I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15722

How Do Immigrants Promote Exports? Networks, Knowledge, Diversity.

Gianluca Orefice
Hillel Rapoport
Gianluca Santoni

NOVEMBER 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15722

How Do Immigrants Promote Exports? Networks, Knowledge, Diversity.

Gianluca Orefice
University Paris-Dauphine, PSL and CESifo

Hillel Rapoport
Paris School of Economics, CEPII, LISER and IZA

Gianluca Santoni
CEPII

NOVEMBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

How Do Immigrants Promote Exports? Networks, Knowledge, Diversity.*

How do immigrants promote exports? To answer this question we propose a unified empirical framework allowing to identify and disentangle the main mechanisms put forth in the literature: the role of networks in reducing bilateral transaction costs, and the productivity shifts arising from migrationinduced knowledge diffusion and increased workforce diversity. While we find evidence supporting all three channels (at both the intensive and the extensive margins of trade), our framework allows to gauge their relative importance. When focusing on diversity, we find stronger results in sectors characterized by more complex production processes and more intense teamwork cooperation. This is consistent with theories linking the distribution of skills to the comparative advantage of nations. The results are robust to using a theoretically-grounded IV approach combining three variations on the shift-share methodology.

JEL Classification: F14, F16, F22, O47

Keywords: international trade, birthplace diversity, migration, productivity

Corresponding author:
Hillel Rapoport
Paris School of Economics
48 Boulevard Jourdan
75014 Paris
France
E-mail: Hillel.Rapoport@psemail.eu

* We are grateful to Andrea Ariu, Simone Bertoli, Matilde Bombardini, Thomas Chaney, Frederic Docquier, Runjuan Liu, Elisabetta Lodigiani, Thierry Mayer, Ariell Reshef, and to participants at the CESifo "Global Economy Area" workshop, the OECD-CEPII Immigration Conference, and at seminars held at CEPII, University of Luxembourg, Paris-Dauphine University, Universitat Jaume I, University of Ancona and University of Bari for insightful comments. The views expressed in this article are those of the authors and do not reect those of CEPII. Hillel Rapoport acknowledges support from Agence Nationale de la Recherche under the framework of the Investissements davenir program (ANR-17-EURE-001).

1 Introduction

The trade-creating effect of migration is well documented. It has been largely attributed to the role of migration networks in reducing informational barriers and transaction costs between home and host countries.[1] This interpretation in terms of networks and information hinges on the fact that the analysis is conducted at the bilateral level (which is the relevant dimension for network effects to manifest themselves) and shows stronger results for trade in differentiated goods, that is, where informational frictions in the buyer-seller relationship are most relevant. While we concur with most of this literature, we note that the gravity framework makes it almost tempting to overlook the export-enhancing effects of immigration that operate at the aggregate level. Indeed, these are absorbed by the country-time fixed effects in the empirical gravity specification. In particular, recent research has emphasized that immigration can boost productivity (and, hence, exports) through channels such as immigration (or birthplace) diversity as well as through migration-induced knowledge diffusion that affect productivity and exports *to any country*. There is ample historical (e.g., Hornung 2014, Moser and San 2019) and contemporary (e.g., Bahar and Rapoport 2018; Bahar et al. 2019) evidence of migration-driven knowledge transfers between countries. The same holds for the birthplace diversity channel (Ager & Brueckner 2013; Ortega & Peri 2014; Alesina, Harnoss & Rapoport 2016; Docquier, Turati, Valette & Vasilakis 2020). Note that the stronger effects found in previous literature for differentiated goods, which support the information channel, are also consistent with these alternative channels, as differentiated goods have certain characteristics (e.g., they require combining more tasks, or tasks that are more intensive in cognitive abilities and teamwork) that make them more prone to benefit disproportionately from increased diversity and knowledge in production.

People originating from a diverse set of countries bring at destination a more diverse set of skills, experiences, ideas, expertise and problem-solving capabilities. Such diversity has been shown to improve the efficiency of production and the overall performance of firms, as if workers from different countries were *de facto* different factors of production (Lazear 1999, Hong & Page 2001, Horwitz & Horwitz 2007).[2] In particular, the diversity in the birthplace of immigrants, by improving the skill dispersion of workers, is expected to promote productivity in sectors relying heavily on complex tasks, where problem solving capabilities are relatively more important. In these sectors, a more diverse distribution of workers' types is more valuable due to sub-modularity in production processes, and shapes the comparative advantage of nations (Maggi & Grossman 2000).[3]

This paper makes four main contributions. First, to the best of our knowledge it is first to jointly test for the three channels through which immigration affects export performance – networks, knowledge diffusion, and

[1] See Gould (1994) and Head & Ries (1998) for early contributions; Rauch (2001); Rauch & Trindade (2002); Felbermayr & Toubal (2012) for cross-country comparisons; and Parsons & Vezina (2018) for a recent assessment exploiting a natural experiment. See also Kugler & Rapoport (2007), Leblang (2010), Javorcik, Ozden, Spatareanu & Neagu (2011), Kugler, Levintal & Rapoport (2018) and Burchardi, Chaney & Hassan (2019) who make a similar argument for FDI and other financial investments.

[2] Even within narrowly-defined skill-cells, immigrants and native workers appear as imperfect substitutes in production (Ottaviano & Peri 2012).

[3] Maggi & Grossman (2000) theoretically demonstrate that the dispersion of skills may represent a source of comparative advantage in such sectors.

diversity – in a unified empirical framework and to do so at both the *intensive* and *extensive* margins of trade. Second, we explore the mechanisms through which immigrants' birthplace diversity affects the comparative advantage of countries. We do so using both bilateral and aggregate country-sector regressions, and test the heterogeneous effects of birthplace diversity across sectors. We conjecture that sectors relying more heavily on problem-solving capabilities (i.e., that could be modeled with sub-modular production functions *à la* Maggi and Grossman 2000) will benefit relatively more from the greater dispersion in the distribution of skills and abilities that diversity brings about.

Third, we use the predicted country-sector-year fixed effects of a structural gravity equation (i.e., the Exporter Multilateral Resistance Terms - MRT) to build a synthetic measure of a country ex ante revealed comparative advantage (RCA) in the vein of Costinot, Donaldson & Komunjer (2012). The main advantage of this strategy is that it allows for packing the exporter MRTs and disentangling the effect of knowledge diffusion and of workforce diversity on comparative advantage. These channels have been largely overlooked in most previous studies due to their adopting a strict bilateral trade perspective (hence these effects are completely absorbed in the country-year fixed effects).[4]

Fourth and finally, we address the endogeneity of immigrants' location decisions: based on a Random Utility Model (RUM) for migration, we propose three theoretically-grounded extensions of the shift-share IV à la Card (2001) aimed to address the identification challenges in the shift-share approach highlighted in Borusyak, Hull & Jaravel (2021). Our first IV relies on the supply-driven component of migration stocks; by removing any demand-driven factors from a predicted bilateral stock of migrants, we are able to use the exogenous variation in the settlement of immigrants across destinations. The second IV complements the first one in that it adds the feedback effect emphasized in Jaeger, Ruist & Stuhler (2018). Finally, the third IV hinges on the (supply-driven) inflows of immigrants following natural disasters in the origin countries.

The rest of the paper is organized as follows. In sections section 2 and 3 we use bilateral trade data to test the effect of the three migration-related channels on the extensive and intensive margins of trade. In particular, section 2 describes the empirical strategy and section 3 discusses the results. In section 4 we use aggregate country-level data to focus on the role of birthplace diversity in affecting export performance. The last section concludes.

2 Immigration and export performance: a unified framework

Previous literature highlighted networks and knowledge diffusion as main explanations for the well documented trade-creating effect of migration. At the same time, birthplace diversity has been shown to have positive effects

[4]The exporter MRTs are based on a first stage bilateral trade regression; we include the bilateral migration stock in the first stage regression to purge the RCA index from the transaction cost channel that cannot be explicitly controlled for in country-sector aggregate regressions. These synthetic measures of revealed comparative advantages are freely available here.

on countries' productivity (Alesina et al. 2016, Ottaviano & Peri 2006) and is therefore expected to have an impact on exports too. This paper tests these three channels in a unified empirical framework, with the goal of i) properly assessing their statistical significance, which makes it crucial to have them jointly, and ii) gauging their relative importance. We thus run the following gravity model:[5]

$$y_{ikjt} = \beta_1 Mig_{ijt} + \beta_2 KD_{ikt} + \beta_3 BD_{it} + \mathbf{X}_{ijkt} + \theta_{ij} + \theta_{jkt} + \theta_{rckt} + \varepsilon_{ijkt} \tag{1}$$

where the dependent variable y_{ijkt} is either a dummy variable equal to one if the country i exports to j in a given SIC 3-digit sector k at time t (*extensive* or *participation* margin) or the total exports of country i to j for sector k and time t, conditioned on being already serving the market jk at time $(t-1)$ - *intensive margin*.[6]

Three main explanatory variables characterize the empirical exercise. First, the stock of immigrants (in ln) in destination i from origin j and time t, (Mig_{ijt}), aims at capturing the transaction cost channel. The presence in country i of immigrants coming from country j is expected to boost exports from i to j ($\beta_1 > 0$). Second, we test the knowledge diffusion channel by including in equation (1) the proportion of immigrants in country i coming from all origins - but j - having a Revealed Comparative Advantage in sector k in 1995 (i.e., $RCA_{jk,1995} > 1$) - KD_{ikt}.[7] We use the Balassa Index in 1995 to approximate the ex-ante comparative advantage of the migrants' origins in a given sector k. A Balassa Index greater (smaller) than one suggests a comparative advantage (*dis*advantage) of a country in sector k. In line with Bahar & Rapoport (2018), in testing the knowledge diffusion channel we exclude j specific migrants in i to capture the *spillover* nature of the knowledge diffusion, and avoid any overlap with the transaction cost channel. Indeed, the diffusion of knowledge by migrants from a given country ($o \neq j$) is expected to affect the export flows towards *all* destinations (including j) and not specifically toward the country of origin of immigrants (o). Moreover, we take the *proportion* of migrants originating from countries with comparative advantage as we want to capture the effect of migrants stock *composition* at destination rather than the simple presence (level) of migrants from a subset of origins.[8] The variable KD_{ikt} therefore captures the effect of migrants originating from countries having a comparative advantage in sector k. Finally, in line with previous literature, we define Birthplace Diversity (BD) as one minus the Herfindahl-Hirschman (HH) concentration index applied to the population of immigrants: $BD_{i,t} = 1 - \sum_{j=1}^{J} s_{ijt}^2$, where s_{ijt} is the share of immigrants originating from country j in the total population of immigrants residing in country i at time t. The index of birthplace diversity $BD_{i,t}$ increases with

[5] See Head & Mayer (2014) for a discussion on the gravity model for trade.

[6] Another potential trade margin, the number of destinations, requires different data aggregation (exporter-sector-year aggregated data) and will be explored in section 4.

[7] $KD_{ikt} = \frac{\sum_{o \neq j} I_{ok,1995} Mig_{iot}}{\sum_o Mig_{iot}}$, with $I_{jo,1995}$ equal to one if origin $o \neq j$ has a Balassa Index greater than one in sector k in 1995, and o stands for the origin country of migrants. To avoid overlap with the transaction cost channel, in the numerator of KD_{ikt} we consider all possible origin countries o but j. By doing so, we mechanically purge the knowledge diffusion proxy from any transaction cost effect.

[8] Indeed, the simple presence of immigrants from a subset of origins is highly correlated with the total number of immigrant in country i (scale effect), here captured by fixed effects.

the diversity in migrants' birthplaces in the country (it is equal to 0 if country i hosts immigrants coming from only one origin country). The birthplace diversity index $BD_{i,t}$ can be interpreted as the probability that two randomly selected *foreign born residents* are from different countries of origins. In a robustness check reported in table C2 we use the ethnic polarization index (Montalvo & Reynal-Querol 2005) as an alternative measure of (inverse) birthplace diversity.

The set of control variables - $\mathbf{X}_{ijkt}$ - includes standard trade policy variables: (i) a dummy for bilateral trade agreement RTA_{ijt} (capturing the effect of a preferential market access), and (ii) the *applied tariff*, included as $log(1+tariff)_{ikjt}$, which controls for the tariff level faced by country i in exporting to j in sector k.[9] Moreover, we include the stock of emigrants from i living in j to control for both the import demand effect (home bias in consumption tastes) of country i's emigrants residing in j and their potential contribution to reducing bilateral information costs.[10]

Three sets of fixed effects are always included in the estimations. First, country pair fixed effects (θ_{ij}) control for any pair-specific time-invariant factor affecting bilateral trade (e.g., geographic distance, common colonial ties, common language). Note that the inclusion of country pair fixed effects implies that the identification of the information/network channel on its within dimension, thus reducing omitted variable concern substantially - see section 2.2. Second, importer-sector-year fixed effects (θ_{jkt}) control for any unobserved importer country-sector-year factor that may affect bilateral exports towards the market jk (i.e., total import demand and/or price in j). In particular, this set of fixed effects controls for the multilateral resistance term on the importer side (Head & Mayer 2014). Since one of the variables of interest ($BD_{i,t}$) is exporter country-year specific, we cannot include fixed effects on this dimension. Namely, exporter country-sector-year that would capture exactly the multilateral resistance term on the exporter side cannot be included. To (partially) address the potential omitted variable problem, on top of the exporter-specific effects subsumed in θ_{ij}, we always include fixed effects specific to the macro region and income level (and sector-year) of the exporter country, θ_{rckt}.[11] A similar strategy is used in Alesina et al. (2016) who include macro regions fixed effects, as country dummies would be perfectly collinear with birthplace diversity. By doing so, any unobserved sectoral shock specific to a macro region within a given income level is captured by fixed effects. As further controls for the (exporter) multilateral resistance term we include: (i) a country remoteness index;[12] and (ii) a series of dummy variables

[9] The applied tariff is the minimum between preferential (if any) and MFN rate. Notice that the effect of MFN tariff imposed in country j in sector k is captured by the importer-sector-year fixed effects. So any significant coefficient on $log(1+tariff)_{ikjt}$ comes from the presence of preferential tariff. Data on applied tariffs are from the WITS-TRAINS database.

[10] Emigrants from i to j can also convey productive knowledge in the host country j, and affect average productivity of j in a given sector – see Bahar & Rapoport (2018). In our empirical framework this is fully captured by importer-sector-year fixed effects.

[11] The macro-region and the income levels of countries are obtained from World Bank classification. For example we have a dummy for South American countries belonging to the same income level (as defined by the World Bank, for the income level we consider year 1995). Table B2 presents a detailed description of each region-income level cell and the number of countries belonging to each cell.

[12] Following Yotov, Piermartini, Monteiro & Larch (2017) we construct the remoteness index for the exporting country as: $ln(Remote)_{it} = ln(\sum_j^J dist_{ij}/E_{jt}/Y_t)$; where E_{jt} and Y_t represent respectively the total expenditure of exporting country i at time t and the world GDP at time t. The remoteness index increases when large destinations markets j (having large expenditure over GDP) are relatively closer than small destination markets. We therefore expect a positive coefficient associated to $ln(Remote)_{it}$.

(bins) for the quartile of total exports of country i in a given sector k at time t. Total exports bins aim at capturing the time-varying export capacity of country i in sector k independently of the specific destination, as suggested by standard gravity equation.[13] In order to reduce any endogeneity concern (i.e., bad control problem), in calculating total export bins of country i we exclude the direct exports towards j and all other destinations belonging to the macro-region of j. This makes export bins plausibly exogenous with respect to the dependent variable which is destination j specific. Therefore, within each region-income cell, and conditional on export bin and market access (i.e., remoteness), exporting countries are assumed to be plausibly homogeneous in terms of sources of comparative advantage other than skill dispersion and composition (i.e., factor endowments, technological level, quality of institutions and infrastructure).

As discussed in Appendix section A, and in line with Maggi & Grossman (2000) and Bombardini, Gallipoli & Pupato (2014), the effect of birthplace diversity is expected to be particularly beneficial for sectors characterised by sub-modular production functions, where having a more disperse distribution of workers types in the labor market constitutes an asset and determines a comparative advantage in the sector (Maggi & Grossman 2000). People migrating from different origin countries bring at destination a *diverse* set of skills, experiences, ideas, expertise and problem-solving capabilities that may be useful to improve the efficiency of the production process and the overall performance of the firm (Lazear 1999, Hong & Page 2001, Horwitz & Horwitz 2007). This theoretical intuition allows us to understand the mechanism through which birthplace diversity may affect the international competitiveness of a country. We dedicate section 4 to carefully test this mechanism but provide a first discussion here. Accordingly, we augment specification (1) and interact the birthplace diversity index with two proxies for problem solving intensity in sector k: (i) abstract tasks intensity - $Abstract_k$ (our baseline), and (ii) teamwork cooperation index - $Team_k$ (main robustness check). The underlying assumption is that sectors intensive in abstract tasks and in teamwork cooperation are more likely to be problem solving intensive and therefore characterized by sub-modular production functions. The abstract tasks intensity is a dummy variable indicating whether sector k is intensive in complex and abstract tasks. Data on abstract intensive sectors are from Autor & Dorn (2013).[14] The teamwork intensity of sectors comes from Bombardini, Gallipoli & Pupato (2012) and is built on the O*NET measure of teamwork intensity (i.e., on the importance of workers' interactions to perform a job - see Bombardini et al. (2012) section IIIA for a more detailed description of this index).[15] In section 4, where we specifically look at the role of birthplace diversity, we provide a battery of alternative proxies for the problem solving intensity of sectors. To test the heterogeneous impact of Birthplace Diversity

[13] In a standard gravity equation the export flow from country i to j depends on the overall international competitiveness of country i (i.e., the marginal cost in Armington model under perfect competition). This may be approximated by bins in overall export performance of country i purged by j specific factors.

[14] More detailed information available here https://www.ddorn.net/data.htm.

[15] In order to build sector specific teamwork intensity based on O*NET, Bombardini et al. (2012) match O*NET data with 2000 US microdata census indicating which occupations are required in each sector. Hence, they compute the average Teamwork index across occupations within each sector. The hypotheses we implicitly make is that the sectors' occupation composition across countries is the same as that in the US.

across sectors with different problem solving intensities we augment specification (1) as follows:

$$y_{ikjt} = \beta_1 Mig_{ijt} + \beta_2 KD_{ikt} + \beta_3 (BD_{it} \times Abstract_k) + \mathbf{X}_{ijkt} + \theta_{ij} + \theta_{jkt} + \theta_{it} + \varepsilon_{ijkt} \tag{2}$$

Our interest is now on the interaction term $BD_{it} \times Abstract_k$ (and on $BD_{it} \times Team_k$ when teamwork intensity is used as proxy for problem solving intensity). This interaction is ikt specific and allows for the inclusion country-year fixed effect (θ_{it}) on top of country-pair and importer-sector-year fixed effects, which are always included in our regressions. By including exporter country-year fixed effects, we considerably reduce endogeneity concerns; in particular, we reduce concerns about high-exporting countries attracting immigrants from a wider range of origins, hence generating a spurious correlation between diversity and international competitiveness.[16] The drawback of this specification is the impossibility to obtain the average effect of BD_{it} as it is perfectly collinear with exporter-year fixed effects. For this reason in the results tables we report both the specification without and with exporter country-year fixed effects. The reader can find informative the results on BD_{it} estimated with a less conservative set of fixed effects as those showed in equation (1).

