

Heterogeneous Immigrants, Exports and Foreign Direct Investment: The Role of Language Skills

Abstract

We investigate the interplay of language skills and immigrant stocks in determining bilateral FDI out-stocks of OECD reporting countries. Applying a Poisson panel estimator to 2004-2011 data, we find robust evidence for a positive effect of bilateral immigrants on bilateral FDI - provided that residents of the two countries have few language skills in common. We find a similar effect for immigrants from third countries that speak the language(s) of the FDI host country, making them potential substitutes for bilateral migrants. Our findings suggest that immigrants facilitate outgoing FDI through their language skills, rather than through other characteristics like cultural familiarity.

Keywords: migration, FDI, foreign languages, globalization

JEL classification: F21, F22, O14

Financial support from Volkswagen Foundation Project "Europe's Global Linkages and the Impact of the Financial Crisis: Policies for Sustainable Trade, Capital Flows, and Migration" is acknowledged. We would like to thank participants in the Lunchtime Seminar in International Economics in Kiel as well and the 2015 SMYE, AEL, and VfS conferences for helpful comments, and in particular Toman Barsbai, Holger Görg, Gordon Hanson, Nicolas Keller, Magnus Lodefalk, Thierry Mayer, Rainer Thiele, the editor, and two anonymous referees. The usual disclaimer applies.

1 Motivation

One important feature of many international labor migrants is their transnationalism - that is, their familiarity with the cultures of both, their countries of origin and destination. Therefore, the presence of immigrants may facilitate international economic transactions, especially foreign direct investment and trade, by reducing international communication costs. Indeed, over the past two decades, many empirical studies have found a positive effect of immigrant stocks on foreign direct investment in the immigrants' countries of origin (Javorcik et al., 2011; Aubry et al., 2012) as well as on bilateral trade (Gould, 1994; Head and Ries, 1998; Rauch and Trindade, 2002). However, migrants are no homogeneous group. In addition to differences in education levels that are often studied, they are highly heterogeneous in terms of the languages they speak.

Most existing studies are based on a gravity model of bilateral trade or FDI where the bilateral migrant stock is added as an explanatory variable to reflect its possible dampening impact on the cost of international economic transactions. One concern with this approach is that migration, FDI, and trade are driven in part by common, unobserved, country-pair-specific determinants. Many existing studies do not address the resulting omitted variable bias; therefore, they probably overstate the positive effect of the migrant stock on bilateral outward FDI.

In this paper, we extend the empirical literature on the impact of immigrants on FDI in two directions. First, we estimate a gravity equation with panel data and country-pair-specific fixed effects because this approach has been shown to avoid omitted-variable bias (Parsons, 2012). This is a particularly demanding test of our underlying hypothesis. We address the impact of short-term, year-to-year variations in the number of immigrants on outward FDI, rather than the long-term effect that may be identified through cross-section analysis. While our short-term effect is economically significant in its own right, we also view it as a lower bound on the total migration effect, following the full social and economic integration of immigrants in the reporting country.

Second, we investigate how the impact of immigrants on bilateral outward FDI depends on the ease with which residents of the two countries can communicate, given their

common language skills. “Transnational” immigrants are familiar with the cultures and languages of both countries. We hypothesize that their presence is associated with higher FDI particularly when only few residents of the two countries can otherwise communicate in a common language - be it the official language of either country or any other language. By contrast, if most residents of the two countries have a language in common, the presence of immigrants may not reduce communication costs much further. Empirically, we try to eliminate all variation except for changes within country pairs over time. We thus try to look well beyond the simple effect of migration or language and conduct a very demanding test of the correlation.

While our focus on the interaction between immigrants and common language skills at the country-pair level is novel (to the best of our knowledge), we draw on several strands of literature. First, irrespective of migrant stocks, many studies have found that a common language increases bilateral trade (Egger and Lassmann, 2012) or FDI (Selmier and Oh, 2013); recent work such as Egger and Lassmann (2014) presents robust evidence on the direction of causality from language to trade.¹

Melitz and Toubal (2014) develop a global language dataset that includes not only the official language(s) of each country, but also any other language spoken by at least four percent of the population (such as English as a foreign language). The authors estimate a conventional gravity model for bilateral trade flows and find separate positive effects for migrant stocks and language variables (common official and spoken languages). They interpret this finding as a reflection of the role of migrants as translators and facilitators in international trade. This is similar to our hypothesis that migrants are more important when there is less knowledge of common languages.

Gould (1994) and others have argued informally that the language skills of immigrants explain the positive empirical association between the number of immigrants and bilateral exports. More generally, language skills are one example of a wide range of skills and characteristics of immigrants that may help to ease informational asymmetries and solve agency problems that are central to the proximity-concentration trade-off (Felbermayr and

¹ Kugler et al. (2013) report a similar positive effect on bilateral financial flows for their common official language dummy.

Jung, 2009; Rauch, 2001; Portes and Rey, 2005). Therefore, immigrants may facilitate various international economic transactions that involve an element of investment, including bilateral trade (because of the fixed cost of market entry), portfolio, and direct investment. The trade and FDI facilitating roles of migrants may thus work through a language channel but they may also be based on other skills or characteristics that are beyond the scope of this paper (e.g. culture).

