$, where is $\beta_0(\tau)$ the true quantile coefficient. Thus, we choose quantile levels $\tau_k = k/(K_n+1)$, with $k = 1, ..., K_n$, where K_n is the number of quantile levels and we approximate the conditional density function $f(\ln y|\ln m, \ln x)$ by

$$\hat{f}\{\ln y|\ln m, \ln x, \hat{\beta}(\tau)) = \sum_{k=1}^{K_n} \frac{\tau_{k+1} - \tau_k}{(\ln m, \ln x^T)\hat{\beta}(\tau_{k+1}) - (\ln m, \ln x^T)\hat{\beta}(\tau_k)}\mathcal{I}.$$

Where $\mathcal{I}$ is the indicator function: $\mathcal{I} = I\{(\ln m, \ln x^T)\hat{\beta}(\tau_k) \leq \ln y < (\ln m, \ln x^T)\hat{\beta}(\tau_{k+1})\}$.

Step 2b: Estimate the conditional density $f(\ln m|\ln x)$. We model $\ln m$ given $\ln x$ parametrically as $f(\ln m|\ln x, e)$. Under the missing at random assumption, we estimate $\widehat{e}$ based on the complete data, and the estimated conditional density of $\ln m$ given x as $\hat{f}(\ln m|\ln x, \widehat{e})$

Step 2c: Estimate the conditional density $f(\ln m|\ln y, \ln x)$ as

$$\hat{f}(\ln m|\ln y, \ln x) \propto \hat{f}\{\ln y|\ln m, \ln x, \hat{\beta}(\tau))\hat{f}(\ln m|\ln x, \hat{\varepsilon}),$$

and impute the missing $\ln m$ accordingly. Each missing $\ln m$ is simulated from $\hat{f}(\ln m|\ln y, \ln x)$ by randomly drawing a $\mathrm{Un}(0,1)$ random variable, and inserting it into the quantile function $F^{-1}(u|\ln y, \ln x)$ for $u \in (0,1)$.

Step 3: We re-estimate $\beta(\tau)$ including the imputed data. It is possible to assemble a new objective function including the completely observed data and the ℓth imputed dataset as

$$\begin{aligned} S_{n(\ell)}(\beta) = &\sum_{i,j=1}^{n_1} \rho_\tau \left(\ln y_{ij} - \ln m_{ij}\beta_1(\tau) + \beta_2(\tau)\ln x_{ij}\right) \\ &+ \sum_{i,j=n_1+1}^{n} \rho_\tau \left(\ln y_{ij} - \ln m_{ij(\ell)}\beta_1(\tau) + \beta_2(\tau)\ln x_{ij}\right), \end{aligned}$$

and define $\widehat{\beta}_{*(\ell)} = \arg\min_\beta S_{n(\ell)}(\beta)$ as the estimated coefficient using the ℓth assembled complete data. We repeat this imputation-estimation step M times, and the multiple imputation estimator is $\widetilde{\beta}(\tau) = M^{-1}\sum_{\ell=1}^{M}\widehat{\beta}_{*(\ell)}$.