2.1 Data and Descriptive evidence

All the migration related variables (i.e., bilateral migration stocks, knowledge diffusion and birthplace diversity) are based on ij specific bilateral stocks of migrants from United Nations (2015). This dataset provides information on bilateral migration stocks for a 195*195 matrix of origin-destination combinations, for the years 1990, 1995, 2000, 2005, 2010, 2015.[17] The main advantage of this dataset, with respect to other sources (such as the IMD-OECD), is the balanced nature of the data which include all OECD and non-OECD destination countries. For periods prior to 1990 (used to build our instrumental variable) we use data from the World Bank *Global Bilateral Migration Database*, see Ozden, Parsons, Schiff & Walmsley (2011). In table 1, for a sub-sample of the countries covered in our empirical analysis, we report the stock of immigrants from all origins, and the value of birthplace diversity in the years 1995, 2005 and 2015.

Export-based measures of international competitiveness (i.e., total exports, intensive and extensive margins) are based on the BACI (CEPII) dataset. We have information on bilateral export flows to/from 195 countries over the period 1995-2015 at product HS 6-digit level. However, since the problem solving intensity measures

[16] The set of exporter-year fixed effects θ_{it} also controls for the quality of institutions in the exporting country, which has been highlighted as empirically relevant in analyzing the social consequences of birthplace diversity at destination (Arbatli, Ashraf, Galor & Klemp Forthcoming). This set of fixed effects also controls for the income level of the destination country. Indeed, Alesina & La Ferrara (2005) show that the GDP per capita at destination is important in assessing the role of ethno-linguistic fractionalization on productivity and other indicators of economic performance. In this respect, it must be noticed that ethno-linguistic diversity is conceptually and statistically different from the diversity in birthplaces considered in this paper. While birthplace diversity considers people born in different countries and educated in different schooling systems, ethno-linguistic diversity builds on people born and raised in the same country but with different ethnic or linguistic backgrounds. And indeed, the two indices are empirically almost totally uncorrelated. See Alesina et al. (2016).

[17] The dataset *Trends in International Migrant Stock: The 2015 Revision* (United Nations database, POP/DB/MIG/Stock/Rev.2015) is available at: `http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates15.shtml`.

discussed in the previous section are available at the SIC 3-digit level, we aggregate the trade data at the country-pair-sector-year level where the sector is defined as SIC 3-digit. Data on the presence of Preferential Trade Agreements and on bilateral distances are from CEPII databases, while tariffs are from WITS. Data on GDP per capita (used to calculate the remoteness measures), as well as income and regional classifications, are from World Bank Development Indicators data.

After merging 5-year windows UN migration stock data with BACI (CEPII) trade flows and other control variables data, we end up with a panel of 195 exporting/immigration destination countries, 176 importing/emigration origin countries, 142 sectors, and observations every 5 years.[18] Out of the 24,367,200 potential observations, because of missing data, our extensive margin regression analysis (including zero trade flows) is based on 20,156,093 observations. The intensive margin analysis, being based on positive trade flows only, relies on 4,575,395 observations. In Table 2 we show in-sample descriptive statistics for the main variables included in our intensive margin estimations. Figure 1 shows the simple correlations between the total exports of a country and two migration-related channels at the core of our empirical exercise: (i) the total stock of immigrants (plot on the left), and (ii) birthplace diversity (plot on the right).[19] Figure 1 is suggestive of a positive correlation between the total stock of immigrants and the exports of country i (transaction cost channel); and also of a positive correlation between birthplace diversity and total exports of country i. Although unconditional and potentially plagued by important omitted variable biases, these positive correlations are consistent with expectations and will be confirmed by our empirical analysis.

2.2 Endogeneity

Equations (1) and (2) will be consistently estimated if the covariance between our variables of interest and the error component ε_{ijkt} is null (conditioned on controls and fixed effects).[20] This condition is verified in absence of omitted (unobserved) variables and reverse causality problem. The inclusion of the three sets of fixed effects described in the previous section, by controlling for any country-pair and country-year determinant of bilateral exports, strongly reduces the omitted variable concern in eq. (1) and (2). Also, country-pair fixed effects, by implying the identification of migration-related channels on the within dimension, partially address the reverse causality concern.[21] However, it may still be the case that unobserved country-sector ik specific shocks affect contemporaneously the export performance of a country and the settlement of immigrants coming from different origins (i.e., positive productivity shocks boosting the export of country i and attracting immigrants from several

[18] We lose the year 1990 available on UN migration data because BACI trade data start in 1995; we also lose a few importing countries due to missing values in the tariff data.

[19] Providing graphical evidence of the knowledge diffusion channel is difficult because it is based on sector-specific spillover effects.

[20] Formally if: $Cov\left(MIG_{ijt}, \varepsilon_{ijkt} | \mathbf{x}_{ijkt}, \theta_{ij}, \theta_{jkt}, \theta_{it}\right) = 0$. The same condition must hold for the knowledge diffusion and birthplace diversity measures.

[21] The fact that the average *level* of bilateral trade may shape the bilateral stock of migrants is captured by country-pair fixed effects. The reverse causality argument must play in *deviation* form country-pair averages.

origins). Moreover, reverse causality may produce biased Ordinary Least Squared (OLS) estimations if *changes* in the international competitiveness of a country (and, thus, its exports) have an impact on the labor demand for immigrants workers.

These endogeneity concerns are addressed here by adopting an Instrumental Variable approach that uses (in turn) two *original* IVs and a third instrumental variable approach in the vein of Jaeger et al. (2018). The three IVs proposed here are theoretically based on a Random Utility Model for migration developed in section 2.2.1. The first IV is based on the predicted supply-driven migration stocks purged from any demand-driven effect - see section 2.2.2. The second IV is based on the main idea in Jaeger et al. (2018), and removes the feedback effect in the predicted supply-driven migration stocks - see section 2.2.3. Finally, the last IV builds on the predicted supply-driven migration but uses the time variation in immigration flows coming from origins that experienced natural disasters (i.e., an exogenous shock in the push factors) - see section 2.2.4.

2.2.1 Theoretical foundation of the Instrumental Variable

This section provides a theoretical foundation for the Instrumental Variables adopted in this study. In the vein of the enclave approach developed by Card (2001), we adapt a Random Utility Model for migration (RUM) to highlight the role of previous settlement of immigrants in affecting contemporaneous bilateral migration flows. From the estimation of such a theoretically grounded bilateral migration equation, we subsequently: (i) extract the predicted value, (ii) purge it from every endogenous destination-specific factors (that *de facto* are a cause of concern about the validity of the standard enclave approach), and (iii) use the purged predicted supply-driven migrant stocks to build the IVs for our three variables of interest.

Consider the situation in which a representative individual h, currently residing in country j, has to decide the optimal location i^* from a set of possible destinations $i \in I$, with I containing also the current country of residence j (i.e., no-migration option). The optimal destination i^* is obtained by maximizing: $i^* = argmax_{i \in I} U_{hjit}$. In line with previous papers deriving RUM models for migration, the utility of opting for destination i (originating from j) has the following form:

$$U_{hjit} = A_{it} - c_{ji} + D_{jit} + \xi_{hjit}. \tag{3}$$

The observable component of the utility of individual h in equation (3) includes: (i) the overall attractiveness of destination i at time t, A_{it} (often proxied in the literature by pull factors, or magnets, such as the expected income or the employment rate at destination); (ii) the bilateral migration cost c_{ji} here assumed to be time invariant for simplicity;[22] and (iii) the benefit from having a large community of migrants from the same origin

[22] Over the period covered in this study the cost of migration associated to distance and other bilateral geographic factors can be

j at destination i (*Diaspora* term D_{jit}). The positive effects of the existing set of migrants from the same origin on later migration flows has been extensively documented in the empirical migration literature, and explicitly introduced in a RUM model for migration by Buggle, Thoenig, Mayer & Sakalli (2020). The unobservable component of the utility ξ_{hjit} captures all the individual specific unobservable factors affecting the location decision. We assume that ξ_{hjit} following a type 1 Extreme Value distribution. In this setting, McFadden (1974) shows that the probability that an individual h will find it optimal to move from j to i has the following form:

$$p_{jit} = \frac{e^{A_{it} - c_{ji} + D_{jit}}}{\sum_d^I e^{A_{dt} - c_{dj} + D_{jdt}}} \tag{4}$$

At the aggregate level, since all individuals at origin j extract the same utility from migrating to i (except for the random component ξ_{hjit}) the probability in equation (4) corresponds to the proportion of individuals in j that find it optimal to migrate to i. Hence, the predicted migration flow from j to i can be expressed as the product between the total population in j, (N_{jt}), and the probability of migrating from j to i: $M_{jit} = p_{jit} \times N_{jt}$. Hence, the logarithm of bilateral migration flows can be expressed as follows:

$$ln(M_{jit}) = A_{it} - c_{ji} + D_{jit} - ln\left(\Omega_{jt}\right) + ln\left(N_{jt}\right) \tag{5}$$

with the term $\Omega_{jt} = \sum_d^I A_{dt} - c_{dj} + D_{jdt}$ representing the aggregate utility associated to all destinations $d \in I$ available for migration in country j. The higher the value of Ω_{jt}, the lower the migration flows from j to a specific country i.[23] By estimating equation (5) we obtain the theoretically consistent imputed bilateral migration flows as a base for our shift-share IVs. In the empirical counterpart of equation (5), destination-year fixed effects δ_{it} will capture the overall attractiveness of the destination country at time t (A_{it}); country pair fixed effects δ_{ij} will capture the time-invariant migration cost c_{ij}; and origin-time fixed effects δ_{jt} will absorb the population at origin N_{jt} and the Ω_{jt} term. Hence, the only component to be specified before turning to estimating equation (5) is D_{jit} (i.e., the size of the diaspora community from j in destination i at time t). This has often been approximated in previous literature by the stock of migrants in destination i from the same origin j until year t (not included) - $MigStock_{ji,t-1}$. Here we slightly depart from the previous literature and follow Buggle et al. (2020): we approximate the diaspora term D_{jit} in relative terms, as the share of the migrant stock from origin j at time $t-1$ over the total population residing at destination at the start of the sample period t_0:

considered invariant. Notice that considering time-variant bilateral migration costs is straightforward from a theoretical point of view but complicated in the empirics when it comes to find a proxy for them.

[23] The term Ω_{jt} mimics the outward multilateral resistance term in a gravity for trade but applied to migration. See Anderson (2011) and Bertoli & Fernandez-Huertas Moraga (2013) for a detailed discussion of the multilateral resistance to migration terms and how they can be derived from a RUM for migration.

$D_{jit} = \frac{MigStock_{ji,t-1}}{Pop_{i,t0}}$.[24]

Estimating equation (5) raises two empirical issues. The first relates to whether one should take the log of bilateral migration as dependent variable and run an OLS model, or estimate it in levels and run a PPML model. Here we follow Silva & Tenreyro (2006) and run a PPML model on levels to address potential heteroskedasticity in the error term.[25] The second issue concerns endogeneity. Since our objective is to use the fit of the empirical counterpart of equation (5) as the base for our IVs, the diaspora component must not be endogenous with respect to the international competitiveness of the destination country. To this end, rather than using the *observed* stock of immigrants $MigStock_{ji,t-1}$, in the vein of Card (2001) we use the *imputed* stock of immigrants $\widetilde{MigStock}_{ji,t-1}$ computed as follows:

$$\widetilde{MigStock}_{ji,t-1} = \frac{MigStock_{ji,t0}}{\sum_i^I MigStock_{ji,t0}} \times MigStock_{j,t-1} \tag{6}$$

This is based on the idea that *contemporaneous* outflows of migrants from a given origin ($MigStock_{j,t-1}$) are allocated across destinations based on the *historical* geographical distribution of migrants from the same origin country (we use 1960 as $t0$ in equation 6). Therefore, the empirical counterpart of equation (5) can be written as:[26]

$$M_{jit} = \exp\left[\delta_{it} + \delta_{jt} + \delta_{ji} + \gamma_1 \frac{\widetilde{MigStock}_{ji,t-1}}{Pop_{i,t0}}\right] * \varepsilon_{jit} \tag{7}$$

The fit of equation (7) is the predicted bilateral migration *flows* between country i and j at time t. However, our three variables of interest (Mig_{ijt}, KD_{ikt}, BD_{it}) are based on the *stock* of immigrants at destination; we therefore need an exogenous variation in bilateral migration *stocks* (rather than flows) to instrument Mig_{ijt}, KD_{ikt} and BD_{it}. To this end, we simply note that stocks are recursive additions of net bilateral migration flows on existing stocks and are determined by the same forces that shape flows over time (i.e., attraction and pull-factors δ_{it}, push-factors δ_{jt}, migration costs δ_{ij} and the diaspora effect). Therefore, the same covariates in (7) can be used to estimate bilateral migration *stocks*. Notice that equation (7) is very similar to the bilateral migration stock equation derived in Burchardi et al. (2019) to estimate the stock of residents in a given

[24] Rescaling the stock of migrants for the overall size of the destination takes into account how diluted is the origin-specific migrant community over the entire population of the destination country. We take the population at destination at the start of the sample to avoid any spurious correlation with contemporaneous migration flows.

[25] Guimares, Figueirdo & Woodward (2003) show that if the discrete choice model does not include decision-maker choice-specific variables – as in our equation (5) – then the PPML log-likelihood is identical to the multinomial logit that is routinely used for discrete choice problems. See Buggle et al. (2020) for more discussion on this point. Hence, the PPML estimator is also consistent with the discrete choice nature of our model.

[26] Since we estimate the equation for migration flows with PPML we report its exponentiated version.

destination from a specific origin (ancestry) and time.[27] For these reasons, as a baseline strategy we use the fit of equation (7) estimated on stocks to instrument the variables Mig_{ijt}, KD_{ikt} and BD_{it}.[28] However, we are aware of the fact that using stocks is not fully consistent with a RUM model of migration; therefore, as a robustness check reported in table (7), we estimate equation (7) on migration flows, and use the *cumulated* fit (i.e., predicted migration flows) to impute the stock of migrants and instrument Mig_{ijt}, KD_{ikt} and BD_{it}.[29]

The final step, before using the fit of equation (7) - $\widehat{M_{ijt}}$ - as a source of exogenous variation to instrument the migration related measures Mig_{ijt}, KD_{ikt} and BD_{it}, is to purge it from the destination-year fixed effects:

$$\widehat{AdjImmi}_{ijt} = \widehat{M_{ijt}} - \widehat{\delta_{it}}. \tag{9}$$

By doing so, we explicitly exclude from the predicted bilateral stocks of migrants any "problematic" destination-specific labor demand component that may invalidate our IV and obtain the predicted *supply-driven* stock of immigrants from j in destination i. This represents an important contribution with respect to previous literature that uses the shift-share IV introduced by Card (2001).[30] Indeed, a valid identification for the shift-share IV requires that any confounding factor affecting the economic outcomes of the destination country i (such as productivity shocks, economic growth, export performance or natives' wages) does not simultaneously affect the interaction of the past geographic distribution of immigrants in i with the total number of migrants from j.[31] If destination-specific shocks attract immigrants from specific origins, the shift component of a standard shift-share IV would be endogenous and, hence, not valid.[32] We address this problem by explicitly removing every possible destination-time specific factor from the predicted bilateral migration stocks, while keeping the original Card (2001) enclave intuition on how emigrants distribute across destinations. Therefore, in our case the key identifying assumption becomes:

[27] See equation (2) in Burchardi et al. (2019).

[28] Notice also that with the inclusion of country pair fixed effects in equation (7), our identification is based on deviation from country-pair averages which can be considered an (imperfect) approximation of flows.

[29] The supply driven migrant stock predicted by cumulated flows is given by:

$$\widehat{AdjImmi}_{jit}^{Cumul} = MigStock_{ji,1980} + \sum_{t=1995,..,2015} \left(\widehat{M_{ijt}} - \widehat{\delta_{it}}\right) \tag{8}$$

where $MigStock_{ji,1980}$ is the observed stock of migrant in 1980 (used as a base to cumulate predicted flows), and $\widehat{M_{ijt}}$ and $\widehat{\delta_{it}}$ are respectively the fit and the exporter-year component of equation (7) estimated on migration flows rather than stocks. $\widehat{AdjImmi}_{jit}^{Cumul}$ is therefore used to build IV for Mig_{ijt}, KD_{ikt} and the birthplace diversity index. Results from this robustness check are reported in columns 4 and 8 of table 7 and largely confirm our baseline results.

[30] Previous papers on the labor market effects of immigration have often adopted the shift share instrument à la Card (2001) to address endogeneity. See for example Ottaviano & Peri (2006); Peri & Requena-Silvente (2010); Card (2009); or Edo & Rapoport (2019).

[31] As discussed in Borusyak et al. (2021), the validity of the IV in our empirical framework is challenged by the presence of potential unobserved common components (e.g., productivity or technological shocks) driving both the settlement of immigrants from several origins in the exporting country (shift component of the IV) and its export performance. The removal of the exporter-time component in equation (9) reduces such a concern.

[32] For this reason we cannot directly use the shift-share IV in our empirical framework: labor demand shocks related to the international competitiveness of country i may directly attract migrants from a specific origin country (reverse causality).

$$Cov\left(\widehat{AdjImmi}_{ijt}, \varepsilon_{ijkt}|\mathbf{x}_{ijkt,\theta_{ij},\theta_{jkt},\theta_{it}}\right) = 0 \tag{10}$$

Since the predicted supply-driven stock of migrants $\widehat{AdjImmi}_{ijt}$ is purged from destination specific shocks, the exclusion restriction assumption in eq. (10) is likely to be valid; and so are the IVs for Mig_{ijt}, KD_{ikt} and BD_{it} built on the predicted supply-driven bilateral migration stocks $\widehat{AdjImmi}_{ijt}$.

The structure of the IV presented so far echoes the gravity-based IV widely used in the trade literature to instrument trade openness measures (see Frankel and Romer 1999). This approach has been successively adopted in the migration literature to instrument migration flows at destination - see Ortega & Peri (2014) and Docquier, Lodigiani, Rapoport & Schiff (2016), among others. In particular, Ortega & Peri (2014) and Docquier et al. (2016) estimate a bilateral migration gravity equation including several proxies for geographic and cultural distance as explanatory variables. The fit of this equation is then used as an instrumental variable for the total stock of migrants at destination. While geographic and cultural distance can fairly be assumed exogenous with respect to the economic performances of the destination country (exclusion restriction), both papers include other gravity-related variables (i.e., population and immigration policy at destination) that may be affected by the economic outcomes of the destination country.[33] Under this circumstance, the exclusion restriction is not satisfied. It is therefore important to remove from the gravity-based predictor, the estimated destination country-time specific component δ_{it}.