Kleinert and Toubal (2010) derive a gravity equation for FDI from a heterogeneous firm model of international trade and FDI similar to Helpman et al. (2004). The firm's decision to undertake FDI, rather than to export or to produce only for the domestic market, is driven by the fixed cost of exporting vs. the fixed cost of setting up production abroad. It is natural to think that the presence of immigrants may reduce both kinds of fixed costs.² On this basis, migrant stocks become part of a rigorously derived gravity equation for FDI.³

We explain our approach to estimating the gravity equation for FDI in Section 2 below. In Section 3, we discuss data sources, particularly the use of the language matrices from Melitz and Toubal (2014). We present our econometric results in Section 4 and discuss robustness checks in Section 5. In Section 6, we summarize our findings and point out the implications for the role of migrants from developing countries in facilitating FDI from high-income countries in their countries of origin.

2 Empirical approach

Several recent papers estimate a single-period cross-sectional gravity model of FDI with the stock of migrants included as an “exogenous” variable (most notably Leblang, 2010). However, this approach will yield biased estimates for at least two reasons: First, there is likely to be reverse causality between FDI and migrant stocks. Second, both FDI and

²In related work, Oldenski (2012) shows that communication-intensive industries are more likely to market their products abroad through FDI than exports. This finding points to the importance of communication with customers, adjusting products to local preferences, etc. in successfully undertaking FDI. It is plausible to think that “transnational” immigrants help with these crucial tasks.

³Aubry et al. (2012) follow a similar approach but use the ratio of FDI stocks (as a proxy for FDI sales) to exports as their dependent variable, as pioneered by Helpman et al. (2008). This approach has been criticized by Santos Silva and Tenreyro (2015) for requiring unrealistically strong normality and homoskedasticity assumptions.

migrant stocks are probably subject to unobserved heterogeneity. Unfortunately, there is no easy way to address these concerns: There are no known instruments for *bilateral* migration in an FDI or trade context with a valid exclusion restriction at the global level. Furthermore, the bias does not occur at the country level but at the country-pair level; for example, specific historical relations between two countries may be a reason for both high bilateral migration and high bilateral FDI flows. Therefore, cross-sectional country-level estimates are potentially severely biased.

Thus our analysis of the impact of immigrant stocks on FDI involves a difficult trade-off between bias reduction and the time horizon of our estimates. We estimate a panel model with bilateral fixed effects in order to eliminate the likely omitted variable bias from unobserved historical ties that may affect all economic interactions - bilateral migration, FDI, and trade. Our bilateral fixed effects are time-invariant - which is a sensible assumption only over a relatively short time horizon. In order to reduce the scope for reverse causality even further, we use one year lagged immigrant stocks to explain FDI stocks. Thus we estimate a correlation that can be seen as the short term "effect" of the immigrant stock on outward FDI as precisely as possible; we consider this short-term effect a lower bound on the total, long-term effect that may be realized only as immigrants integrate in the reporting country economically and socially. We can see no way of estimating the long-term effect remotely as cleanly as the short-term effect and therefore do not attempt it.

When we choose our estimator, we take into account that bilateral trade and FDI data contain many zeros and are highly heteroskedastic (cf. Silva and Tenreyro, 2006, and follow-up papers). We use a fixed effects Poisson model with cluster-robust standard errors (Wooldridge, 1999; Stock and Watson, 2008). This estimator requires only minimal assumptions, particularly that the conditional mean is correctly specified and that the independent variables are strictly exogenous.⁴

The explanatory variables in our gravity model for bilateral FDI out-stocks include log migrant stocks, time-variant gravity variables such as GDP, GDP per capita, and free

⁴These assumptions are discussed in the robustness section below. A Monte Carlo study is provided by Westerlund and Wilhelmsson (2011). Note that the estimator should not be confused with either the negative binomial MLE or the panel qMLE.

trade agreements, time-invariant bilateral fixed effects, and year fixed effects. Thus the direct effect of all time-invariant factors is subsumed under the bilateral fixed effect. Instead of log-linearizing the outcome variable, our estimator directly estimates the exponential regression function in its multiplicative form. Using the notation of Westerlund and Wilhelmsson (2011), we are interested in the simple gravity equation with bilateral fixed effects

$$E(FDI_{ijt}|Y_{it}, Y_{jt}, D_{ijt}, \theta_{ij}) = \exp(\theta_{ij} + \gamma D_{ijt}) Y_{it}^{\beta_1} Y_{jt}^{\beta_2}. \quad (1)$$

Y_{it} and Y_{jt} contain time-variant, contry-specific gravity variables for reporting and partner countries, respectively. D_{ijt} contains variables such as the one year lagged bilateral log migrant stock which vary over time as well as across country pairs. θ_{ij} denotes the time-invariant bilateral fixed effects.

We then directly estimate the exponential regression function using maximum likelihood with the fixed-effects-conditional mean given by

$$\gamma_{ijt} = \exp(\theta_{ij} + \gamma D_{ijt} + \beta_1 \ln(Y_{it}) + \beta_2 \ln(Y_{jt})). \quad (2)$$

With this approach, we neither need to truncate the dependent variable (which includes many zeros) nor subject it to an arbitrary non-linear transformation that could lead to severely biased and inconsistent results (cf. Silva and Tenreyro, 2006; Westerlund and Wilhelmsson, 2011).

3 Data

Overview

Of the data required for our analysis, immigrant stocks represent the bottleneck. The widely used World Bank Bilateral Migration Database relies on census data and is decadal. This is insufficient for our analysis where we seek to eliminate unobserved heterogeneity at the country-pair level as much as possible and therefore rely on much of this unobserved

heterogeneity being time-invariant. This assumption can be argued to hold with yearly data in our context but is unlikely to hold across decades. Therefore we use yearly data from the OECD international migration database that provides good coverage of immigrants in OECD countries from all countries of the world (Dumont and Lemaitre, 2004; Widmaier and Dumont, 2011; Arslan. et al., 2014).