2.2.2 IV 1: the modified shift-share based instrumental variable

The predicted supply-driven bilateral migration stocks $\widehat{AdjImmi}_{ijt}$ are directly used to instrument transaction costs (i.e., the bilateral stocks of migrants Mig_{ijt}) in equations (1) and (2), and are aggregated as done for the KD_{ikt} variable to also instrument for knowledge diffusion. Finally, we build the instrumental variable for the birthplace diversity index using $\widehat{AdjImmi}_{ijt}$:

$$\widehat{BD_{it}^{PPML}} = 1 - \sum_{j=1}^{J}\left(\frac{\widehat{AdjImmi}_{ijt}}{\sum_{j=1}^{J}\widehat{AdjImmi}_{ijt}}\right)^2 \tag{11}$$

Notice that $\left(\frac{\widehat{AdjImmi}_{ijt}}{\sum_{j=1}^{J}\widehat{AdjImmi}_{ijt})}\right)$ is the share in the total population of the *supply driven* predicted number of migrants in country i originating from country j.[34] Therefore, our instrumental variable $\widehat{BD_{it}^{PPML}}$ is built using the pure supply-driven component of the bilateral migration stocks, and can be used safely as an instrumental variable. In this case, the exclusion restriction assumption is that the diversity index based on the *predicted supply-driven* migration stocks ($\widehat{BD_{it}^{PPML}}$) affects the competitiveness of a country only through the BD_{it} index

[33] Negative economic shocks may translate into a change in the destination country's population or of its immigration policy.

[34] A relevant property of the PPML model used to estimate equation (7) is the fact that the fit corresponds exactly to the size of the *predicted* immigrant stock.

based on the *observed* migration stocks. This is plausible because the variability of $\widehat{BD_{it}^{PPML}}$ bases on j specific outflows of immigrants and not on the i specific component of bilateral migrant stocks. Another usual criticism of the standard shift-share instrument is the non-orthogonality of the initial distribution of immigrants used to allocate subsequent migration inflows. By using the distribution of immigrants in 1960 (35 years before the initial year of our estimations), this concern is reduced.[35]

In RUM models for migration, potential migrants make their decision as to whether and where to migrate based on the expected wages (or attractiveness) of all potential destinations - see Bertoli & Fernandez-Huertas Moraga (2013). In presence of segmented labor markets, the wage considered by the potential migrant depends on his/her origin, and specifically on the cultural similarity with the destination country. Hence, removing the destination country-year component of the predicted migration flows may not be sufficient to remove endogeneity concerns. Therefore, in a first robustness check, we remove from the predicted migration stock $\widehat{M_{ijt}}$ any destination-year-origin group specific component that may endogenously affect the settlement of migrants across destinations. Namely, for each destination country we identify four groups of origins based on quartiles of language similarity,[36] and augment equation (7) with destination-year-origin group fixed effects. Our adjusted set of IVs for Mig_{ijt}, KD_{ikt} and BD_{it} is therefore based on eq. (9) where we subtract such destination-origin group specific components from the predicted bilateral migration. Results from this robustness check are reported in columns 3 and 7 of table 7 and largely confirm our baseline results.

2.2.3 IV 2: a modified shift-share based instrumental variable controlling for feedback effects

As discussed in Jaeger et al. (2018), the country of origin mix for a given destination is likely to be invariant over time (i.e., high persistence of immigrants' settlement across destination countries). This implies a high degree of autocorrelation in the shift-share instrument, that therefore captures both the short- and the long-term effect of immigration at destination.[37] If the short- and the long-term effect have opposite expected signs on the outcome variable (international competitiveness here), then the resulting estimates using a standard shift-share approach have an unclear interpretation. To address this potential bias, in the spirit of Jaeger et al. (2018), we remove from the predicted bilateral migration stocks $\widehat{M_{ijt}}$ the long run component of the term $\frac{\widehat{MigStock}_{ji,t-1}}{Pop_{i,t0}}$.[38] Namely, we obtain the *predicted supply-driven* stock of immigrants in each country i based on the estimation of the following structural gravity model for migration:

[35] Notice that given our bilateral setting, we cannot apply the procedure suggested by Goldsmith-Pinkham, Sorkin & Swift (2020) aiming at identifying the relevant "shares" driving the estimates - unless replicating the procedure for each of the 195 destination countries. We reckon that the validity of our instruments relies on the fact that, conditional on local "demand pull factors" (i.e. $\widehat{\delta_{it}}$), the distribution of immigrant shares in 1960 is plausibly orthogonal to trade flows in 1995-2015.

[36] Language similarity data are from Melitz & Toubal (2014).

[37] Jaeger et al. (2018) show a positive correlation between shift-share instruments and their lags equal to 0.96.

[38] We do not follow exactly the multiple instrumentation proposed by Jaeger et al. (2018) as a higher number of instruments in presence of a large set of fixed effects would produce a low-efficient estimator. However, we definitely follow Jaeger et al. (2018) in spirit by removing from the predicted immigration flow the long-term component $Immi\ Sh_{ij,60} * ln(Immi)_{jt-2}$.

$$M_{jit} = \exp\left[\delta_{it} + \delta_{jt} + \delta_{ji} + \gamma_1 \frac{\widetilde{MigStock}_{ji,t-1}}{Pop_{i,t0}} + \gamma_2 \frac{\widetilde{MigStock}_{ji,t-2}}{Pop_{i,t0}}\right] * \varepsilon_{jit} \tag{12}$$

where the variables have the same meaning as in equation (7) and we split the diaspora component (i.e., the past distribution approach à la Card 2001) into short- and long-term components (i.e., $\widetilde{MigStock}_{ji,t-1}$ and $\widetilde{MigStock}_{ji,t-2}$ respectively). From equation (12) we take the predicted value $\widehat{M_{jit}}$ (fit of the regression) and subtract the destination-year fixed effect and the long-term component ($\widetilde{MigStock}_{ji,t-2}$) as follows:

$$\widehat{AdjImmi}^{short}_{ijt} = \widehat{M_{ijt}} - \widehat{\delta_{it}} - \widehat{\gamma_2} \frac{\widetilde{MigStock}_{ji,t-2}}{Pop_{i,t0}}. \tag{13}$$

By doing so, we purge the predicted *supply driven* bilateral stock of immigrants from any demand-driven effects and from the long-term component highlighted by Jaeger et al. (2018). Finally, we use $\widehat{AdjImmi}^{short}_{ijt}$ to build the IVs for Mig_{ijt}, KD_{ikt} and for the Birthplace Diversity index.

2.2.4 IV 3: a natural disaster based instrumental variable

An alternative to address endogeneity is to rely on natural experiments. Natural disasters (tsunamis, earthquakes, floods, etc.) have been shown to be a strong predictor of human mobility in many developing countries (Gray & Mueller 2012, Beine & Parsons 2017).[39] We therefore compute the birthplace diversity index based on the *predicted supply-driven* stocks of immigrants induced by countries that experienced (at least one) natural disaster in the pre-treatment period, i.e 1985-1990.[40] To do this, we use the supply-driven predicted stock of immigrants ($\widehat{AdjImmi}_{ijt}$) from equation (9) only for the sub-sample of origins j with at least one natural disaster over the period 1985-1990. This variable is then used to instrument Mig_{ijt} and KD_{ikt} and the birthplace diversity index.

Notice that for a precise calculation of the diversity measure, we cannot omit migrants from origin countries that did not experience natural disasters during the period considered (i.e., 1985-1990). Indeed, if we computed the BD index using the $\widehat{AdjImmi}_{ijt}$ only for the subsample of countries that have experienced natural disasters, we would miss a large number of origins and the resulting BD would be strongly biased.[41] We therefore use the bilateral stock of immigrants in 1960 to include the origin countries that did not experience natural disasters in the period 1985-1990. The formula below describes our alternative instrumental variable for the birthplace diversity index:

[39] Beine & Parsons (2017) show that while natural disasters *per se* have a null (slightly negative) effect on overall emigration, they considerably boost emigration towards destinations with low migration costs.
[40] See Appendix B for details on natural disaster data used in the paper.
[41] The total number of countries j affected by a catastrophic natural event in our sample is 41.

$$BD_{it}^{\widehat{PPML},ND} = 1 - \sum_{j=1}^{J} \left[\left(I_{j,85-90} \frac{\widehat{AdjImmi}_{ijt}}{\sum_{j=1}^{J} \widehat{AdjImmi}_{ijt}} \right)^2 + \left((1 - I_{j,85-90}) \frac{Immi_{ij,60}}{\sum_{j=1}^{J} Immi_{ij,60}} \right)^2 \right] \quad (14)$$

where $Immi_{ij,60}$ is the stock of immigrants in i from origin j in 1960, and $I_{j,85-90}$ is a dummy variable equal to one if country j experienced a natural disaster in the period 1985-1990. The first term of the squared bracket in equation (14) applies to origins with natural disasters and uses the predicted supply-driven component of bilateral migration ($\widehat{AdjImmi}_{ijt}$) to compute the BD index. The second term of the squared bracket applies to countries without natural disaster and uses bilateral stock of immigrants in 1960 to compute the squared share of immigrant from j (taken in 1960 to reduce endogeneity concerns in the time variation of the stocks of migrants coming form disasters-free origins). Figure 2 qualitatively supports the identification strategy used in this case. It shows a clear positive relationship between natural disaster events which occurred in the period 1985-1990 in origin countries j and subsequent outward migration (univariate R-square 0.79).

3 Results

Estimation results are reported in Tables 3 and 4 respectively for the extensive and intensive margins of trade. The structure of the two tables is similar. In columns (1)-(4) we show OLS results discussed in section 3.1, while in columns (5)-(6) we report 2SLS results using the IV strategy described above. 2SLS results are then discussed in section 3.2. In terms of specifications, in columns (1)-(2) and (5) we show results concerning the baseline (eq. 1) and augmented specification (eq. 2) without exporter-year fixed effects, while in columns (3), (4) and (6) we report estimations with exporter-year fixed effects. Specifications in columns (3) and (4) differ for the level of clusters in standard errors (i.e., country-year in column (3) and the more demanding country-specific clustering in column (4)).

3.1 OLS Results

Extensive margin results. In line with previous literature, OLS results show a strong and robust positive effect of bilateral migration stocks on the extensive margin of trade of country i towards country j - see columns (1)-(4) in Table 3. This is the standard transaction cost channel, already highlighted in several previous papers. It captures the fact that migrants from a specific origin provide additional information to firms at destination on how to export to j (i.e., consumers' tastes, regulations, business contacts to establish a distribution network, etc.). In particular, using the specification in column 4 of Table 3, a 10 percent increase in the stock of immigrant from j increases the probability that country i exports to j by 0.18%. The knowledge diffusion channel is also supported by our results. When the composition of the immigrants in i is twisted toward origins having a comparative advantage in sector k, then exports to the rest of the world of that sector is boosted further (this is as if immigrants originating from countries $o \neq j$ with a comparative advantage in sector k were diffusing knowledge and best practices learned at home, making host countries more productive in those sectors). See columns (1)-(4) in Table 3. On top of the transaction costs and knowledge diffusion channel, the diversity in the origins of migrants has a positive and statistically significant effect on the extensive margin of exports in all specifications (see columns 1 and 2). The birthplace diversity effect is stronger in abstract tasks intensive sectors (here used as a proxy for problem solving intensity) as revealed by the interaction term ($BD_{it} \times Abstract_k$) - see columns 2, 3 and 4. This last result is particularly relevant because it is obtained after including exporting country-year fixed effects that considerably reduce omitted variables concerns.

We may therefore conclude, at least tentatively, that the three migration-related channels play at the same time a significant role in affecting the international competitiveness of countries at the extensive margin. In line with *Conjecture 1* discussed in Appendix A, we may also conclude that the effect of birthplace diversity works through a productivity increase in sectors intensive in problem solving capabilities (i.e., abstract sectors), where a more diverse set of skills shapes the sector's comparative advantage (Maggi & Grossman 2000). In order to provide a quantitative interpretation of the relative magnitudes of the three migration-related channels, we follow Helpman, Melitz & Yeaple (2004) and report in Table 5 the standardised coefficients for the baseline estimations.[42] It emerges that the transaction cost represents quantitatively the most important channel in affecting the extensive margin, while birthplace diversity and knowledge diffusion - though highly significant - have smaller magnitudes.

Among the control variables included in the specifications, all have the expected sign. The only exception is tariffs, showing positive coefficient in columns (1), (2) and (5). This is likely to reflect an omitted variable bias; indeed, when we include exporter-year fixed effects in columns (3), (4) and (6), tariffs turn to have the expected negative and significant coefficient. However, notice that the role of tariffs in affecting the extensive margin of trade is not key. Indeed, a change in the variable cost of trade (tariffs) is expected to have a null/slightly negative effect on the extensive margin of trade (the effect of tariff is expected to be much more relevant at the intensive margin estimation discussed below - see Chaney 2008).

Intensive margin results. In Table 4 we report the estimation results for the intensive margin of trade. The structure of the table is the same as Table 3. The transaction costs and knowledge diffusion channels have both a positive and significant effect at the intensive margin. The higher the stock of immigrant in i from origin j, the higher the total exports of country i towards j in sector k- see Table 4. Using the specification in column

[42]Standardised or "beta" coefficients are obtained by employing standardised variables of interest in the estimations (i.e., as the product between the original variable and its standard deviation, divided by the standard deviation of the dependent variable). See Wooldridge (2012). Such a standardization converts the original regression coefficients into units of sample standard deviations.

(4), a 10 percent increase in the bilateral stock of immigrants increases exports from i to j by 1 percent. The intensive margin of trade is also positively affected by birthplace diversity, which has a significant positive effect after controlling for transaction costs and for knowledge diffusion: a one standard deviation increase in the birthplace diversity index implies a 4.4 percent increase in the export flows between i and j in sector k at time t - see column 2.[43] This effect is magnified in abstract intensive sectors as shown by the positive and significant interaction between birthplace diversity and $Abstract_k$ in columns 2, 3 and 4.

All control variables included in equation (1) have the expected sign. Coherently with expectations, the presence of a Regional Trade Agreement increases bilateral exports among partner countries,[44] and remoteness has the expected positive coefficient. Interestingly, the positive coefficient on $ln(Emigrants)$ suggests that the presence of emigrants from i to j stimulates the import demand of j from i - preference channel in import demand (in line with evidence in Gould 1994). Tariffs always have always the expected negative sign. In particular, a 1% increase in bilateral (sector specific) tariffs reduces exports by 1.6-2.4%.[45]

3.2 2SLS Results

The baseline 2SLS estimation results using the IV discussed in section 2.2.2 are reported in columns (5)-(6) of Tables 3 and 4 (first stage results are reported in Table 6). Robustness checks with the alternative IVs discussed in sections 2.2.3 and 2.2.4 are reported in Table 7.

We instrument the bilateral stock of immigrants and the knowledge diffusion channel with the predicted supply-driven bilateral migration flows. We then instrument the birthplace diversity index with a diversity measure based on the *predicted* (rather than observed) bilateral migration stocks. These instruments are all based on the exogenous variation of the supply of immigrants in origin country j, $\widehat{AdjImmi}_{ijt}$. Any i-specific labor demand effect has been removed from the IVs (see the detailed discussion in Section 2.2). In column (5)-(6) of Table 3 we report 2SLS estimations on the extensive margin of exports. The coefficient for the bilateral stocks of migrants and for knowledge diffusion are both positive and significant, with overall similar magnitudes as for the OLS estimations (if anything, there is a very small downward bias in the OLS estimations). The coefficient for birthplace diversity remains positive and statistically as well, including when transaction costs and knowledge diffusion are explicitly controlled for. In Table 4 we report 2SLS estimations at the intensive margin of exports. The results confirm those obtained using OLS: the three channels (i.e., networks, knowledge diffusion and diversity) positively affect the export flows from country i to market kj. Birthplace diversity has a magnified positive effect in the case of abstract intensive sectors, suggesting once again that the role of diversity is particularly relevant for sectors with high problem-solving intensity.

[43] The standard deviation of birthplace diversity index is 0.18 (see table 1).

[44] The point estimate on the RTA dummy is smaller than that obtained (on average) in previous literature (Head & Mayer 2014) because the inclusion of country pair fixed effects absorbs part of the variation of the RTA dummy.

[45] The coefficient on tariffs' elasticity is coherent with many previous studies (Buono & Lalanne 2012, Fitzgerald & Haller 2018).

In Table 5 we show the 2SLS-based "beta" coefficients in order to get a sense of the relative magnitudes of the three migration channels affecting both the extensive and the intensive margins of trade. As expected, the transaction costs channel has the largest effect on trade (at both the extensive and intensive margins), with birthplace diversity and knowledge diffusion still playing an important role in affecting the international competitiveness of host countries. More precisely, the effect of birthplace diversity equals to respectively 25% and 15% of that of networks/transaction costs at the extensive and intensive margins. As for knowledge diffusion, these figures stand at 3.7 percent 16 percent, respectively. To put this differently: a one standard deviation increase in network size (bilateral migration), knowledge diffusion (KD) and birthplace diversity (BD) raises: i) the likelihood of exports (extensive margin) by respectively 16.3%, 0.6% and 4% of a standard deviation and ii) the logarithm of exports (intensive margin) respectively by 21%, 3.4% and 3.3% of a standard deviation.

A simple back-of-the-envelope calculation provides an alternative way of getting a sense of the relative magnitudes of the three migration channels. By combining the observed variation in Mig_{ijt}, KD_{ikt} and BD_{it} in the period 2005-2015, with the baseline 2SLS point estimates reported in column 4 of table 4, we obtain country i's *expected* export growth due to changes in the *observed* transaction cost, knowledge diffusion and birthplace diversity channels. Table C1 shows such a back-of-the-envelope calculation for a selected number of countries (the same set of countries are reported in Table 1). The first column reports the observed export growth of countries in the period 2005-2015; while the other columns report the *expected* export growth implied by observed changes in transaction costs (i.e., bilateral migration stocks), knowledge diffusion and birthplace diversity variables.[46] In line with the "beta" coefficients reported in Table 5, Table C1 shows the clear predominant role of the transaction costs channel in affecting the growth of country's exports. Given the observed change in US migration stocks which occurred in the period 2005-2015, the expected US export growth due to the transaction costs, knowledge diffusion and birthplace diversity variables are respectively 3.5%, -0.7% and 0.34%.[47] The cross-country export growth induced by the three migration-related channels are reported in figures C2, C3 and C4 for the transaction costs, knowledge diffusion and birthplace diversity channels, respectively.[48]

Since we rely solely on variation in the imputed number of immigrants $\widehat{AdjImmi}_{ijt}$ to instrument the three migration-related channels, it is important to show that each migration channel is properly instrumented by its IV counterpart. For example, the interacted birthplace diversity variable has to be identified by the interacted *diversity* index based on the imputed number of immigrants ($\widehat{AdjImmi}_{ijt}$), and not by the variable $\widehat{AdjImmi}_{ijt}$ per se. This is shown in Table 6 where we report the details for the first stage of the baseline equation using the supply-driven predicted stock of migrants to build the IVs (specification 6 of Tables 3 and 4.[49] Reassuringly,

[46]The *expected* export growth due to *observed* changes in migration-related variables are based on country-invariant elasticities to Mig_{ijt}, KD_{ikt} and BD_{it} and must therefore be interpreted with caution.