Most reporting countries define immigrants either by country of birth or nationality; only a few report both definitions. Therefore, if we were to limit our analysis to data based on only one definition, we would exclude a substantial number of countries. In the interest of external validity, we include each country's immigrant stock according to the country's preferred definition. If both are available, such as in the case of Norway, we use the country of birth. Since we only rely on within-bilateral-tie variation and eliminate differences between countries, it seems unlikely that these differing definitions will affect our findings. However, we ensure that definitions and reporting of migration numbers do not change within bilateral tie since this would lead to purely artificial time-variation in migration numbers. Between 2002 and 2004 several OECD countries adjusted their definition of migrants in these statistics, for example by cleaning up register data, regularizing and thus suddenly counting irregular migrants or changing reporting standards. Therefore, we restrict our data to the years from 2004 that are unaffected by these changes. Otherwise our analysis would have been based partly on purely statistical or artificial fluctuations in the migration data, resulting in low comparability even within countries. Furthermore, we use variables from the Doing Business Report⁵ to account for the fact that the investment climate in some countries may have been so poor as to preclude most FDI except to exploit natural resources.

Our final dataset includes all OECD countries except Luxembourg and Turkey.⁶ Data range from 2004 to 2011. In our Poisson panel estimations, where values for consecutive years are required, results are based on 26 host countries with an average of 82 partners

⁵Doing Business, The World Bank (<http://www.doingbusiness.org>)

⁶Luxembourg is missing because the language scarcity variable (see below) is only available for "Belgium and Luxembourg" combined. Inspecting the data and comparing independent estimates of the share of Dutch, French and German speakers in Belgium and Luxembourg, we decided to treat the "Belgium and Luxembourg" values as representative of Belgium, but not Luxembourg. In the case of Turkey, migration data exist only for a single year, making panel estimation impossible.

each. Average bilateral data points per reporting country range from 3 to 5.6 years with an average of 3.9 years.⁷

We match our immigrant stock data with FDI outstocks and outflows from the OECD International Direct Investment Statistics (OECD, 2014) and to UN Comtrade data on exports and imports⁸. Our definition of FDI thus comprises vertical and horizontal FDI. Language data are from Melitz and Toubal (2014); we discuss the construction of our "common language scarcity" variable in the following sub-section. We add standard gravity variables and take GDP and population data from the World Bank's World Development Indicators⁹.

The final dataset consists of 11,999 bilateral observations in 3,089 bilateral ties (Table 1). Naturally, FDI, exports, and immigrant stocks are highly skewed with the mass of observations close to zero.

[Table 1 about here.]

Language ties and the bridging role of migrants

We hypothesize that the scarcer are the language skills needed to do business with another country, the more valuable is the presence of immigrants from that country. As a first test, we allow the coefficient of the bilateral immigrant stock in our gravity model to vary with the share of individuals in the two countries who have no spoken language in common. Thus the bilateral immigrant stock appears among our explanatory variables not only on its own, but also interacted with our "common language scarcity" variable.¹⁰

We calculate "common language scarcity" from the "common spoken language" variable in Melitz and Toubal (2014). For every country pair, Melitz and Toubal estimate the chance that two randomly selected individuals, one from each country, have a language in common. They compile a global database of all 42 languages that are spoken by at least 4% of the population in at least two countries. For every country i they report the share L_{li} of the population who speak language l . Then, for every country pair (i, j) and language

⁷The reductions in years per bilateral tie is due to incomplete migration data for some countries.

⁸<https://comtrade.un.org/db/default.aspx>

⁹<http://data.worldbank.org/data-catalog/world-development-indicators>

¹⁰ Here we follow Ginsburgh and Prieto-Rodriguez (2011) in our reasoning.

$l = 1, \dots, L$, they calculate the chance that two randomly selected residents both speak language l ($L_{li}L_{lj}$). They sum this term over all languages l for each country pair to yield an estimate of the chance that two randomly selected residents have a language in common: $\alpha_{ij} = \sum_l L_{li}L_{lj}$.¹¹ Finally they standardize their common spoken language variable to lie between 0 and 1.

For our analysis, we are interested in the share of residents of the two countries who *cannot* communicate in any language. Therefore, we use $(1 - \alpha_{ij})$ as a proxy for the scarcity of language ties. With standardization, our common language scarcity variable takes a value of 1 for two countries whose residents have no language skills in common whatsoever. By contrast, when all residents can communicate with one another (say, because two countries share the same official language and this is spoken by everyone), the language scarcity variable takes a value of 0. When we estimate our gravity equation for bilateral FDI out-stocks, we expect that common language scarcity interacted with the (lagged) immigrant stock will have a positive effect on outgoing FDI. Within the short time span of the panel, there is hardly any change in common languages through other channels than migration. The time invariant effect of a shared language between two countries is eliminated by the bilateral fixed effect.

In our data, "common language scarcity" is distributed between 0 and 1 and provides sufficient common support (Figure A1 in our Online Appendix) to make inferences on a cross-country basis. The average is 0.78 and the median 0.91. As one might expect, English language skills are the most important contributor to bilateral communication.

We also take a step further and investigate whether immigrants from third countries with similar language skills act as substitutes for immigrants from the FDI host country. Take the example of United Kingdom FDI in Mozambique. Residents from the two countries have almost no language skills in common (common language scarcity is close to 1) and there are few immigrants from Mozambique in the UK. However, maybe immigrants in the UK from other Portuguese-speaking countries such as Brazil or Portugal facilitate UK FDI in Mozambique in a similar fashion - provided that language skills, rather than

¹¹ This is an overestimate of the true language overlap to the extent that some random pairs of residents may be able to communicate in more than one language.

other factors like familiarity with local institutions and culture, are at the heart of the link between immigrants and outward FDI.

To capture the possible effect of third-country immigrants in reporting country i who speak one of the languages of FDI host country j ("same language migrants"), we estimate their number based on the languages spoken in the immigrants' countries of origin:

$$SameLanguageMigrants_{ijt} = \sum_k (\alpha_{jk} \cdot migrants_{ikt}), \text{ where } k \neq i, j \quad (3)$$

We include "same language migrants" in the explanatory variables of our gravity model in parallel with bilateral immigrants.

4 Results

As expected, our decision to estimate the gravity model from panel data with bilateral fixed effects constitutes a demanding test of our hypotheses (Table 2). To facilitate comparison with other literature, we begin by estimating our basic gravity equation for both exports and FDI out-stocks with bilateral migrant stocks, but without interaction terms, using three different specifications: the widely used PPML estimator with reporter, partner and year fixed effects (Columns 1 and 2); our preferred panel Poisson setup with bilateral and year fixed effects and gravity variables (detailed in table notes) that will be used in the rest of the paper on different samples (Columns 3 to 8); a robustness check combining bilateral, reporter-year, and partner-year fixed effects (Columns 9 and 10). In addition, we provide evidence of our results' robustness using alternative language data (Columns 11 and 12).

[Table 2 about here.]

Under PPML without bilateral fixed effects, the bilateral immigrant stock has a statistically significant and positive coefficient for both exports and FDI as the literature suggests (Columns 1 and 2). With our preferred panel Poisson estimator, which relies only on variation over time and also eliminates bilateral tie specific time-invariant differences, we find a positive significant coefficient for bilateral immigrant stocks only for exports (Column 3),

but not FDI (Column 4). These results suggest that the positive association between immigrant stocks and FDI that is typically found in conventionally estimated gravity models (similar to Columns 1 and 2) may be severely upward-biased.¹²

We analyze our main hypothesis by allowing the impact of the bilateral immigrant stock on FDI to vary with our "common language scarcity" variable in columns 5 through 8. The coefficient of the corresponding interaction term is significantly positive for FDI (Columns 6, 8, and 10), but not for exports (Columns 5, 7, and 9). While Panel A includes the full available migration data from 2000-2011 including the changes in definition mentioned in the data section, in Panel B we restrict our working sample to the more reliable migration data from 2004 onwards. The stylized results are very similar and results become statistically more significant when excluding low quality data. The results in Panel B are therefore our preferred estimates. The stylized finding holds both for the specification with bilateral and year effects and the the very rigid specification with bilateral, reporter-year and partner-year fixed effects in columns 9 and 10. As hypothesized, bilateral immigrants facilitate FDI the more, the scarcer are the common language skills between the two countries.

By contrast, although we have earlier found that exports are facilitated by bilateral immigrants, the size of this effect does not significantly depend on common language scarcity. Therefore, in columns 5, 7, and 9 this effect is split into two components that are both positive, yet statistically insignificant, or close to zero and add approximately to the single coefficient in column 3 when applied to countries with average or median language scarcity. This pattern indicates that the correlation between migration and exports does not vary across the range of the interacted variable. This finding is support by the specification with an alternative language scarcity proxy from Dow and Karunaratna (column 11).¹³

¹²The sign of the correlation between migration and outward FDI in the restrictive panel Poisson fixed effects estimation heavily depends on the inclusion of cases with very high common language values. While in the full sample there is no statistically significant correlation between migration and FDI (Column 4), when excluding the less than 10 percent of the sample where more than 80 percent speak a common language, the coefficient on FDI is very similar to the PPML setting and statistically significant. For exports, conditioning on language scarcity does not affect the coefficient in a similar way. These findings suggest that the contribution of migrants to bilateral exports may work less through their language skills than in the case of FDI.

¹³This proxy is only available for a subsample of countries. The stylized result also holds with this subsample in the other specifications of Panel B.

The theoretical argument for the role of migrants in facilitating exports and FDI relate to variable and fixed trade/FDI costs. However, the scope for reducing these costs by employing migrants is not unlimited because these costs are largely driven by a country's business climate. In Table 2, Panel C we therefore test the role of the cost of doing business by splitting the sample along the median value of export-/FDI-receiving countries for two indicators frequently used to study the conditions for international trade and FDI: time to import and property registration time. The results suggest that, while there is no clear pattern for exports, the statistically significant link between FDI, migration and language scarcity only exists if business conditions are not too difficult. By contrast, in countries with a poor business climate, the related administrative costs apparently limit FDI irrespective of how migrants might otherwise contribute to lowering communication and other international transaction costs.

The combination of a straightforward positive effect of bilateral immigrants on exports and a more nuanced effect on FDI can be motivated in the framework of a heterogeneous firm model (e.g. Helpman et al., 2004). Exporting means adapting domestically produced goods to foreign preferences and marketing these goods in an unfamiliar environment. It seems plausible that greater access to immigrants as potential employees with extensive cultural knowledge of both countries may reduce the cost of exporting even when communication is not impeded by the scarcity of common languages.