[47]Negative values for the expected export growth in Table C1 depend on negative changes in the migration-related explanatory varables (Mig_{ijt}, KD_{ikt} and BD_{it}) observed over the decade 2005-2015.

[48]Export growth in figures C2, C3 and C4 have been re-scaled to the mean export growth to assure the cross-country comparability.

[49]Recall that the omitted variable problem is considerably reduced by the inclusion of exporter-year fixed effects.

we find that each endogenous variable is explained by its respective IV. For both the intensive and extensive margin channel estimations, each IV is strongly correlated with the relevant variable of interest – see columns 1-3 and 4-6 in Table 6. It is only in columns (3) and (4) that we have two IVs which are simultaneously positive and significant predictors of one endogenous variable.[50] This may raise a small concern of unclear identification of the transaction cost channel in the 2SLS intensive margin specification, and of the interacted diversity in the extensive margin specification.

The F-stat of the first stage regressions reported at the bottom of Tables 3 and 4 support the absence of weak IVs. Note that the validity of the instruments cannot be tested with a Sargan test (exact identified model); however since they are based exclusively on the supply of immigrants from country j, they are plausibly valid (i.e., unrelated by construction to any country i specific shock). Since the labor demand component of country i has been explicitly removed from the predicted migration flows, the exclusion restriction here is that immigrants residing in i because they were "pushed away" from country j affect the export performance of country i only through their effect on the size and structure of its immigration. In other words, bilateral export performances are expected to be orthogonal with respect to the push component of emigration from j. Recall in addition that the allocation of exogenous "push" migration is made on the distribution of immigrants from j across destinations i in 1960 - see equation (6). With a lag of thirty years we are confident about the validity of our IVs.

These results are robust to the two alternative Instrumental Variables described in Sections 2.2.3 and 2.2.4 (columns 1-2 and 4-5 in Table 7) and to the exclusion of any destination-origin group specific labor demand shock (i.e., fixed effects) from the predicted immigrants $Adj\widehat{Immi}_{ijt}$ (columns 3 and 6 in Table 7). The results reported in Table 7 show again positive and significant coefficients for the (instrumented) transaction costs, knowledge diffusion and (interacted) birthplace diversity variables. The strength of the three instrumental variables is reported at the bottom of Table 7. The instruments are highly relevant and do not suffer any problem of potential weak instrument (F-stat above 10). As for the baseline IV discussed above, the validity of the alternative instruments cannot be tested with a Sargan test, however since they are also based on the pure supply of immigrants from country j, they are plausibly valid. The orthogonality of the bilateral migration flows is even stronger for the natural disasters-based IV where emigration is pushed by purely exogenous factors (see columns 2 and 6). It is also reassuring that our results are robust to the IV inspired by Jaeger et al. (2018) (see columns 1 and 5).

Overall, the above results are supportive of a positive, causal effect of birthplace diversity on the export performance of countries (alongside the other channels). However, it could be that the presence of too many, culturally distant migrant communities at destination lead to high coordination cost in production, generating a non-linear

[50] Namely, in column 4, the instrument for the interacted birthplace diversity ($BD_{it} \times Abstract_k$) has a small but significant correlation with the Mig_{ijt} variable.

relationship between birthplace diversity and productivity (and, hence, exports). We test this hypotheses in figure 3 by plotting the OLS estimation of birthplace diversity by quartile in the degree of dissimilarity between destination i and the its set of origins j (as approximated by the average language dissimilarity index between i and all origins j). It clearly emerges that the positive effect of birthplace diversity decreases for destinations countries at the top-quartile of the language dissimilarity index, where the very large diversity in the origins of migrant workers may imply a problem of coordination in production.

3.3 Robustness checks

As discussed in section 2, the O*NET-based teamwork intensity of occupations composing the sectors ($Team_k$) can be used as an alternative proxy for the problem solving intensity of sector k. Results for this robustness check are reported in Table 8 and confirm our baseline results. Transaction costs, knowledge diffusion and birthplace diversity have a positive and significant effect on the extensive and intensive margins of trade. Importantly, the positive effect of birthplace diversity is magnified for sectors intensive in teamwork collaboration: the availability of a more horizontally diverse set of workers originating from different countries is particularly beneficial for sectors characterized by a high degree of teamwork interactions. This result holds at both the extensive (columns 1-3) and intensive margins (columns 4-6).

In line with results in Alesina et al. (2016) and Docquier et al. (2020), problem solving capabilities are more likely to be transmitted at destination by high-skilled migrants. In Table 9 we use the OECD DIOC-E database (providing information on the education of immigrants) on bilateral stock of migrants in years 2000 and 2010 to test whether the effect of birthplace diversity is specific to tertiary educated migrants. We basically estimate empirical specification (1) but use three different versions of the BD_{it} variable, one for each level of education of migrants (primary, secondary and tertiary) in destination i year t.[51] As expected, the diversity effect seems entirely driven by highly-educated immigrants.

As a further robustness check in Table C2 we report results using an alternative measure of birthplace diversity. Based on Montalvo & Reynal-Querol (2005), instead of using (one minus) the Herfindahl-Hirschman index, we approximate birthplace diversity by computing the polarization index. In this case an *increase* in the polarization indicates a *reduction* in the diversity of migrant communities. Results reported in Table C2 confirm our baseline evidence that a wider diversity in the countries of origin of immigrants helps the international competitiveness of exporting countries through both the intensive and the extensive margin of exports.

[51] With only two available years from DIOC-E database, we could not apply our IV procedure (too small time variation to remove exporter-year fixed effect from the PPML fit) and we rely on OLS estimations only.

4 The role of diversity

This section focuses on the birthplace diversity channel and provides a direct empirical test of the underlying mechanism at play, namely the productivity effect of birthplace diversity.

4.1 Theoretical motivation

The effect of diversity on the productivity of production teams has been shown in many papers. Hong & Page (2001) theoretically show that a group of more diverse problem solvers may perform better than a group of homogeneous but more able problem solvers. Hoogendoorn & van Praag (2012) use a randomized field experiment to show that more culturally diverse teams have better performance than culturally homogeneous teams: in diverse teams the coordination costs from diversity are offset by the wider availability of relevant skills. A recent research published by McKinsey&Company find a significant positive relationship between culturally diverse teams and the financial performance of firms. Companies at the top quartile of diversity are 35 percent more likely to outperform their national industry median (Vivian, Dennis & Sara 2015). Relatedly, Kahane, Longley & Simmons (2013) analyze the national composition of National Hockey League teams in the US and find that more diverse teams have a better performance. Interestingly, Kahane et al. (2013) conclude that the "productivity" premium provided by diverse teams is driven by complementarity between native and foreign-born players' skills.[52] Trax, Brunow & Suedekum (2015) use German establishment level data to show that diversity of foreign born workers increases the productivity of plants. Accordingly, Parrotta, Pozzoli & Sala (2016) test the effect of diversity on the export performance of Danish firms, finding a strong positive effect of firm's workforce diversity on the extensive margin of exports (participation and number of export markets).

At the aggregate local labor market level, Ottaviano & Peri (2006) find that multicultural urban environments raise the productivity of US-born citizens. Recent studies by Ager & Brueckner (2013) and Rodriguez-Pose & Berlepsch (2019) on US counties identifies the presence of a strong positive impact of population diversity on county-level economic development: counties that received migrants from more diverse set of origins over the late 19^{th} century are nowadays significantly richer than counties with a more homogeneous population at the time. At the cross-country level, the positive effect of diversity on growth has been shown empirically in Alesina et al. (2016). Using a comprehensive 195 x 195 matrix of bilateral migration stocks for the years 1990 and 2000, the authors find that increasing the diversity of skilled immigration by 1 percentage-point increases long run economic output by about 2%. Similarly, Docquier et al. (2020) use US states data over the period 1960-2010 to show that diversity among college-educated immigrants has a positive effect on economic growth; namely a 10% increase in high-skill diversity raises GDP per capita by about 6%.[53] Finally, Bahar, Rapoport & Turati

[52] Peri & Sparber (2009) provide empirical evidence of a productivity effect from the complementarity among immigrant and native workers.

[53] Ortega & Peri (2014) as well as Alesina et al. (2016) adopt an instrumental variable approach to support the positive effect of

(2021) show that immigrants' birthplace diversity positively affects a host country's "complexity" (defined by the Hidalgo and Hausmann index – Hidalgo & Hausmann 2009) through a diversification of its exports basket.

To our knowledge, only few papers directly link population diversity, the structure of countries' comparative advantage and international trade.[54] Maggi & Grossman (2000) develop a theoretical model in which countries with a more diverse population have a comparative advantage in the production/export of goods characterized by high substitutability among employees in production (i.e., when the presence of highly-talented workers is relatively more important). Indeed, countries endowed with a more diverse distribution of workers' abilities have higher possibilities for matching extremely brilliant workers with more standard ones. This implies a comparative advantage in the sectors characterized by sub-modular technologies, where creativity and problem-solving capabilities are relatively more needed. From an empirical point of view, a first test of the relevance of skill-dispersion (diversity) on the comparative advantage of countries is provided by Bombardini et al. (2012). By combining IALS scores (purged of observable characteristics such as education, age and gender) with O*NET-based measures of skill complementarity, the authors show that countries with a more dispersed (residual) skill distribution tend to specialize in sectors with low skill-complementarity in production.

Our theoretical rationale for the positive impact of birthplace diversity on the export performance of countries follow this line of reasoning. Workers originating from a more diverse set of countries may have similar hard-skills (formal education) but different soft-skills in production. Hence, beyond the average productivity-boost induced by a diverse set of workers (as in Hong and Page 2001), host countries endowed with a mode diverse set of migrants (which translates into a higher birthplace diversity index as defined in section 1) will have a comparative advantage in sectors relying heavily on creativity and problem-solving tasks.[55] See Appendix section A for a more detailed discussion on this theoretical rationale. We test this specific mechanism in what follows. In particular, we test the effect of birthplace diversity on the export competitiveness and comparative advantage of countries, with a focus on sectors intensive in creative and problem-solving tasks (approximated here by a battery of sector characteristics).

4.2 Empirical Strategy and Results

The three migration-related channels discussed in the previous sections have different identifying variations. While the interacted diversity ($BD_{it} \times Abstract_k$) and the knowledge diffusion (KD_{ikt}) channels are country-

trade openness and diversity of immigration on long-run income per capita and productivity.

[54] It must be noted that with a CES production function and many imperfectly substitutable origin-specific workers, the production of the firm is maximized when hiring a perfectly equal share of workers across origins (i.e., perfect diversity in production).

[55] As discussed in Appendix section A, theoretical models in Maggi & Grossman (2000) or Bombardini et al. (2012) are based on the concept of *vertical* dispersion of workers (i.e., where workers are ranked by degree of ability and the dispersion of such a distribution matters in affecting the comparative advantage of the host country). Here we rely on the *horizontal* diversification of abilities. In other words, immigrants arriving from different origins are not ranked by their abilities (or education), but are horizontally differentiated based on the imperfect substitutability in production among immigrants from different origins.

sector-year specific, the transaction cost channel is country-pair specific.[56] This may lead to a potential aggregation bias in the estimation of the effect of birthplace diversity at the core of this section when estimated using bilateral data. We address this concern by estimating the trade effect of birthplace diversity at the same level of aggregation as the two variables of interest here (i.e., BD_{it} and its interaction with the sector problem solving intensity ($BD_{it} \times Abstract_k$)). Hence, in this section we aggregate trade-related variables at the country-sector-year level (i.e., total exports and number of destinations served by country i in sector k). We then calculate country-sectors comparative advantage indices and test the effect of birthplace diversity $BD_{i,t}$ and its interaction with a sector's problem-solving intensity. Given that the variable of interest is now country-sector-year specific, we can include both country-year (θ_{it}) and sector-year (θ_{kt}) fixed effects. With country-sector-year aggregated data at hand, we run the following econometric specification:

$$y_{ikt} = \beta_1(BD_{i,t} \times Abstract_k) + \beta_2 KD_{i,k,t} + \theta_{it} + \theta_{kt} + \epsilon_{i,k,t} \tag{15}$$

where the dependent variable y_{ikt} is either: (i) the total exports (in ln) of country i in sector k and time t; (ii) the number of destination countries reached by i on sector k and time t (in ln); or (iii) the Revealed Comparative Advantage (RCA) index à la Costinot et al. (2012). $BD_{i,t}$ is the birthplace diversity measure for country i at time t and $Abstract_k$ is the proxy for the problem solving intensity of sector k as described above. Fixed effects θ_{it} and θ_{kt} respectively control for any country-year and sector-year specific determinants of competitiveness. In particular, country-year fixed effects control for the transaction cost channel (total stock of migrants in country i time t). The main drawback of including country-year fixed effects is the impossibility to estimate the effect of diversity on the average sector (abstracting from its problem solving intensity). Therefore, in order to estimate both the average effect of diversity and its interaction with $Abstract_k$, in two initial specifications we omit the country-year fixed effects and include the total stock of immigrants residing in country i to control for the effect of migration-related shocks on competitiveness other than diversity. We also control for the number of preferential trade agreements in force for country i (as proxy for its average market access), and for the GDP of the country (in ln). Finally, we explicitly control for the knowledge diffusion channel by including the KD_{ikt} variable described above in all specifications.

One may question the inclusion of country-year fixed effects as a compelling way of controlling for the transaction cost channel. Indeed, it may be the case that the availability of immigrants in the country (transaction costs channel) might affect international competitiveness heterogeneously across sectors. In this case, the coefficient on the interaction between BD_{it} and the sector problem solving intensity would also capture some of the effect of bilateral migration (i.e., the transaction costs channel). The RCA index à la Costinot et al. (2012)

[56] Birthplace diversity *per se* is country-year specific but its effect depends also on sector characteristics (i.e., complexity and problem solving intensity).

allows to address this concern. As in Costinot et al. (2012), we compute a synthetic measure of the export performance of country i in sector k and time t conditioned on the effect of bilateral migration. This is obtained by keeping the country-sector-year fixed effect in the following auxiliary regression:

$$Export_{ijkt} = \delta_{ikt} + \delta_{jkt} + \delta_{ijk} + \beta_1 Mig_{ijt} + \beta_2 PTA_{ijt} + \mu_{ijkt} \tag{16}$$

where δ_{ikt}, δ_{jkt} and δ_{ijk} are respectively exporter-sector-year, importer-sector-year and country pair-sector fixed effects; whereas Mig_{ijt} controls for the effect of bilateral migration on exports (transaction costs channel). From eq. (16) we recover the estimated exporter-sector-year fixed effects ($\widehat{\delta_{ikt}}$) which represents a synthetic measure of the Revealed Comparative Advantage (export performance) of country i in sector k at time t.[57] Notice that equation (16) is estimated using a PPML model to control for the heteroscedasticity of trade flows and the incidence of zeros as suggested in Silva & Tenreyro (2006). In equation (16) we also include a dummy for the presence of a common Preferential Trade Agreement so as to control for possible preferential market access boosting bilateral trade. Being conditioned on bilateral migration, the RCA measure discussed above is purged from the transaction costs channel. We can therefore claim that the coefficient on ($BD_{it} \times Abstract_k$) precisely captures the effect of diversity and not that of other migration-related channels. We rely on the RCA index as main dependent variable for country-sector-year aggregated estimations.

The results are reported in Table 10. The first column aims at presenting the effect of birthplace diversity on the average sector (country-year fixed effects are therefore not included so as to allow for estimating BD_{it}). In column 2, with the same set of fixed-effects, we introduce the interaction between birthplace diversity and the problem solving intensity measure ($BD_{it} \times Abstract_k$). In all specifications we explicitly control for the knowledge diffusion channel. Coherently with previous results, we find that birthplace diversity has a positive and significant effect on the export performances (RCA index) of the country-sector, and more so for problem-solving intensive sectors. In columns (3)-(5) we estimate the augmented equation (15) including country-year fixed effects and focus on the heterogeneous effects of birthplace diversity ($BD_{it} \times Abstract_k$). Even after controlling for country-year fixed effects, we find that birthplace diversity has a strongly significant effect on the RCA index in sectors characterized by high problem-solving intensity (see column 3). The coefficient of the interacted term is positive and significant also on total exports (column 4) and on the number of destinations served by country i in sector k (column 5). Coherently with previous results, knowledge diffusion is always positive and significant for the three export margins considered here. Moreover, Table C3 in the Appendix presents the results obtained after instrumenting both birthplace diversity and knowledge diffusion, using the

[57] In a Ricardian-type model of trade, Costinot et al. (2012) show that exporter-sector-year fixed effects from a reduced form model as in equation (16) exactly mirror the ex-ante Ricardian comparative advantage of a country. This measure of revealed comparative advantage is made freely available for the interested scholars and practitioners on a dedicated webpage here (see appendix section D for a description of the dataset).

instruments detailed in Section 2.2.2. The results presented so far are largely confirmed.

In Appendix Table C4 we show further robustness tests using alternative proxies for the problem solving intensity of sector k, such as: (i) job complexity as defined by Costinot (2009), (ii) O*NET-based teamwork intensity, (iii) knowledge intensity by Bahar (2020),[58] (iv) a dummy for differentiated goods based on the Rauch (2001) classification, (v) the skill-intensity of the sector (dummy variable based on UNCTAD classification), and (vi) the technology intensity of the sector (dummy variable based on UNCTAD classification). Independently of the variable used to approximate the problem solving intensity of a sector, we find that the interacted birthplace diversity index has a positive and significant effect on the revealed comparative advantage index, with the exception of column (6) where the interaction coefficient is imprecisely estimated. Interestingly, as showed in Table C5, the birthplace diversity effect increases in the income level of the host country (high-income dummy interaction), while the knowledge diffusion channel helps in particular less developed countries and has an almost null effect on high-income exporters.

The evidence discussed so far points to the fact that birthplace diversity is particularly beneficial for sectors in which the variety of ideas and problem-solving capabilities are particularly relevant (differentiated goods, high-skill workers and technology-intensive sectors). In line with the theoretical background discussed above (discussed in more details in Appendix section A), we may therefore conclude that the birthplace diversity of immigrants translates into improved export performance through an increase in the efficiency of production processes characterized by problem solving intensive tasks (i.e., production processes plausibly characterized by sub-modular production functions).

4.3 Sector specific birthplace diversity

The lack of data on the number of migrant workers by *sector*,[59] forces us to implicitly assume an homogeneous distribution of immigrant workers across sectors of a given country (implying country-year specific birthplace diversity measure BD_{it}). In reality the diversity of origins among migrant workers across sectors (within a country) might differ considerably, and ideally the index of birthplace diversity should be calculated at the country-sector level. To do so and partially address this concern, we rely on French 1990 Census data providing the number of migrant workers by district and sector, and build a district-sector specific measure of birthplace diversity to test the robustness of our baseline results.[60]

We combine 1990 French Census data with national statistics on district-sector specific exports, and test the effect of *sector* specific birthplace diversity on the international competitiveness of French districts. We run

[58] The index captures the *tacit* knowledge intensity of an economic activity, based on the average (accumulated) experience and training of the workforce in an industry. The occupational characteristics are defined according to the O*NET dataset.

[59] To our knowledge, there is a lack of available migration databases providing information on the number of migrant workers by country of origin and sector over time (in particular for the full set of destination countries considered in this paper).

[60] The French districts are called "*départements*" and correspond to the NUTS3 classification of Eurostat.

this robustness check at the district-sector level because we do not have data on i) destination specific exports of districts, and on ii) change over time in the number of migrant workers by sector-district. Nevertheless, this represents an important robustness check supporting the validity of our results when the assumption of homogeneous distribution of immigrant workers across sectors is relaxed. This robustness check and the data sources are discussed in detail in Appendix section B. The results (reported in Table C6) confirm that the birthplace diversity of immigrants (now built at the sector level) has a positive and significant effect on the competitiveness of French districts.[61]

5 Conclusion

Immigration affects the economy in many ways. For one thing, immigration makes host countries demographically more diverse, more connected to the rest of the world, and more permeable to knowledge coming from overseas. As such, immigration has the potential to make host countries more productive, especially in sectors in which immigrants bring with them valuable knowledge from their home countries, or where diversity is a key ingredient for productivity. Migration-induced productivity shifts combined to lower access/transaction costs generated through the web of networks linking immigrants to their home countries eventually materialize in the form of into better export performance.

This paper demonstrates the joint workings of productivity-related effects of immigration (which translate into higher aggregate export performance to the rest of the world) and of the well-established network-based information channel (which translate into higher export performance bilaterally). We use a unified empirical framework to account for both the bilateral and aggregate channels and address various sources of endogeneity in the migration and trade relationship. We show that all three channels – networks, diversity, and knowledge diffusion – are simultaneously at play (at both the extensive and intensive margins) and we gauge their relative importance. When focusing on diversity and in line with theoretical intuition,[62] we find stronger positive effects of birthplace diversity (again, at both the extensive and intensive margins) on export performance in sectors relying more intensively on problem-solving tasks and teamwork. Given the growing importance of these sectors in all advanced economies, one can safely conjecture that immigration will become an even stronger strategic determinant of countries' comparative advantage and overall economic performance.

[61] District and sector fixed effects always included.

[62] Especially Maggi & Grossman (2000), for whom a more dispersed distribution of worker types in production is particularly beneficial for sectors characterized by sub-modular production process.

Bibliography

Ager, P. & Brueckner, M. (2013), 'Cultural diversity and economic growth: Evidence from the US during the age of mass migration', *European Economic Review* **64**(C), 76–97.

Alesina, A., Harnoss, J. & Rapoport, H. (2016), 'Birthplace diversity and economic prosperity', *Journal of Economic Growth* **21**(2), 101–138.

Alesina, A. & La Ferrara, E. (2005), 'Ethnic diversity and economic performance', *Journal of Economic Literature* **43**(3), 762–800.

Anderson, J. E. (2011), 'The Gravity Model', *Annual Review of Economics* **3**(1), 133–160.

Arbatli, C., Ashraf, Q. H., Galor, O. & Klemp, M. (Forthcoming), 'Diversity and conflict', *Econometrica* .

Autor, D. H. & Dorn, D. (2013), 'The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market', *American Economic Review* **103**(5), 1553–1597.

Bahar, D. (2020), 'The hardships of long distance relationships: time zone proximity and the location of mnc's knowledge-intensive activities', *Journal of International Economics* **125**, 103311.

Bahar, D., Özgüzel, C., Hauptmann, A. & Rapoport, H. (2019), Migration and Post-Conflict Reconstruction: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia, IZA Discussion Papers 12412, Institute of Labor Economics (IZA).

Bahar, D. & Rapoport, H. (2018), 'Migration, knowledge diffusion and the comparative advantage of nations', *Economic Journal* **128**(612), F273–F305.

Bahar, D., Rapoport, H. & Turati, R. (2021), 'Birthplace diversity and economic complexity: Cross-country evidence', *Research Policy* **forthcoming**.

Beine, M. & Parsons, C. R. (2017), 'Climatic factors as determinants of international migration: Redux', *CESifo Economic Studies* **63**(4), 386–402.

Bertoli, S. & Fernandez-Huertas Moraga, J. (2013), 'Multilateral resistance to migration', *Journal of Development Economics* **102**(C), 79–100.

Bombardini, M., Gallipoli, G. & Pupato, G. (2012), 'Skill Dispersion and Trade Flows', *American Economic Review* **102**(5), 2327–2348.

Bombardini, M., Gallipoli, G. & Pupato, G. (2014), 'Unobservable skill dispersion and comparative advantage', *Journal of International Economics* **92**(2), 317 – 329.

Borusyak, K., Hull, P. & Jaravel, X. (2021), 'Quasi-experimental Shift-Share Research Designs', *Review of Economic Studies* **forthcoming**.

Buggle, J., Thoenig, M., Mayer, T. & Sakalli, S. O. (2020), The Refugee's Dilemma: Evidence from Jewish Migration out of Nazi Germany, Cahiers de Recherches Economiques du Département d'économie 20.01, Université de Lausanne.

Buono, I. & Lalanne, G. (2012), 'The effect of the Uruguay round on the intensive and extensive margins of trade', *Journal of International Economics* **86**(2), 269–283.

Burchardi, K. B., Chaney, T. & Hassan, T. A. (2019), 'Migrants, ancestors, and foreign investments', *The Review of Economic Studies* **86**(4), 1448–86.

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.

Chaney, T. (2008), 'Distorted gravity: the intensive and extensive margins of international trade', *American Economic Review* **98**(4), 1707–1721.

Costinot, A. (2009), 'On the origins of comparative advantage', *Journal of International Economics* **77**(2), 255–264.

Costinot, A., Donaldson, D. & Komunjer, I. (2012), 'What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas', *Review of Economic Studies* **79**(2), 581–608.

Docquier, F., Lodigiani, E., Rapoport, H. & Schiff, M. (2016), 'Emigration and democracy', *Journal of Development Economics* **120**(C), 209–223.

Docquier, F., Turati, R., Valette, J. & Vasilakis, C. (2020), 'Birthplace diversity and economic growth: Evidence from the us states in the post-world war ii period', *Journal of Economic Geography* **20**(2), 321–354.

Edo, A. & Rapoport, H. (2019), 'Minimum wages and the labor market effects of immigration', *Labour Economics* **61**, 101753.

Felbermayr, G. J. & Toubal, F. (2012), 'Revisiting the Trade-Migration Nexus: Evidence from New OECD Data', *World Development* **40**(5), 928–937.

Fitzgerald, D. & Haller, S. (2018), 'Exporters and shocks', *Journal of International Economics* **113**, 154 – 171.

Frankel, J. A. & Romer, D. H. (1999), 'Does Trade Cause Growth?', *American Economic Review* **89**(3), 379–399.

Goldsmith-Pinkham, P., Sorkin, I. & Swift, H. (2020), 'Bartik instruments: What, when, why, and how', *American Economic Review* **110**(8), 2586–2624.

Gould, D. M. (1994), 'Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows', *The Review of Economics and Statistics* **76**(2), 302–316.

Gray, C. L. & Mueller, V. (2012), 'Natural disasters and population mobility in bangladesh', *Proceedings of the National Academy of Sciences* **109**(16), 6000–6005.

Guimares, P., Figueirdo, O. & Woodward, D. (2003), 'A Tractable Approach to the Firm Location Decision Problem', *The Review of Economics and Statistics* **85**(1), 201–204.

Head, K. & Mayer, T. (2014), Gravity Equations: Workhorse,Toolkit, and Cookbook, *in* G. Gopinath, . Helpman & K. Rogoff, eds, 'Handbook of International Economics', Vol. 4 of *Handbook of International Economics*, Elsevier, chapter 0, pp. 131–195.

Head, K. & Ries, J. (1998), 'Immigration and Trade Creation: Econometric Evidence from Canada', *Canadian Journal of Economics* **31**(1), 47–62.

Helpman, E., Melitz, M. J. & Yeaple, S. R. (2004), 'Export Versus FDI with Heterogeneous Firms', *American Economic Review* **94**(1), 300–316.
URL: *https://ideas.repec.org/a/aea/aecrev/v94y2004i1p300-316.html*

Hidalgo, C. A. & Hausmann, R. (2009), 'The building blocks of economic complexity', *Proceedings of the National Academy of Sciences* **106**(26), 10570–10575.

Hong, L. & Page, S. E. (2001), 'Problem Solving by Heterogeneous Agents', *Journal of Economic Theory* **97**(1), 123–163.

Hoogendoorn, S. M. & van Praag, M. C. (2012), Ethnic Diversity and Team Performance: A Field Experiment, IZA Discussion Papers 6731, Institute for the Study of Labor (IZA).

Hornung, E. (2014), 'Immigration and the diffusion of technology: The huguenot diaspora in prussia', *American Economic Review* **104**(1), 84–122.

Horwitz, S. & Horwitz, I. (2007), 'The effects of team diversity on team outcomes: A meta-analytic review of team demography', *Journal of Management* **33**(6), 987–1015.

Jaeger, D. A., Ruist, J. & Stuhler, J. (2018), Shift-Share Instruments and the Impact of Immigration, NBER Working Papers 24285, National Bureau of Economic Research, Inc.

Javorcik, B. S., Özden, C., Spatareanu, M. & Neagu, C. (2011), 'Migrant networks and foreign direct investment', *Journal of Development Economics* **94**(2), 231–241.

Kahane, L., Longley, N. & Simmons, R. (2013), 'The Effects of Coworker Heterogeneity on Firm-Level Output: Assessing the Impacts of Cultural and Language Diversity in the National Hockey League', *The Review of Economics and Statistics* **95**(1), 302–314.

Kugler, M., Levintal, O. & Rapoport, H. (2018), 'Migration and Cross-Border Financial Flows', *World Bank Economic Review* **32**(1), 148–162.

Kugler, M. & Rapoport, H. (2007), 'International labor and capital flows: Complements or substitutes?', *Economics Letters* **94**(2), 155–162.

Lall, S. (2000), 'The Technological Structure and Performance of Developing Country Manufactured Exports, 1985-98', *Oxford Development Studies* **28**(3), 337–369.

Lazear, E. P. (1999), 'Globalisation and the Market for Team-Mates', *Economic Journal* **109**(454), 15–40.

Leblang, D. (2010), 'Familiarity Breeds Investment: Diaspora Networks and International Investment', *American Political Science Review* **104**(3), 584–600.

Leromain, E. & Orefice, G. (2014), 'New revealed comparative advantage index: Dataset and empirical distribution', *International Economics* **139**(3), 48–70.

Maggi, G. & Grossman, G. M. (2000), 'Diversity and Trade', *American Economic Review* **90**(5), 1255–1275.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka, ed., 'Frontier in Econometrics', Vol. 1, Academic Press, New York, chapter 5.

Melitz, J. & Toubal, F. (2014), 'Native language, spoken language, translation and trade', *Journal of International Economics* **93**(2), 351–363.

Montalvo, J. G. & Reynal-Querol, M. (2005), 'Ethnic polarization, potential conflict, and civil wars', *The American Economic Review* **95**(3), 796–816.

Moser, P. & San, S. (2019), Immigration, Science, and Invention: Evidence from the Quota Acts, mimeo.

Ortega, F. & Peri, G. (2014), 'Openness and income: The roles of trade and migration', *Journal of International Economics* **92**(2), 231–251.

Ottaviano, G. I. P. & Peri, G. (2012), 'Rethinking The Effect Of Immigration On Wages', *Journal of the European Economic Association* **10**(1), 152–197.

Ottaviano, G. I. & Peri, G. (2006), 'The economic value of cultural diversity: evidence from US cities', *Journal of Economic Geography* **6**(1), 9–44.

Ozden, C., Parsons, C. R., Schiff, M. & Walmsley, T. L. (2011), 'Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000', *World Bank Economic Review* **25**(1), 12–56.

Parrotta, P., Pozzoli, D. & Sala, D. (2016), 'Ethnic diversity and firms' export behavior', *European Economic Review* **89**(C), 248–263.

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

Peri, G. & 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.

Peri, G. & Sparber, C. (2009), 'Task Specialization, Immigration, and Wages', *American Economic Journal: Applied Economics* **1**(3), 135–169.

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

Rauch, J. E. (2001), 'Business and Social Networks in International Trade', *Journal of Economic Literature* **39**(4), 1177–1203.

Rauch, J. E. & Trindade, V. (2002), 'Ethnic Chinese Networks In International Trade', *The Review of Economics and Statistics* **84**(1), 116–130.

Rodriguez-Pose, A. & Berlepsch, V. (2019), 'Does Population Diversity Matter for Economic Development in the Very Long Term? Historic Migration, Diversity and County Wealth in the US', *European Journal of Population* **35**(5), 873–911.

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

Trax, M., Brunow, S. & Suedekum, J. (2015), 'Cultural diversity and plant-level productivity', *Regional Science and Urban Economics* **53**(C), 85–96.

United Nations, P. D. (2015), Trends in international migrant stock: Migrants by destination and origin. the 2015 revision, Technical report, United Nations Department of Economic and Social Affairs Population Division, New York.

Vivian, H., Dennis, L. & Sara, P. (2015), Diversity matters, Technical report, McKinsey&Company.

Wooldridge, J. (2012), *Introductory Econometrics. A Modern Approach*, South-Western.

Yotov, Y. V., Piermartini, R., Monteiro, J.-A. & Larch, M. (2017), *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, United Nations publication, World Trade Organization.

Tables and Figures

Table 1: Migration Stocks and Diversity.

	Year 1995		Year 2005		Year 2015	
	Mig Stock$_{it}$	Birth. Div.$_{it}$	Mig Stock$_{it}$	Birth. Div.$_{it}$	Mig Stock$_{it}$	Birth. Div.$_{it}$
United States	24.41	0.91	34.27	0.90	41.28	0.90
Germany	7.07	0.92	10.00	0.93	11.81	0.93
Russia	11.73	0.84	11.47	0.85	11.44	0.85
Saudi Arabia	4.92	0.90	6.28	0.90	9.84	0.90
UK	4.03	0.95	5.68	0.97	8.14	0.97
UAE	1.77	0.81	3.16	0.78	7.97	0.76
Canada	4.81	0.96	6.00	0.96	7.69	0.96
France	6.06	0.93	6.70	0.92	7.62	0.93
Australia	4.09	0.91	4.75	0.92	6.56	0.94
Spain	1.01	0.93	4.09	0.95	5.81	0.95

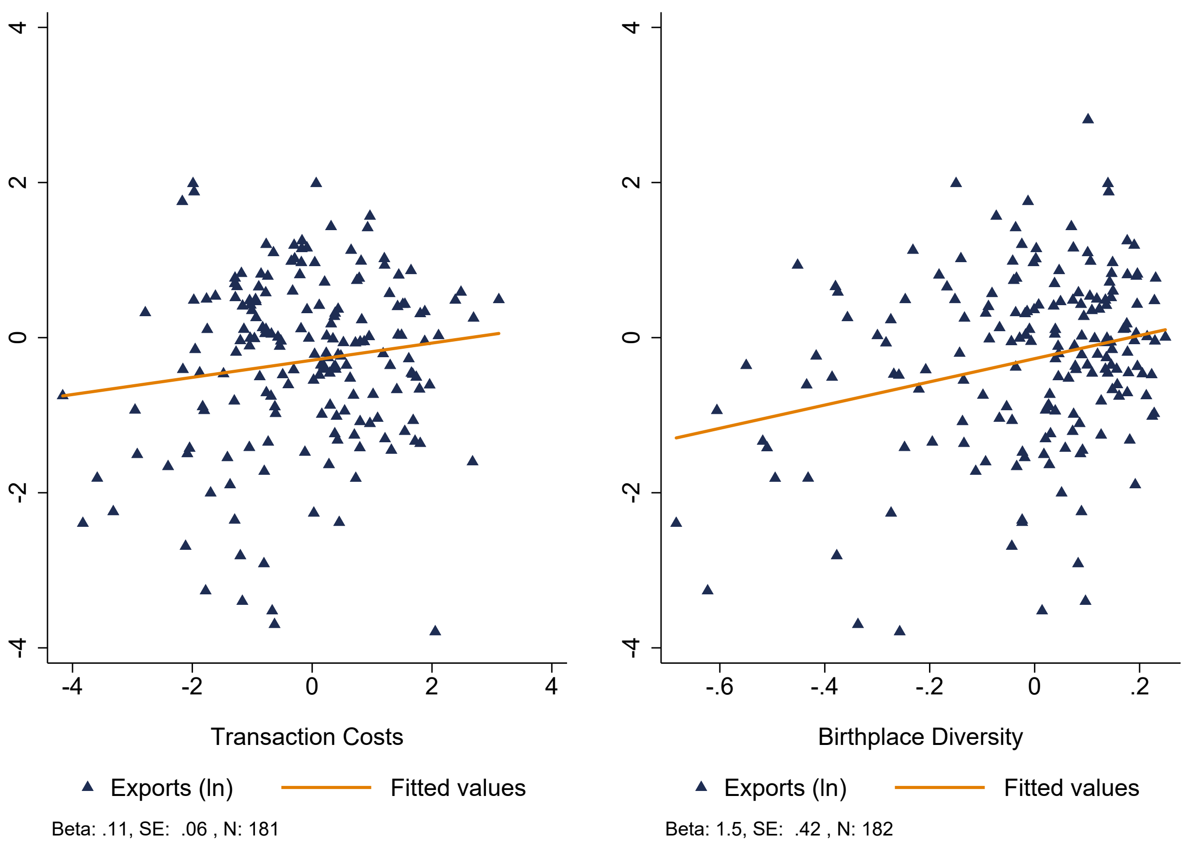
Note: Mig Stock$_{it}$ reports the stock of foreign born residents in millions. Birthplace Diversity calculated as reported in section 2.

Table 2: In-sample descriptive statistics.