By contrast, both vertical and horizontal FDI implies reporting country staff managing substantial part of the activities that are carried out by the local staff in the FDI host country. This is very communication intensive and may be infeasible if few in the FDI host country speak the sending country's language or a substitute such as English. Bilateral immigrants may thus provide highly specialized communication tasks in the FDI sending country without which there might be less scope for FDI to their country of origin. However, since most other tasks in production and marketing are carried in the FDI host country - after all, this is why firms decide to use FDI rather than exports - there may be little other use for immigrants' origin-country specific skills left.

Should this pattern hold in the data, we would expect bilateral migrants not only to matter more for FDI if there is little language overlap in general but also to be more positively associated with FDI in receiving country where there is little knowledge of English, the world's business language. In Table 3 we analyze whether bilateral migrants are indeed more strongly associated with FDI the lower the share of English speakers. Indeed, point estimates of the corresponding coefficient are substantially higher if we restrict the estimation to, for example, partner countries with fewer than 10% English speakers.¹⁴ For exports, there is no such pattern, suggesting that common language skills or English as a substitute are not the main driver of the correlation between bilateral immigrants and exports.

[Table 3 about here.]

Our estimates imply a negative effect of immigrants on FDI unless common languages are scarce (negative coefficient for the direct effect of the lagged immigrant stock in Table 2, Column 8). This result is probably an artifact due to our linear specification of the variable coefficient model, combined with the fact that about 75 percent of observations on "common language scarcity" lie between 0.75 and 1 (Figure 1). If the mass of observations lies in an area where the true overall effect increases and there is little statistical weight close to the origin, a linearity assumption on the effect can imply that the estimate undershoots the true effect of 0 to indicate negative values. Within the relevant range of values of common language scarcity, the composite effect of immigrants on FDI is predicted to be positive: given the baseline coefficient estimates (Table 2, Column 8), the composite effect of the immigrant stock on FDI exceeds 0 if common language scarcity exceeds 0.65.

[Figure 1 about here.]

In order to formalize this argument, we test alternative specifications of our gravity model that only enter the bilateral immigrant stock weighted by common language scarcity if the latter exceeds a specific threshold. This lifts the linearity assumption on the relationship between language scarcity and the log migrant stock. Thus, the hypothesis that

¹⁴Since sample sizes are substantially lower, standard errors are larger.

bilateral migrants only become relevant for outgoing FDI as common language skills become scarce can be tested using a single migration-related variable. The hypothesized relationship between immigrants and FDI is nil up to the respective threshold and linear thereafter.¹⁵ This way, the function has no jumps and the estimated coefficients is the predicted effect for a bilateral tie without any common language. Figure 2 shows the estimated coefficient and 95% confidence interval for each threshold $p \in [0, 1)$, which is shown on the vertical axis. For thresholds above 0.65, there is strong evidence of a statistically significant, linear relationship. This includes the mass of our observations on common language scarcity and is consistent with our main estimate. We are confident, therefore, that the negative direct effect of the immigrant stock on FDI is merely a technical consequence of our linear variable coefficient model and does not indicate that immigrants actually reduce FDI for low values of common language scarcity (see also Footnote 2).

[Figure 2 about here.]

Our estimates suggest that the impact of the bilateral immigrant stock on outward FDI is sizable: for a country pair with median common language scarcity at 0.91 (cf. Table 1), the composite effect is 0.21 for the full sample: approximately, a 1 percent increase in the bilateral immigrant stock implies a 0.21 percent increase in the FDI outstock. When residents in the two countries share almost no language skills (common language scarcity = 0.99), the corresponding increase in FDI is 0.27 percent.

Now we turn to our second language-related hypothesis regarding the impact of immigrants on outward FDI: FDI may be affected by immigrants from third countries who speak the same language(s) as FDI host country residents (our "same language migrants" - Section 3). It would be natural to add "same language migrants" to our gravity model exactly in parallel with bilateral migrants, i.e. both on its own and interacted with common language scarcity. However, there is too little meaningful variation left when all four

¹⁵For interpretation, we standardize the new covariate depending on threshold $p \in [0, 1]$ as

$$f(p) = \begin{cases} 0 & \text{if } 1 - \alpha_{ij} < p \\ ((1 - \alpha_{ij}) - p)/(1 - p) \cdot \text{lagged } \log(\text{migrants}_{ij}) & \text{if } 1 - \alpha_{ij} \geq p. \end{cases}$$

migrant-related variables are included. With an average of only 5 data points for each country pair and many zeros, our panel data do not permit reliable within-estimates for so many regressors that vary over time and across country pairs.

Therefore, we include "same language migrants" in our gravity equation under three alternative specifications: (i) for the whole sample (Table 4, Columns 1 and 2); (ii) only for country pairs with common language scarcity above 0.9 (Columns 3 and 4); and (iii) only for country pairs with common language scarcity above 0.95 (Columns 5 and 6). For FDI, the corresponding coefficient increases along with the threshold on common language scarcity (Columns 2, 4, and 6) and becomes significant at the 0.95 threshold; the latter specification still covers 46 percent of all observations. The effect from "same language migrants" is large, with a coefficient of 0.78 for the 0.95 threshold (Column 6); it comes on top of the broadly unchanged effect for bilateral immigrants - a composite effect of 0.18 when common language scarcity is set to 0.95 which is close its median value. Thus for country pairs with few common language skills, third-country immigrants with the "right" languages may act as substitutes for bilateral immigrants.¹⁶ This substitutive relationship also remains statistically significant when we use a kinked functional form similar to the one discussed earlier to model the scarcity threshold (cf. Figure 2).