Variable	N	Mean	Sd	p25	p50	p75
Bilateral Sample:						
log(Exports)	4,575,395	4.33	3.33	2.01	4.31	6.65
log(Immigrants)	4,575,395	4.81	4.04	0.00	5.40	8.06
KD	4,575,395	0.29	0.26	0.08	0.22	0.44
Diversity	4,575,395	0.78	0.18	0.69	0.84	0.92
Aggregate Sample:						
log(Exports)	116,268	7.78	3.99	4.88	7.92	10.81
log(Immigrants)	116,268	12.11	2.10	10.67	12.27	13.57
Diversity	116,268	0.71	0.20	0.60	0.76	0.85

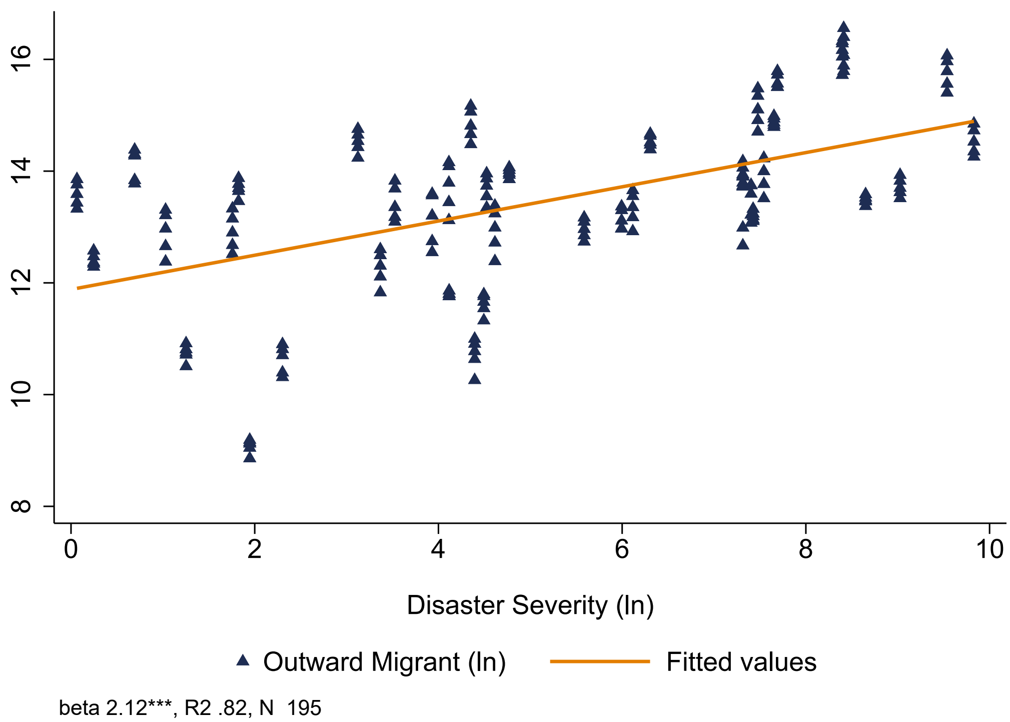
Note: Data based on 5-year intervals from 1995 to 2015.

Figure 1: Un-conditional correlations between exports and the migration-related competitiveness channels (i.e transaction cost approximated by migrants stock and birthplace diversity). Year 2015.



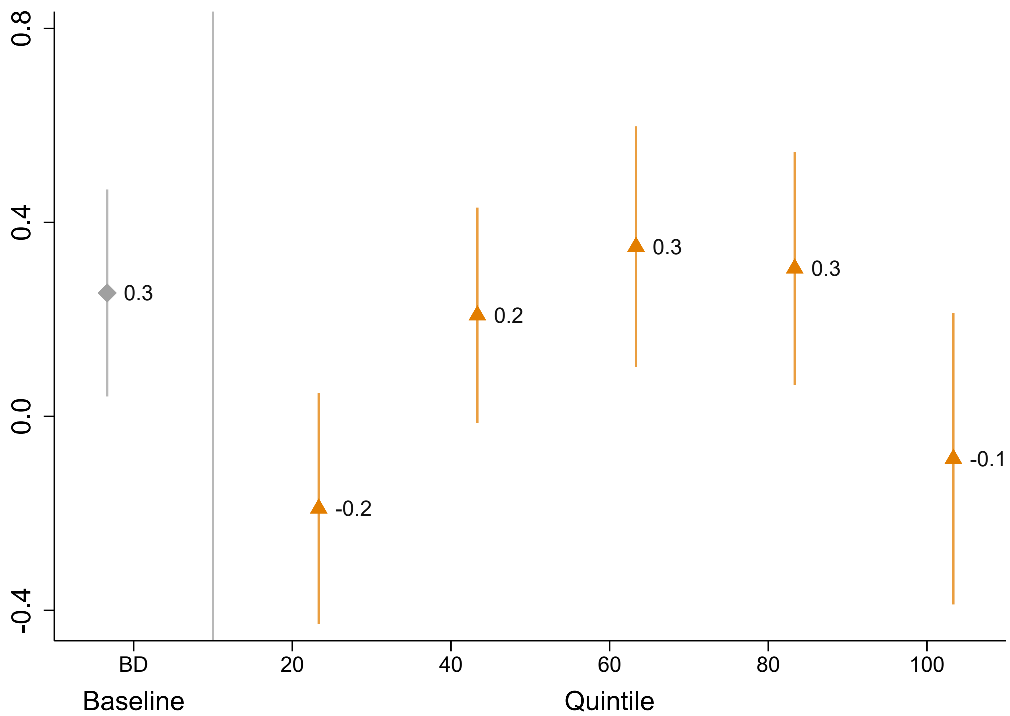
Note: scatter plot between aggregate country specific exports and immigrant-related channels. From Equation 1 transaction costs are captured by $ln(Mig_{ijt})$ - scatter on the left; Birthplace diversity by BD_{it} - scatter on the right. Both graphs control for country size, i.e. $ln(GDP)_{i,t-5}$. Source: Authors calculation on BACI (CEPII) and United Nations (2015) data.

Figure 2: Natural Disasters and Outward Migration.



Note: regression of outward migration stock on number of natural disasters in the previous decade (log-log specification, 41 countries, 5 years). Source: EM-DAT: The Emergency Events Database.

Figure 3: Non-linear relationship between exports and birthplace diversity. OLS BD_{it} estimations by quartile in language dissimilarity index between destination i and all its origins js.



Note: quintiles based on country-year specific average index of language dissimilarity.

Table 3: Baseline estimation results, OLS and 2SLS. Extensive margin.

Dep var	$Dummy = 1$ if $Export_{ijkt} > 0$					
	(1)	(2)	(3)	(4)	(5)	(6)
Mig_{ijt}	0.026***	0.026***	0.018***	0.018***	0.023***	0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Know. Diffusion_{ikt}	0.013***	0.012***	0.036***	0.036***	0.011**	0.037***
	(0.003)	(0.003)	(0.002)	(0.004)	(0.005)	(0.005)
Birth. Divers._{it}	0.076***	0.076***			0.099***	
	(0.012)	(0.012)			(0.024)	
Birth. $\text{Divers.}_{it} \times \text{Abstract}_k$		0.014***	0.033***	0.033***		0.036***
		(0.003)	(0.004)	(0.008)		(0.009)
RTA_{ijt}	0.024***	0.024***	0.012***	0.012**	0.025***	0.013**
	(0.005)	(0.005)	(0.004)	(0.006)	(0.007)	(0.006)
$\ln(1+\text{Tariff})_{ijkt}$	0.044***	0.045***	-0.033**	-0.033*	0.041*	-0.033*
	(0.016)	(0.016)	(0.014)	(0.018)	(0.022)	(0.018)
$\ln(\text{Remoteness})_{it}$	0.240***	0.240***			0.235***	
	(0.033)	(0.033)			(0.064)	
$\ln(\text{Emigrants})_{ijt}$	0.024***	0.024***	0.014***	0.014***	0.024***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Quartile II $\text{Exports}_{ikt;-jr}$	0.070***	0.070***	0.022***	0.022***	0.071***	0.022***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.006)	(0.003)
Quartile III $\text{Exports}_{ikt;-jr}$	0.176***	0.176***	0.054***	0.054***	0.177***	0.054***
	(0.005)	(0.005)	(0.003)	(0.006)	(0.010)	(0.006)
Quartile IV $\text{Exports}_{ikt;-jr}$	0.361***	0.360***	0.126***	0.126***	0.363***	0.126***
	(0.007)	(0.007)	(0.005)	(0.010)	(0.014)	(0.010)
Estimator	OLS	OLS	OLS	OLS	2SLS	2SLS
IV					Predicted Supply	Predicted Supply
FE: θ_{ij}	yes	yes	yes	yes	yes	yes
FE: θ_{jkt}	yes	yes	yes	yes	yes	yes
FE: θ_{rckt}	yes	yes	no	no	yes	no
FE: θ_{it}	no	no	yes	yes	no	yes
Cluster std err	it jt	it jt	it jt	i j	i j	i j
Observations	20,156,093	20,156,093	20,156,093	20,156,093	20,156,093	20,156,093
R-squared	0.578	0.578	0.595	0.595	0.138	0.018
F-stat First Stage					870.7	572.1

Note: $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 195$, Importers $j = 176$, Sectors (SIC72) $k = 142$, Year $t = 5$ (five-year window). IV for migration stock, knowledge diffusion and birthplace diversity based on the supply-driven predicted migration stocks.

Table 4: Baseline estimation results, OLS and 2SLS. Intensive margin.

Dep var	$Ln(export_{ijkt})\|_{export_{ijk(t-1)}>0}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Mig_{ijt}	0.153***	0.153***	0.106***	0.106***	0.174***	0.145***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.008)	(0.009)
Know. Diffusion_{ikt}	0.434***	0.420***	0.828***	0.828***	0.444***	0.857***
	(0.054)	(0.055)	(0.052)	(0.099)	(0.112)	(0.100)
Birth. Divers._{it}	0.247**	0.247**			0.614***	
	(0.123)	(0.123)			(0.235)	
Birth. $\text{Divers.}_{it} \times \text{Abstract}_k$		0.168**	0.748***	0.748***		0.785***
		(0.081)	(0.080)	(0.152)		(0.163)
RTA_{ijt}	0.214***	0.214***	0.181***	0.181***	0.188***	0.154***
	(0.032)	(0.032)	(0.031)	(0.040)	(0.044)	(0.039)
$\ln(1+\text{Tariff})_{ijkt}$	-1.568***	-1.563***	-2.371***	-2.371***	-1.557***	-2.319***
	(0.171)	(0.171)	(0.217)	(0.351)	(0.252)	(0.352)
$\ln(\text{Remoteness})_{it}$	2.793***	2.781***			2.813***	
	(0.402)	(0.402)			(0.729)	
$\ln(\text{Emigrants})_{ijt}$	0.170***	0.170***	0.099***	0.099***	0.173***	0.111***
	(0.004)	(0.004)	(0.004)	(0.007)	(0.008)	(0.007)
Quartile II $\text{Exports}_{ikt;-jr}$	1.382***	1.380***	0.981***	0.981***	1.365***	0.983***
	(0.042)	(0.042)	(0.038)	(0.065)	(0.070)	(0.065)
Quartile III $\text{Exports}_{ikt;-jr}$	2.672***	2.671***	1.972***	1.972***	2.631***	1.972***
	(0.053)	(0.053)	(0.059)	(0.110)	(0.091)	(0.109)
Quartile IV $\text{Exports}_{ikt;-jr}$	4.117***	4.116***	3.296***	3.296***	4.063***	3.292***
	(0.061)	(0.061)	(0.080)	(0.154)	(0.104)	(0.153)
Estimator	OLS	OLS	OLS	OLS	2SLS	2SLS
IV					Predicted Supply	Predicted Supply
FE: θ_{ij}	yes	yes	yes	yes	yes	yes
FE :θ_{jkt}	yes	yes	yes	yes	yes	yes
FE: θ_{rckt}	yes	yes	no	no	yes	no
FE: θ_{it}	no	no	yes	yes	no	yes
Cluster std err	it jt	it jt	it jt	i j	i j	i j
Observations	4,575,395	4,575,395	4,575,395	4,575,395	4,575,395	4,575,395
R-squared	0.709	0.709	0.706	0.706	0.169	0.098
F-stat First Stage					617.2	401.9

Note: $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 195$, Importers $j = 176$, Sectors (SIC72) $k = 142$, Year $t = 5$ (five-year window). IV for migration stock, knowledge diffusion and birthplace diversity based on the supply-driven predicted migration stocks.

Table 5: Bilateral regressions. OLS and 2SLS using standardized Variables.

Dep var	$Dummy = 1$ if $Export_{ijkt} > 0$		$\text{Ln(export)}\|_{export(t-1)>0}$	
	(1)	(2)	(3)	(4)
Standardized Mig_{ijt}	0.179***	0.163***	0.186***	0.211***
	(0.010)	(0.010)	(0.008)	(0.010)
Standardized Know. Diffusion_{ikt}	0.007**	0.006**	0.033***	0.034***
	(0.003)	(0.003)	(0.008)	(0.009)
Standardized Birth. Divers._{it}	0.031***	0.040***	0.013	0.033***
	(0.008)	(0.010)	(0.011)	(0.012)
Estimator	OLS	2SLS	OLS	2SLS
IV		Predicted Supply		Predicted Supply
Controls: $\mathbf{X}_{ijkt}$	yes	yes	yes	yes
FE: θ_{ij}	yes	yes	yes	yes
FE: θ_{jkt}	yes	yes	yes	yes
FE: θ_{rckt}	yes	yes	yes	yes
Cluster std err	i j	i j	i j	i j
Observations	20,156,093	20,156,093	4,575,395	4,575,395
R-squared	0.578	0.138	0.709	0.169
F-stat First Stage		870.7		617.2

Note: $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 195$, Importers $j = 176$, Sectors (SIC72) $k = 142$, Year $t = 5$. All regressions include the full set of bilateral controls included in eq. (1). IV for migration stock, knowledge diffusion and birthplace diversity index are based on the supply-driven predicted migration stocks.

Table 6: First stage results of baseline 2SLS estimation using IVs based on the supply-driven predicted migration stocks. Specification with exporter-year fixed effects.

Dep var	$Dummy = 1$ if $Export_{ijkt} > 0$			$\text{Ln(export)}\vert_{export_{ijk(t-1)}>0}$		
	Mig_{ijt}	Know. Diffusion_{ikt}	Birth. Div._{it} $\times$ Abstract_k	Mig_{ijt}	Know. Diffusion_{ikt}	Birth. Div._{it} $\times$ Abstract_k
	(1)	(2)	(3)	(4)	(5)	(6)
IV: Mig_{ijt}	0.753***	-0.000***	0.000	0.651***	-0.000	0.000
	(0.018)	(0.000)	(0.000)	(0.018)	(0.000)	(0.000)
IV: Know. Diffusion_{ikt}	-0.011***	0.996***	0.003*	-0.054***	0.993***	0.000
	(0.003)	(0.002)	(0.002)	(0.009)	(0.003)	(0.002)
IV: Birth. Divers._{it} $\times$ Abstract_k	0.000	-0.001	0.956***	0.040***	-0.001	0.971***
	(0.001)	(0.001)	(0.011)	(0.007)	(0.002)	(0.010)
Controls: $\mathbf{X}_{ijkt}$	yes	yes	yes	yes	yes	yes
FE: θ_{ij}	yes	yes	yes	yes	yes	yes
FE: θ_{jkt}	yes	yes	yes	yes	yes	yes
FE: θ_{it}	yes	yes	yes	yes	yes	yes
Cluster std err	i j	i j	i j	i j	i j	i j
Observations	20,156,093	20,156,093	20,156,093	4,575,395	4,575,395	4,575,395
Kleibergen-Paap F-stat		572.1			401.9	

Note: This table reports the first stage regressions for specification in column (6) in tables 3 and 4. In all regressions standard errors in parentheses are double clustered at exporter and importer country level. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 195$, Importers $j = 176$, Sectors (SIC72) $k = 142$, Year $t = 5$ (5-year window).

Table 7: 2SLS regressions. Alternative Instruments.

Dep var	$Dummy = 1$ if $Export_{ijkt} > 0$				$\text{Ln(export)}\vert_{export_{ijk(t-1)}>0}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mig_{ijt}	0.017***	0.015***	0.017***	0.036**	0.146***	0.147***	0.125***	0.191**
	(0.001)	(0.002)	(0.001)	(0.016)	(0.009)	(0.012)	(0.009)	(0.085)
Know. Diffusion_{ikt}	0.037***	0.034***	0.036***	0.077***	0.857***	0.917***	0.843***	1.377***
	(0.005)	(0.007)	(0.004)	(0.023)	(0.100)	(0.125)	(0.099)	(0.284)
Birth. $\text{Divers.}_{\cdot it} \times \text{Abstract}_k$	0.036***	0.044***	0.035***	0.066*	0.786***	0.987***	0.754***	2.808***
	(0.009)	(0.013)	(0.008)	(0.038)	(0.163)	(0.224)	(0.155)	(1.025)
IV	No feedback effect	Natural Disaster	Segmented Labor mkt	Cumulated imputed mig flows	No feedback effect	Natural Disaster	Segmented Labor mkt	Cumulated imputed mig flows
Controls: $\mathbf{X}_{ijkt}$	yes	yes	yes	yes	yes	yes	yes	yes
FE: θ_{ij}	yes	yes	yes	yes	yes	yes	yes	yes
FE: θ_{jkt}	yes	yes	yes	yes	yes	yes	yes	yes
FE: θ_{it}	yes	yes	yes	yes	yes	yes	yes	yes
Cluster std err	i j	i j	i j	i j	i j	i j	i j	i j
IV: Mig_{ijt}	0.752***	0.542***	0.789***	0.171***	0.650***	0.437***	0.703***	0.181***
IV: Know. Diffusion_{ikt}	0.996***	0.375***	1.003***	0.153***	0.993***	0.452***	1.000***	0.136***
IV: Birth. $\text{Div.}_{\cdot it} \times \text{Abstr.}_k$	0.955***	0.604***	-0.979***	0.178***	0.970***	0.713***	-0.982***	0.210***
Observations	20,156,093	20,156,093	20,156,093	20,156,093	4,575,395	4,575,395	4,575,395	4,575,395
R-squared	0.018	0.018	0.018	0.010	0.098	0.097	0.098	0.056
F-test	575.3	179.8	733.7	9.3	401.4	150	471.1	14.0

Note: In all regressions standard errors in parentheses are double clustered at exporter and importer country level. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 195$, Importers $j = 176$, Sectors (SIC72) $k = 142$, Year $t = 5$ (5-year window).

Table 8: Results using teamwork intensity as a proxy for problem solving intensity. OLS and 2SLS results.