[Table 4 about here.]

It is remarkable that we find a robust effect on FDI exactly for those third-country immigrants that have the "right" language skills to communicate with FDI host country residents. It is nevertheless conceivable that factors other than language skills may enable migrants that do not come from the FDI host country to nevertheless facilitate FDI. One possible such factor is cultural familiarity, which is often proxied by geographical proximity. We test this hypothesis by including immigrants from the FDI host country's neighbors in our gravity equation. It turns out that there is no significant effect from "neighborhood" immigrants on bilateral FDI in general (Table 5, columns 1 and 2). However, if neighboring countries have a large language overlap with the FDI host country (e.g. Chile and Peru) and

¹⁶The negative effect on exports, which is significant in Column 3, may simply reflect more third-country migrants making exports to their home country more profitable.

if common language skills between FDI origin and host country are also scarce, there is a small but significant positive correlation between immigrants from neighboring countries and bilateral outward FDI (Table 5, Columns 4 and 6). Once again, the coefficients for bilateral immigrants are almost unchanged from our basic results (Table 2).

[Table 5 about here.]

So far, we have assumed a double-log-linear relationship between immigrants and outward FDI. We now explore whether FDI is subject to decreasing returns to immigrants (Table 6, Columns 1 to 4). First, we exclude the 25 percent largest bilateral immigrant stocks from our sample because we expect immigrants to have a larger effect in less developed migration corridors (Columns 1 and 2). Second, we exclude the favorite corridor of each country of origin from the sample (Columns 3 and 4). Under both specifications, the composite effect of the bilateral immigrant stock on outward FDI at median "common language scarcity" is substantially smaller (at 0.08 and 0.10, respectively) than in our basic regression (0.21; Table 2). In the presence of decreasing returns, we would expect the composite effect to be larger, rather than smaller, than for the full sample. Hence our estimates provide no evidence of decreasing returns in the relationship between bilateral immigrants and outward FDI.

Finally, we use the limited data on bilateral migrant stocks broken down by skill level from DIOC 2000 to distinguish between high-skilled and other migrants (Table 6, Column 5 and 6). In line with much of the immigration literature, we defined high-skilled migrants as those with ISCED levels 5 or 6, i.e. those with at least some tertiary education.¹⁷ This is a considerable simplification, because the effect of migration on FDI is more likely to be tied to occupations or tasks than to education levels. With this caveat in mind, our findings suggest that low-to-medium-skilled immigrants have no significant effect on outward FDI, whereas high-skilled immigrants have a higher composite impact (0.31) than total bilateral

¹⁷We assume that the skill ratio observed in 2000 is representative of our whole period of observation (2004 to 2011), knowing that in the long term maturing migration networks can lead to slowly falling skill ratios due to falling migration costs that benefits low skilled migrants in particular. Our regression results are therefore highly tentative (Table 6, Columns 5 and 6) but additional tests using DIOC 2010/11 data indicate their robustness.

immigrant stocks according to our baseline results. However, without proper annual data on immigrant stocks by skill levels, it is impossible to corroborate this hypothesis further.

[Table 6 about here.]

5 Further robustness checks

In most of our reporting countries, annual immigrant stock data come from administrative sources such as population registers or residence permits. A few countries, including the United States and the United Kingdom, rely on census data or even surveys such as the Current Population Survey or the Labor Force Survey. Survey-based data may be unreliable for small groups of immigrants (defined by country of origin); hence regression estimates based on within-variation in such data may be problematic. On the other hand, census data as such tend to be fairly precise, but the gaps in between censuses are sometimes filled by national statistical offices using procedures that are not fully reported. Therefore, to assess the robustness of our findings to different sources of migration data, we restrict our sample to the best available administrative data (Table 6, Columns 7 and 8). The composite effect of the bilateral immigrant stock on outward FDI is now larger (at 0.39) than in our baseline regression as could be expected if measurement error was present. The direct effect of the immigrant stock is still negative, but no longer significant. We consider these estimates as support for the robustness of our baseline results. We also stick with our baseline results because we do not wish to exclude several important migrant destination countries from our sample, as we would have to do in order to use administrative data only.

One important requirement for a causal interpretation of our results is the exogeneity of immigrant stocks, conditional on the covariates and fixed effects. This is potentially problematic because changes in FDI outstocks could be correlated with changes in immigrant stocks due to some underlying factor or due to reverse causality. This concern can be assessed through placebo regressions that rearrange the chronological order of events in our presumed causal chain. If correlations remain similar in a framework without lags or a specification with leads, this indicates simultaneity bias, reverse causality or severe autocorrelation. Reassuringly, when we replace the lags of the immigrant stock and inter-

action variables with leads, our gravity regression does not produce the same significant effects of immigrants on FDI. Likewise, when we leave out the lag structure altogether, we obtain only statistically insignificant estimates. These findings support our hypothesis that the direction of causality runs from immigrants to FDI because we find that the positive influence of immigrants on FDI takes some time to take effect. Our findings also imply that immigration is not caused by outward FDI. Furthermore, we confirm that immigration is not significantly correlated with FDI in-stocks; thus immigrants do not come to OECD countries in tandem with, following, or leading inward investment from their origin countries. In spite of this tentative evidence, our identification strategy only permits conclusions about correlations, not causal relationships.

Our analysis might also be subject to a systematic bias if past bilateral immigrants cause current immigration due to network or similar effects (Munshi, 2003). If this led to autocorrelation in the inflow of migrants that is correlated with time-invariant factors, this would have the potential to spill over into the within variation on which we rely for our estimates with bilateral fixed effects. We therefore show through additional tests that the growth of the log immigrant stock is not higher when there is a larger immigrant presence - as a network effect would imply (Figure A2 in our Online Appendix). Therefore, Card-style shift-share instrumental variables lack instrumental relevance in this study - in contrast to destination specific studies where they can work well (cf. Peri and Requena-Silvente, 2010). Also we would find it inconvincing to use the exogenous part of variation in push factors because many push factors simultaneously decrease the attractiveness for incoming FDI (e.g. instability) and factors that shift the importance of given levels of push factors (e.g. migration policy in Mayda, 2009) are often determined specifically with the aim of improving economic ties between countries. There is also no evidence of significantly more bilateral immigration based on common spoken languages in the short run (Figure A3). Furthermore, the results are robust to the exclusion of longstanding bilateral migration ties, based on per capita bilateral immigrant stocks in 1960 that are taken from the 'Worldbank Bilateral Migration Database 1960-2000' (Özden et al., 2011).

We have tested extensively whether the immigrant stock - language scarcity interaction

might reflect any other correlated factor. In Table A1 in Online Appendix, we test whether similar patterns to language scarcity can be found for several gravity and language-related variables, including distance between the two countries, past colonial ties, common official languages, and language proximity (a proxy of how much overlap there is between two languages in linguistic terms, yielding for instance a high value for Dutch and English). If we had found a pattern similar to that of migration and language scarcity, it might have indicated a high level of collinearity between language scarcity and the respective other variable, which would substantially weaken our main results regarding migrants and scarce languages. Reassuringly, we do not find any similar pattern to that for common language scarcity among the four variables that we have tested. Furthermore, we show in Table A2 that including these additional interaction terms does not substantially affect our baseline estimates.

Our results continue to hold when we exclude country pairs with common borders from our sample. In this case, the estimated coefficient on the immigrant stock - language scarcity interaction term is (insignificantly) smaller than in the full sample. It appears (plausibly) that contiguous countries whose residents cannot communicate easily benefit substantially from the presence of immigrants with respect to their FDI.

Furthermore, our estimates are robust to excluding all zero bilateral migrant stocks or the smallest quarter of bilateral migrant stocks. We can also control for inflows of asylum seekers without affecting our results. Finally, our results are robust to using language shares that have been cleaned of bilateral immigrants.¹⁸

6 Conclusion

We conduct a demanding test of the hypothesis that immigrants facilitate FDI from their country of destination to their country of origin. We use panel data for OECD reporting countries with bilateral and year fixed effects to estimate a gravity model for bilateral FDI out-stocks in OECD and other host countries. We find a robust positive impact of bilateral

¹⁸For example, the likelihood that a French and a UK resident can communicate in a common language will increase if UK residents include many French immigrants, and French residents include many immigrants from the UK.

immigrants on FDI only if residents of the two countries have few language skills in common. While earlier studies have concluded that immigrants may facilitate FDI, our finding suggests that immigrants' language skills play a crucial role in mediating this effect.

Quantitatively, we find that a 1 percent increase in bilateral immigrants in a country pair with median "common language scarcity" implies a 0.21 percent higher FDI out-stock according to our baseline estimates. When residents in the two countries share almost no language skills (common language scarcity = 0.99), the FDI outstock increases by 0.27 percent. However, if FDI host countries' business climate is too adverse, we do not find evidence of a positive relationship. In this case, other constraints to FDI thus seem to outweigh the possible positive effect of migrants.

Our conclusion receives strong support from an additional set of regressions that focus on those third-country immigrants in the OECD reporting country that have the same language skills as FDI host country residents. Through several specifications, we find that this group has a positive effect on bilateral FDI similar to bilateral immigrants. To further corroborate the key role of language skills, we play devil's advocate and search for a similar effect through other third-country immigrants. We focus on those from the FDI host country's neighbors because geographic proximity is a good proxy for cultural familiarity, which may help to facilitate FDI. However, we fail to find a positive effect on bilateral FDI outstocks through this group. However, these robustness checks cannot definitely rule out that part of the association between FDI, migration and language is caused by cultural similarity or knowledge of the country of origin. Future research at the micro level may be able to disentangle these effects more conclusively.

Our finding has a particular bearing for FDI from high-income to medium and low-income countries. In pairs of rich countries, residents can often communicate either in English or in the countries' official languages, obviating the need for immigrants to facilitate FDI. By contrast, for most country pairs with one rich and one developing country, common language skills are scarce. For these country pairs, immigrants in the rich country with the "right" language skills have an important role to play in facilitating FDI. Promoting foreign (as well as domestic) investment is often thought of as an important development

strategy. Our findings suggest that developing countries, as well as aid donors, should seek to link their investment promotion and Diaspora policies.