Dep var	$Dummy = 1$ if $Export_{ijkt} > 0$			$\text{Ln(export)}\vert_{export_{ijk(t-1)}>0}$		
	(1)	(2)	(3)	(4)	(5)	(6)
Mig_{ijt}	0.026***	0.018***	0.017***	0.153***	0.106***	0.145***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.009)
Know. Diffusion_{ikt}	0.013***	0.037***	0.037***	0.422***	0.827***	0.857***
	(0.003)	(0.002)	(0.005)	(0.055)	(0.052)	(0.100)
Birth. Divers._{it}	0.076***			0.259**		
	(0.012)			(0.124)		
Birth. $\text{Divers.}_{it} \times \text{Team}_k$	0.005	0.122***	0.125***	0.979*	4.152***	4.328***
	(0.020)	(0.023)	(0.048)	(0.509)	(0.456)	(0.911)
Estimator	OLS	OLS	2SLS	OLS	OLS	2SLS
IV			Predicted supply			Predicted supply
Controls: $\mathbf{X}_{ijkt}$	yes	yes	yes	yes	yes	yes
FE: θ_{ij}	yes	yes	yes	yes	yes	yes
FE: θ_{jkt}	yes	yes	yes	yes	yes	yes
FE: θ_{rckt}	yes	no	no	yes	no	no
FE: θ_{it}	no	yes	yes	no	yes	yes
Cluster std err	it jt	it jt	i j	it jt	it jt	i j
IV: Mig_{ijt}			0.753***			0.651***
IV: Know. Diffusion_{ikt}			0.996***			0.993***
IV: Birth. $\text{Divers.}_{it} \times \text{Team}_k$			0.956***			0.972***
Observations	19,591,612	19,591,612	19,591,612	4,501,275	4,501,275	4,501,275
R-squared	0.578	0.597	0.017	0.710	0.708	0.097
F-test			572.3			403.5

Note: In all regressions standard errors in parentheses are double clustered at exporter and importer country level. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 195$, Importers $j = 176$, Sectors (SIC72) $k = 138$, Year $t = 5$ (5-year windows). IV for migration stock, knowledge diffusion and birthplace diversity based on the supply-driven predicted migration stocks.

Table 9: Regression by Skill Level (OECD, DIOC-E database). OLS estimations.

Dep var	*Dummy* = 1 if $Export_{ijkt} > 0$					$\text{Ln(exp)}\vert_{exp_{ijk(t-1)}>0}$
	(1)	(2)	(3)	(4)	(5)	(6)
Mig_{ijt}	0.032***	0.032***	0.031***	0.031***	0.031***	0.177***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.024)
Know. Diffusion_{ikt}	0.041***	0.039***	0.040***	0.039***	0.039***	0.275
	(0.013)	(0.012)	(0.013)	(0.013)	(0.014)	(0.192)
Birth. $\text{Divers.}_{it}^{Primary}$	0.047			0.078	0.078	-0.032
	(0.035)			(0.063)	(0.064)	(1.054)
Birth. $\text{Divers.}_{it}^{Secondary}$		0.022		-0.245*	-0.245	-2.395*
		(0.047)		(0.132)	(0.155)	(1.383)
Birth. $\text{Divers.}_{it}^{Tertiary}$			0.101**	0.278**	0.278**	3.885***
			(0.049)	(0.109)	(0.121)	(1.310)
Controls: $\mathbf{X}_{ijkt}$	yes	yes	yes	yes	yes	yes
FE: θ_{jk}	yes	yes	yes	yes	yes	yes
FE: θ_{rck}	yes	yes	yes	yes	yes	yes
Cluster std err	it jt	it jt	it jt	it jt	i j	i j
Observations	2,001,680	2,001,680	2,001,680	2,001,680	2,001,680	1,036,479
R-squared	0.528	0.528	0.529	0.529	0.529	0.663

Note: $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 114$, Importers $j = 174$, Sectors (SIC72) $k = 142$, Year $t = 2$ (i.e. 2000, 2010).

Table 10: Country-sector aggregate results. Results by abstract intensity of tasks. OLS estimations.

Dep var	RCA_{ikt}			$Export_{ikt}$	$\# \ Dest_{ikt}$
	(1)	(2)	(3)	(4)	(5)
Birth. Divers.$_{it}$	0.438***	0.440***			
	(0.064)	(0.066)			
Birth. Divers.$_{it}$ $\times$ $Abstr._k$		0.200***	0.208***	0.400***	-0.008
		(0.044)	(0.045)	(0.047)	(0.019)
Know. Diff.$_{ikt}$	0.474***	0.466***	0.461***	1.168***	0.237***
	(0.040)	(0.040)	(0.040)	(0.050)	(0.015)
Mig$_{it}$	0.038*	0.039*			
	(0.020)	(0.020)			
$FTA^{\#}_{i,t-5}$	0.006***	0.006***			
	(0.001)	(0.001)			
$ln(GDP)_{i,t-5}$	0.307***	0.307***			
	(0.023)	(0.023)			
FE: θ_i	yes	yes	no	no	no
FE: θ_{it}	no	no	yes	yes	yes
FE: θ_{kt}	yes	yes	yes	yes	yes
Cluster std err	ik kt	ik kt	ik kt	ik kt	ik kt
Observations	114,262	114,262	114,262	114,262	114,262
R-squared	0.742	0.742	0.763	0.848	0.876

Note: In all regressions standard errors in parentheses are double clustered at country-sector and sector-year. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 186$, Sectors (SIC72) $k = 142$, Year $t = 5$.

A Trade and Diversity: insights from Maggi and Grossman (2000)

According to the theoretical model in Maggi & Grossman (2000), the diversity of a country's workforce composition (i.e., dispersion of workers' *skills*) improves the export performance of sectors characterized by sub-modular production functions, that is, where a more dispersed distribution of skills allows high-talented workers to be paired with workers at the opposite end of the ability distribution, thereby maximizing the productivity of the production process. Maggi & Grossman (2000) propose a theoretical model in which two countries have different distributions of workers' abilities (one being more disperse than the other) and produce two types of goods: (i) one good characterized by a *super-modular* production function; and (ii) the other by a *sub-modular* production function. In presence of super-modular technology, performing better at one stage of production raises the marginal value of a better performance in another stage. This is the case for industries in which cost-effectiveness in a long sequence of operations contributes to the success of the overall production process (e.g., the automotive industry is an example of super-modular sector). In presence of sub-modular technology, performing better one step of production mitigates the need for better performance in another step. This is the case of industries requiring creativity and problem solving abilities (such as fashion, design, or cultural goods), where the overall success of the production process strongly depends on the presence of extremely brilliant workers in production (i.e., when the marginal value of having a more able worker increases when the other co-workers in production have a lower ability).

Under these assumptions, Maggi & Grossman (2000) show that the country with a more dispersed skill distribution, by having higher possibilities for matching extremely brilliant workers with more modest workers in production, will have a comparative advantage in the sector characterized by sub-modular technology, where creativity and problem-solving are more needed. In the same vein, Bombardini et al. (2014) propose a theoretical model in which all sectors feature super-modular production functions, but differ in the degree of skill complementarity.[63] In their model, the output of firms in each sector depends on the mass of employees and on a productivity factor based on the distribution of skills in the country. As a result, countries with higher dispersion of skills have a comparative advantage in sectors with lower skill complementarity.

The theoretical predictions in both Maggi & Grossman (2000) and Bombardini et al. (2014) are based on a *vertical* notion of workers' ability dispersion: workers are assumed to be vertically ranked based on their skills (with high-talented workers being more productive than less-talented workers). In the case of birthplace diversity, workers originating from a diverse set of countries are likely to have similar hard-skills (education or technical knowledge) but still be imperfect substitute in production (Ottaviano & Peri 2012) because they are endowed with different problem-solving capabilities and soft-skills. Therefore, host countries with a higher

[63] In Bombardini et al. (2014) the output of a sector depends on a specific skill a, on the mass of workers with the same skill $h(a)$ and a parameter λ measuring the skill complementarity in the sector: $y = \left(\int a^{\lambda} h(a) da \right)^{1/\lambda}$

birthplace diversity index will have a more disperse distribution of *horizontally* differentiated problem solving capabilities and soft-skills (not vertically rankable).[64] We thus imagine a production process of one task performed by two workers having the same hard-skills but potentially different problem solving capabilities and different soft-skills in general.[65] In the case of sectors characterized by sub-modular production functions, overall productivity is maximized when the firms have the possibility to combine horizontally differentiated workers (i.e., when the pool of workers available in the local labor market is more diverse). This is also consistent with the broader idea/intuition by Hong & Page (2001) that a more horizontally differentiated workforce is particularly beneficial in high-complex sectors requiring creativity, different problem solving approaches and soft-skills. So, while we depart from the theoretical framework in Maggi & Grossman (2000) and Bombardini et al. (2014) concerning the nature of the workers' ability dispersion (*vertical* vs *horizontal*), we can still conjecture about the positive effect of birthplace diversity on the international competitiveness (i.e. productivity) of countries, with magnified effect for sectors characterized by sub-modular production functions:

Conjecture 1 *Birthplace Diversity is expected to improve the export performance of host countries, and more so in sectors requiring cognitive and problem solving capabilities.*

Based on Conjecture 1, in Section 4 we test the average effect of birthplace diversity on the competitiveness of countries and the underlying mechanism (controlling for the two other migration-related channels). Namely, we test whether the international competitiveness effect of birthplace diversity is particularly valid for problem solving intensive sectors (as proxied by a sector's abstract and teamwork intensity).[66]

[64] Another simple way to conceptualize the role of migration in improving the soft-skill dispersion of workers at destination is by assuming that workers from different origins are different factors of production, in the vein of the Armington model for trade (see appendix A in Ortega & Peri (2014) for a more detailed discussion). Indeed, migrants from different origins differ in terms of their language, culture and social norms.

[65] This can be conceptually extended to the case of a production process composed of n additive tasks.

[66] We thanks Matilde Bombardini for sharing data on O*NET-based measures of teamwork intensity at the sector level.

B Additional descriptive evidence

Natural disaster data. We detail here the countries affected by a severe natural disaster that we use as source of identification in the natural experiment based instrumental variable discussed in section 2.2.4. We define a disaster as severe if it causes both economic and social disruptions. Using the total damages, number of people affected and total casualties as proxy for the economic and social impact of a catastrophic event, table B1 identifies the 41 countries affected by a severe event during the pre-sample period, 1985-1990. Data on severe natural disaster are from the EM-DAT (Emergency Events Database) and consider only natural events: biological (epidemic), climatological (drought, wildfire), geophysical (mass movement, earthquake, volcanic activity), hydrological (flood, landslide), meteorological (storm, fog, extreme temperature).

Region-income level fixed effects. In order to create fixed effects for the macro-region and income level of each exporting country i, we attached to each exporter its macro-region and income level based on the World Bank classification. We therefore have seven macro-region and 4 income-level to characterize each country. The number of countries belonging to each region-income level cell are reported in table B2.

Table B1: List of countries affected by (severe) natural disasters, 1985-1990

Countries	Events	Damages (US$, Mln)	Affected	Casualties
ARG	6	1640	1416990	44
ATG	1	80	8030	2
AUS	6	265.839	1012	15
BEN	1	4.8	475000	61
BGD	14	2187	57905460	19561
BOL	1	50	310000	29
BRA	13	1886	3752961	884
CAN	2	117	1000	12
CHL	5	1678	1684781	344
CHN	92	13868.94	280067742	11680
COM	1	9	50000	24
CRI	4	88.5	154609	36
DZA	2	1	15000	56
ECU	4	1500	166006	5102
FSM	1	6	203	5
GLP	1	50	11084	5
HKG	4	0.067	3512	12
HND	1	100	48000	5
HTI	5	91.286	873901	81
IDN	22	76.641	285250	832
IND	39	4498.843	21765519	7590
IRN	7	8311.7	884117	40142
ITA	5	2105	2716	27
JAM	1	5.2	300	7
JPN	3	5713	148366	67
KOR	3	547	210000	669
MEX	6	4430.6	2255204	9811
MSR	1	240	12040	11
MWI	3	28	150544	57
NIC	1	400	360278	130
PAN	2	60.35	14732	32
PER	11	60.2	2515946	412
PHL	53	1766.393	22974707	6554
SLV	1	1500	770000	1100
THA	1	452	199000	458
TON	1	2.5	3103	1
TZA	3	0.28	162868	389
USA	34	18574.1	1055222	634
VEN	4	1.8	18029	139
VNM	8	21.725	6929667	1343
YEM	1	33	340000	25

Note: The table reports the total number of (severe) natural disasters over the period 1985-1990 by country, along with the amount of damages, in milions US$, the number of affected residents and total number of casualties. Source: EM-DAT: The Emergency Events Database - Université catholique de Louvain (UCL) - CRED, D. Guha-Sapir - www.emdat.be, Brussels, Belgium.

Table B2: List of countries by region-income cell

	Income Level					
Region	High	Up-Middle	Low-Middle	Low	N.e.s.	Total
East Asia & Pacific	7	4	15	7	0	33
Europe & Central Asia	18	12	19	0	0	49
Latin America & Caribbean	4	10	20	2	0	36
Middle East & North Africa	6	6	9	0	0	21
North America	3	0	0	0	0	3
South Asia	0	0	0	8	0	8
Sub-Saharan Africa	0	2	9	31	0	42
Nes	0	0	0	0	3	3
Total	38	34	72	48	3	195

Note: The table reports the total number of countries per region-income cell. Both regions and income levels are from the World Bank. Income levels refers to the first available year reported in the World Bank database: 1987 (151 countries); between 1987-1994 (37 countries); PLW (1996); SRB (2006); TCA and TUV (2009). The category "Nes" includes 3 countries are not included neither in the region nor in the income database: GIB, NRU, VGB.

C Additional Robustness Checks

Back-of-the-envelope calculation. In order to quantify the relative contribution of each migration-related channel we computed a simple back-of-the-envelope calculation. We compute the expected value of trade for 2015 from the estimates reported in column 4 of table 4 (our baseline) and then the expected value of trade from the same regression but keeping the three variables of interest (i.e. Mig_{ijt}, KD_{ikt} and BD_{it}) at their 2005 level. We then infer the contribution of each channel by comparing the associated change in expected trade with respect to the baseline. Results are reported in table C1, where the contribution of each migration-related channel can be also compared with the observed average export growth (in column 1). Out of the 52% export growth experienced by the US in the period 2005-2015, 2.54% was due to a positive change in the bilateral stock of migrants (*transaction costs* channel), -0.25% to a negative change in the stock of immigrants coming from origins with comparative advantage (*knowledge diffusion* channel) and 0.30% to a positive change in the birthplace diversity of immigrants in the US (*diversity* channel). In figures C1-C4 we show maps on the overall role immigration and that of the three migration related channels on the intensive margin of export of all the countries covered in our exercise.

Alternative measures of birthplace diversity. In table C2 we report results using an alternative measure of birthplace diversity. Based on Montalvo & Reynal-Querol (2005) we approximate birthplace diversity by computing the ethnic polarization index. An *increase* in the polarization indicates a *reduction* in the diversity of migrant communities. Results in table C2 confirm our baseline evidence.

2SLS country-sector aggregated estimations. Table C3 shows the results obtained by adopting a 2SLS estimation approach on country-sector aggregated estimations (i.e. instrumenting both birthplace diversity and knowledge diffusion with the IV detailed in Section 2.2.2). The baseline evidence discussed in Section 4 is largely confirmed.

Alternative measures of problem solving intensity. In this section we present additional evidence for the aggregate cross-country analysis based on different proxies for the problem solving intensity of sectors, such as: (i) job complexity index as defined in Costinot (2009), (ii) O*NET based teamwork intensity (as discussed in Bombardini et al. (2012)), (iii) differentiated goods (Rauch 1999), (iv) sector's degree of skill intensity (UNCTAD, *Trade and Development Report* 2002), (v) sector's degree of technological intensity (UNCTAD, based on Lall (2000)). Consistently with the original product classifications, trade data are aggregated at the SIC 3-digit level. Each dimension is tested in a separate regression using a dummy variable taking the value of 1 if the sector is classified as differentiated, high skill (TDRE) or technology intensive (LDC09 and LDC10).[67] Results reported in Table C4 largely confirm our main findings, birthplace diversity positively affect country competitiveness in sectors characterized by higher degree of differentiation, high skill intensity or high technological intensity.

Birthplace diversity and knowledge diffusion by income level. In table C5 we show the effect of Birthplace Diversity and Knowledge diffusion for destination countries with different income levels. Interestingly, the birthplace diversity effect increases in the income level of the host country, while the knowledge diffusion channel particularly beneficial for less developed countries.

Birthplace diversity at sector level. One major threat to our identification strategy stems from the fact that it is impossible to measure workforce birthplace diversity at the sectoral level across countries. To cope with this limitation of the data we replicate our analysis across French districts. Using 1990 French population census we are able to construct a measure of workforce birthplace diversity at the district d and sectoral k level.[68] We combine this information with the national trade statistics and test the robustness of our results.[69] Table C6 reports the results from different cross-sectional specifications. Results in column 1 to column 5 corroborate our main findings across different years and trade margins. Point estimates in table C6 cannot directly compared with our baseline results because based on pure cross-section variation (while our baseline estimation base on within identification).

Table C1: Migration Stocks, Knowledge Diffusion and Diversity: back of the envelope calculation.

	Average % Change 2005-2015 Exports	$\overline{\Delta X_{2005}^{2015}} * \beta^X$, in %			
		Overall	Mig Stock	Knowledge Diff.	Birth. Div.
United States	52.19	2.60	2.54	-0.25	0.30
Germany	33.41	3.74	3.48	0.46	-0.21
Russia	50.27	-0.06	0.01	-0.07	0.01
Saudi Arabia	69.77	1.32	1.29	-0.01	0.04
UK	26.25	1.88	2.41	-0.52	0.02
UAE	121.77	5.54	6.88	0.14	-1.39
Canada	7.68	3.66	4.20	-0.51	0.00
France	17.25	2.88	2.48	-0.05	0.44
Australia	30.37	11.02	11.49	-1.15	0.73
Spain	39.39	6.58	6.16	0.14	0.25

Note: The intensity of the effect for each country is computed as the average change in the Immigration variable over the period 2005-2015 times the estimated coefficient from the baseline equation (in %).

[67] UNCTAD classifications are available at https://unctadstat.unctad.org/en/classifications.html.
[68] Data Source: Population Census 1990, sampling 1 = 4, INSEE, available at ADISP-CMH.
[69] Data source: French Customs, https://lekiosque.finances.gouv.fr/site_fr/telechargement/telechargement_SGBD.asp.

Table C2: 2SLS regressions. Using the Polarization index (baseline IV).

Dep var	$Dummy = 1$ if $Export_{ijkt} > 0$		$\text{Ln(export)}\vert_{export_{ijk(t-1)}>0}$	
	(1)	(2)	(3)	(4)
Mig_{ijt}	0.024***	0.017***	0.182***	0.145***
	(0.001)	(0.001)	(0.009)	(0.009)
Know. Diffusion_{ikt}	0.011**	0.036***	0.441***	0.843***
	(0.005)	(0.005)	(0.112)	(0.104)
Birth. Polariz._{it}	-0.056**		0.241	
	(0.026)		(0.281)	
Birth. $\text{Polariz.}_{it} \times \text{Abstract}_k$		-0.051***		-0.865***
		(0.009)		(0.140)
IV	Predicted supply		Predicted supply	
Controls: $\mathbf{X}_{ijkt}$	yes	yes	yes	yes
FE: θ_{ij}	yes	yes	yes	yes
FE: θ_{jkt}	yes	yes	yes	yes
FE: θ_{rckt}	yes	no	yes	no
FE: θ_{it}	no	yes	no	yes
Cluster std err	i j	i j	i j	i j
Observations	20,156,093	20,156,093	4,575,395	4,575,395
R-squared	0.138	0.018	0.170	0.100
F-stat First Stage	969.8	572.1	654.6	401.9

Note: In all regressions standard errors in parentheses are double clustered at exporter and importer country level. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 195$, Importers $j = 176$, Sectors (SIC72) $k = 142$, Year $t = 5$. IV for migration stock, knowledge diffusion and birthplace polarization index are based on the supply-driven predicted migration stocks.

Table C3: Country-sector aggregate results. Results by abstract intensity of tasks. 2SLS estimations.

Dep var	RCA_{ikt}		$Export_{ikt}$		$\#\ Dest_{ikt}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Birth. $\text{Divers.}_{it} \times Abstr._k$	0.341***	0.341***	0.877***	0.877***	0.269***	0.269***
	(0.048)	(0.101)	(0.078)	(0.170)	(0.042)	(0.090)
Know. Diff._{ikt}	1.103***	1.103***	2.251***	2.251***	0.465***	0.465***
	(0.073)	(0.159)	(0.119)	(0.272)	(0.040)	(0.085)
FE: θ_{it}	yes	yes	yes	yes	yes	yes
IV: Birth. $\text{Divers.}_{it} \times Abstr._k$	1.453***	1.453***	1.453***	1.453***	1.453***	1.453***
IV: Know. Diffusion_{ikt}	0.998***	0.998***	0.998***	0.998***	0.998***	0.998***
Cluster std err	ik kt	i k	ik kt	i k	ik kt	i k
Observations	114,262	114,262	114,262	114,262	114,262	114,262
F-test	2878	83.08	2878	83.08	2878	83.08

Note: In column 1, 3 and 5 standard errors in parentheses are double clustered at country-sector and sector-year. In column 2, 4 and 6 standard errors in parentheses are double clustered at country and sector. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 186$, Sectors (SIC72) $k = 142$, Year $t = 5$.

Table C4: Country-sector aggregate results. Synthetic export performance (RCA) estimations. OLS estimations.

Dep var	RCA_{ikt} (ln)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth. Div.$_{it}$	0.438***	0.440***						
	(0.064)	(0.065)						
Birth. Div.$_{it}$ × Complex$_k$		0.190***	0.200***					
		(0.040)	(0.040)					
Birth. Div.$_{it}$ × Team$_k$				0.162***				
				(0.039)				
Birth. Div.$_{it}$ × Knowledge$_k$					0.295***			
					(0.044)			
Birth. Div.$_{it}$ × Differ$_k$						0.203		
						(0.143)		
Birth. Div.$_{it}$ × Skill Int$_k$							0.613***	
							(0.094)	
Birth. Div.$_{it}$ × High Tech$_k$								0.836***
								(0.143)
Know. Diff.$_{ikt}$	0.474***	0.466***	0.461***	0.461***	0.460***	0.246**	1.028***	1.031***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.111)	(0.046)	(0.047)
Mig$_{it}$	0.038*	0.038*						
	(0.020)	(0.020)						
$RTA^{\#}_{i,t-5}$	0.006***	0.006***						
	(0.001)	(0.001)						
$ln(GDP)_{i,t-5}$	0.307***	0.307***						
	(0.023)	(0.023)						
FE: θ_i	yes	yes	no	no	no	no	no	no
FE: θ_{it}	no	no	yes	yes	yes	yes	yes	yes
FE: θ_{kt}	yes	yes	yes	yes	yes	yes	yes	yes
Cluster std err	ik kt	ik kt	ik kt	ik kt	ik kt	ik kt	ik kt	ik kt
Observations	114,262	114,262	114,262	114,262	111,440	193,465	196,858	196,858
R-squared	0.742	0.742	0.763	0.763	0.766	0.857	0.607	0.607

Note: In all regressions standard errors in parentheses are double clustered at country-sector and sector-year. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 186$, columns 1-5 Sectors (SIC72) $k = 142$; columns 6-8 Sectors (SITC) $k = 264$; Year $t = 5$.

Table C5: Country-sector aggregate results. Results by abstract intensity of tasks and Income. OLS estimations.

Dep var	RCA_{ikt}	$Export_{ikt}$	$\#\ Dest_{ikt}$
	(1)	(2)	(3)
Birth. Divers.$_{it}$ $\times$ $Abstract_k$	-0.030	0.028	-0.015
	(0.049)	(0.046)	(0.015)
Birth. Divers.$_{it}$ $\times$ $Abstract_k$ $\times$ $HighIncome_i$	0.322***	0.490***	-0.009
	(0.026)	(0.030)	(0.013)
Know. Diff.$_{ikt}$	0.591***	1.417***	0.303***
	(0.043)	(0.055)	(0.016)
Know. Diff.$_{ikt}$$\times$ $HighIncome_i$	-0.595***	-1.174***	-0.356***
	(0.087)	(0.093)	(0.034)
FE: θ_i	no	no	no
FE: θ_{it}	yes	yes	yes
FE: θ_{kt}	yes	yes	yes
Cluster std err	ik kt	ik kt	ik kt
Observations	114,262	114,262	114,262
R-squared	0.764	0.851	0.876

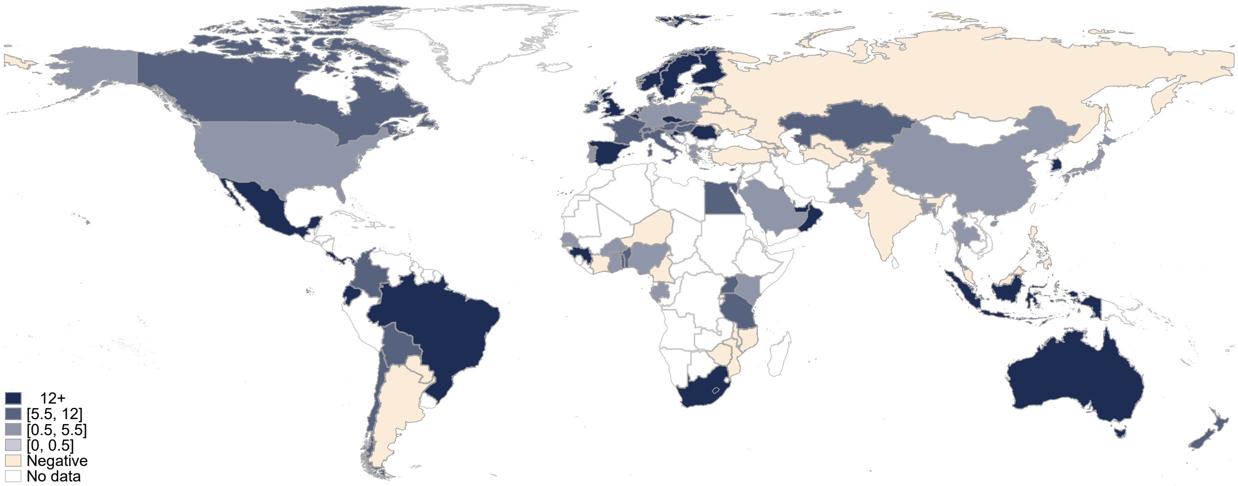
Note: In all regressions standard errors in parentheses are double clustered at country-sector and sector-year. $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Exporters $i = 186$, Sectors (SIC72) $k = 142$, Year $t = 5$.

Table C6: District-sector aggregate results. The case of France. OLS estimations.

Dep var	$Export_{dk}$				RCA_{dk}
	(1)	(2)	(3)	(4)	(5)
Mig$_{dk}$	0.635***	0.626***	0.613***	0.595***	0.584***
	(0.050)	(0.046)	(0.049)	(0.048)	(0.046)
Know. Diffusion$_{dk}$	-0.215	-0.255	-0.196	-0.241	-0.210
	(0.276)	(0.255)	(0.249)	(0.236)	(0.221)
Birth. Diversity$_{dk}$	1.236***	1.241**	1.406***	1.380**	1.296**
	(0.425)	(0.523)	(0.491)	(0.516)	(0.492)
Trade Year	2013	2015	2016	2017	2017
FE: θ_d	yes	yes	yes	yes	yes
FE: θ_k	yes	yes	yes	yes	Yes
Cluster std err	region	region	region	region	region
Observations	1,151	1,151	1,151	1,151	1,151
R-squared	0.752	0.747	0.749	0.749	0.691

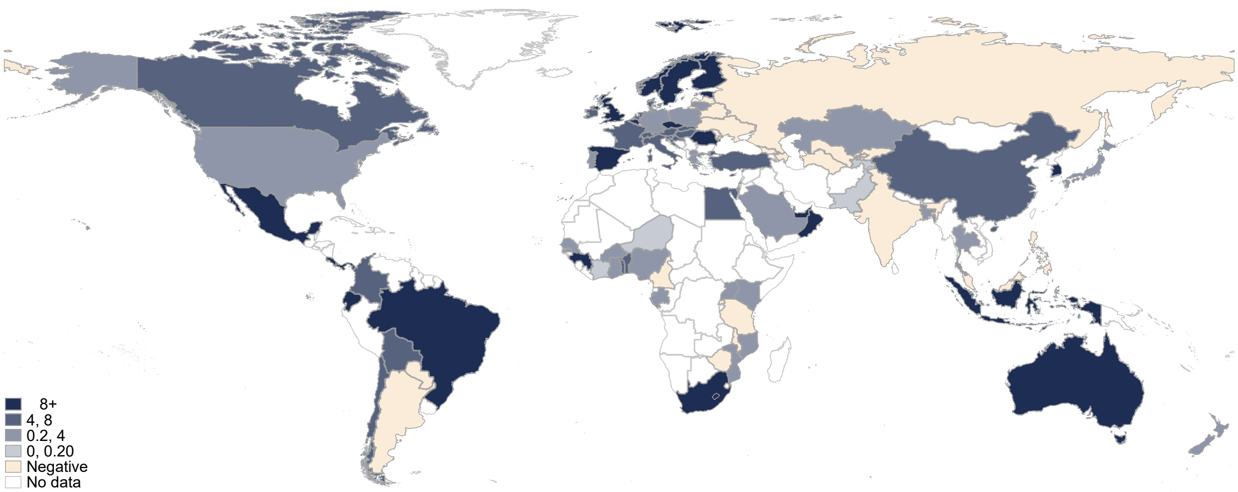
Note: In all regressions standard errors in parentheses are clustered at the Region level (Region $r = 22$). $***$, $**$, $*$ denotes statistically significance at the 1%, 5% and 10% level, respectively. Districts $i = 95$, Sectors $k = 15$.

Figure C1: Back-of-Envelope quantification: Migration Overall Effect, period 2005-2015.



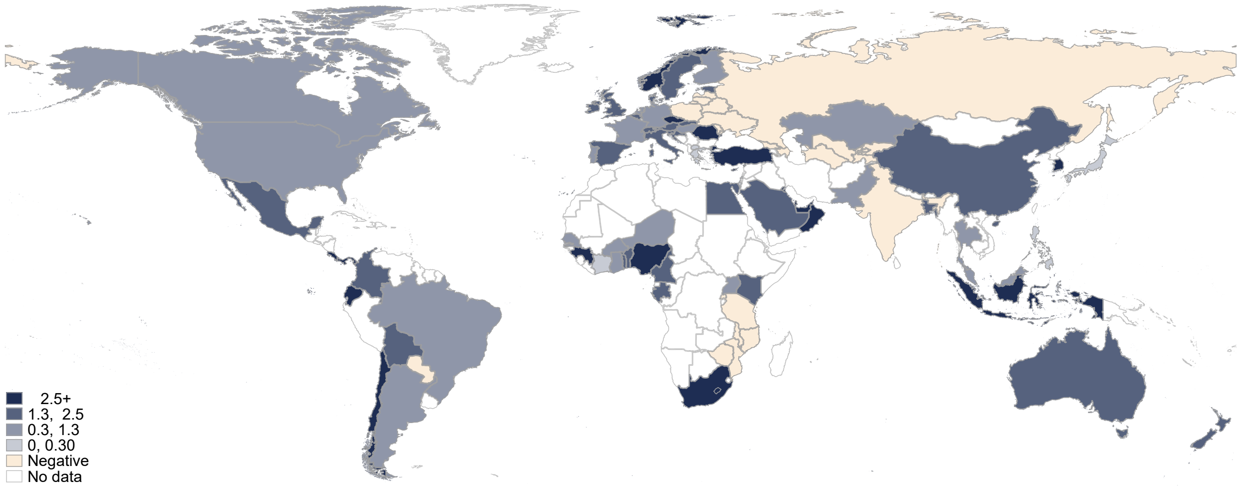
Note: the map reflects the heterogeneous impact of Immigrant. The intensity of the effect for each country is computed by fixing Mig_{ijt}, KD_{ikt} and BD_{it} at their values in year 2005 and projecting trade for year 2015 using the estimated coefficients from the baseline regression. The values refers to the percentage change in 2015 Exports induced by the observed immigration over the period 2005-2015. Negative values, implying a reduction in the stock of foreign born residents, are depicted in a brighter shade.

Figure C2: Back-of-Envelope quantification: Transaction Cost Channel, period 2005-2015.



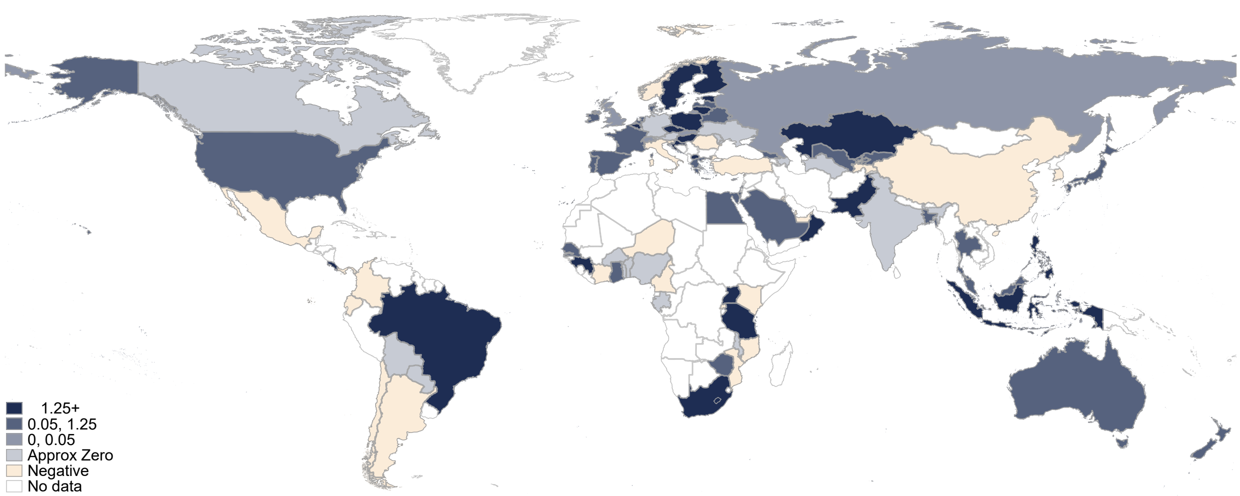
Note: the map reflects the heterogeneous impact of Immigrant Transaction Cost channel. The intensity of the effect for each country is computed by fixing Mig_{ijt} at year 2005 and projecting trade for year 2015 using the estimated coefficients from the baseline regression (i.e. 0.174). The values refers to the percentage change in 2015 Exports induced by the observed immigration transaction cost channel over the period 2005-2015. Values are mean-normalized entries smaller (greater) than 1 identify below (above) the average impacts. Negative values, implying a reduction in the stock of foreign born residents, are depicted in a brighter shade.

Figure C3: Back-of-Envelope quantification: Knowledge Diffusion Channel, period 2005-2015.



Note: the map reflects the heterogeneous impact of Immigrant Diversity channel. The intensity of the effect for each country is computed by fixing KD_{ikt} at year 2005 and projecting trade for year 2015 using the estimated coefficients from the baseline regression (i.e. 0.444). The values refers to the percentage change in 2015 Exports induced by the observed immigration knowledge diffusion channel over the period 2005-2015. Negative values, implying a reduction in the share of foreign born residents coming from countries with a comparative advantage in exports of sector k over total foreign born population, are depicted in a brighter shade. Notice that being a share a reduction in Knowledge Diffusion does not imply a reduction in the levels.

Figure C4: Back-of-Envelope quantification: Diversity Channel, period 2005-2015.



Note: the map reflects the heterogeneous impact of Immigrant Diversity channel. The intensity of the effect for each country is computed by fixing Mig_{ijt}, KD_{ikt} and BD_{it} at their values in year 2005 and projecting trade for year 2015 using the estimated coefficients from the baseline regression. (i.e. 0.614). The values refers to the percentage change in 2015 Exports induced by the observed immigration diversity channel over the period 2005-2015. Negative values, implying a reduction in the birthplace diversity of foreign born residents, are depicted in a brighter shade.

D Database on synthetic RCA index.

A side product of the present paper is a new database on the synthetic revealed comparative advantage obtained by estimating the equation (16). The theoretical foundation for considering estimated country-sector-year fixed effects a good proxy for (Ricardian) revealed comparative advantage is provided by Costinot et al. (2012). This database is intended to update and extend the RCA comparative advantage index proposed by Leromain & Orefice (2014). In order to purge the Revealed Comparative Advantage measure (i.e. country-sector-year fixed effects in equation 16) from any migration-driven transaction cost channel and from any aggregate effect of RTAs, we include in equation (16) the bilateral stock of immigrants and a dummy indicating the presence of an active RTA between country i and j at time t. In the vein of Costinot et al. (2012) the point estimates of country-sector-year fixed effects in equation (16) can be fairly considered valid proxies for the comparative advantage of country i in sector k and time t.

For the scrutiny and use of scholars and practitioners, we make this RCA index freely available on a dedicated webpage `here`. The user may download data for the full sample of 195 countries over the period 1990-2015 at both SIC 3-digit, HS 2-digit and 4-digit product level. The user will find three databases, one for each HS product aggregation. Each database contains four variables: (i) the ISO code of the country (variable *i*); (ii) the year (variable *year*); (iii) the sector of interest (variable *hs-code*)); and (iv) the measure of synthetic revealed comparative advantage (variable *RCA*).