Bibliography

- Arslan., C., Dumont, J.-C., Kone, Z., Moullan, Y., Özden, C., Parsons, C., Xenogiani, T., 2014. A New Profile of Migrants in the Aftermath of the Recent Economic Crisis. OECD Social, Employment and Migration Working Papers, n. 160, OECD Publishing.
- Aubry, A., Kugler, M., Rapoport, H., 2012. Migration, FDI and the Margins of Trade. mimeo.
- Dow, D., Karunaratna, A., 2006. Developing a multidimensional instrument to measure psychic distance stimuli. *Journal of International Business Studies* 37 (5), 578–602.
- Dumont, J.-C., Lemaitre, G., 2004. Counting immigrants and expatriates in OECD countries: a new perspective. mimeo.
- Egger, P. H., Lassmann, A., 2012. The language effect in international trade: A meta-analysis. *Economics Letters* 116 (2), 221–224.
- Egger, P. H., Lassmann, A., 2014. Cultural Integration and Export Variety Overlap Across Countries Cultural. CESifo Working Paper 4800.
- Felbermayr, G. J., Jung, B., 2009. The pro-trade effect of the brain drain: Sorting out confounding factors. *Economics Letters* 104 (2), 72–75.
- Ginsburgh, V. A., Prieto-Rodriguez, J., 2011. Returns to Foreign Languages of Native Workers in the EU. *Industrial and Labor Relation Review* 64 (3), 599–618.
- Gould, D. M., 1994. Immigrant Links to the Home Country: Empirical Implications of U.S: Bilateral Trade Flows. *Review of Economics and Statistics* 76 (2), 302–316.
- 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., Rubinstein, Y., 2008. Estimating trade flows: trading partners and trading volumes. *Quarterly Journal of Economics* 73 (2), 441–487.
- Helpman, E., Melitz, M. J., Yeaple, S. R., 2004. Export Versus FDI with Heterogeneous Firms. *American Economic Review* 94 (1), 300–316.
- Javorcik, B. S., Özden, c., Spatareanu, M., Neagu, C., 2011. Migrant networks and foreign direct investment. *Journal of Development Economics* 94 (2), 231–241.
- Kleinert, J., Toubal, F., 2010. Gravity for FDI. *Review of International Economics* 18 (1), 1–13.

- Kugler, M., Rapoport, H., Levintal, O., 2013. Migration and Cross-Border Financial Flows. IZA Working Paper (7548).
- Leblang, D., 2010. Familiarity Breeds Investment: Diaspora Networks and International Investment. *American Political Science Review* 104 (3), 584–600.
- Mayda, A. M., 2009. International migration: a panel data analysis of the determinants of bilateral flows. *Journal of Population Economics* 23 (4), 1249–1274.
- Melitz, J., Toubal, F., 2014. Native language, spoken language, translation and trade. *Journal of International Economics* 93 (2), 351–363.
- Munshi, K., 2003. Networks in the modern economy: mexican migrants in the U.S. labor market. *Quarterly Journal of Economics* 118 (2), 549–599.
- OECD, 2014. OECD International Direct Investment Statistics 2014, OECD Publishing, Paris.
- Oldenski, L., 2012. Export Versus FDI and the Communication of Complex Information. *Journal of International Economics* 87 (2), 312–322.
- Parsons, C. R., 2012. Do Migrants Really Foster Trade? a Panel Approach 1960 – 2000. World Bank Policy Research Working Paper 6034.
- 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.
- Portes, R., Rey, H., 2005. The determinants of cross-border equity flows. *Journal of International Economics* 65 (2), 269–296.
- Rauch, J. E., 2001. Business and Social Networks in International Trade. *Journal of Economic Literature* 39 (December), 1177–1203.
- Rauch, J. E., Trindade, V., 2002. Ethnic chinese networks in international trade. *Review of Economics and Statistics* 84 (February), 116–130.
- Santos Silva, J. M. C., Tenreyro, S., 2015. Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Rubinstein Model Empirically. *Oxford Bulletin of Economics and Statistics* 77 (1), 93–105.
- Selmier, W. T., Oh, C. H., 2013. The Power of Major Trade Languages in Trade and Foreign Direct Investment. *Review of International Political Economy* 20 (3), 486–514.
- Silva, J. M. C. S., Tenreyro, S., 2006. The Log Of Gravity. *Review of Economics and Statistics* 88 (November), 641–658.

- Stock, J. H., Watson, M. W., 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica* 76 (1), 155–174.
- Westerlund, J., Wilhelmsson, F., 2011. Estimating the gravity model without gravity using panel data. *Applied Economics* 43 (6), 641–649.
- Widmaier, S., Dumont, J.-C., 2011. Are recent immigrants different? A new profile of immigrants in the OECD based on DIOC 2005/06. OECD Social, Employment and Migration Working Papers No. 126, Directorate for Employment, Labour and Social Affairs, OECD Publishing.
- Wooldridge, J. M., May 1999. Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* 90 (1), 77–97.
- Özden, C., Parsons, C. R., Schiff, M., Walmsley, T. L., 2011. Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000. *The World Bank Economic Review* 25 (1), 12–56.

7 Tables

Table 1: *Descriptive statistics for key variables as used in main specification*

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
FDI outstock in mill. USD	11,999	3,180	8.59	18,583	0	595,139
Exports in mill. USD	11,999	2,207	74.5	9,981	0	280,710
Log bilateral immigrant stock	11,999	6.26	6.35	3.40	0.00	16.3
Scarcity of bilateral language ties ($1 - \alpha_{ij}$)	11,999	0.78	0.91	0.27	0.00	1.00
$(1 - \alpha_{ij}) \times$ Log bilateral immigrant stock	11,999	4.70	4.58	3.07	0.00	14.2
Log GDP of reporter	11,999	6.15	5.93	1.42	2.19	9.62
Log GDP of partner	11,999	10.4	10.5	0.48	8.77	11.5
Log GDP per capita of reporter	11,999	3.61	3.60	2.35	-3.84	8.90
Log GDP per capita of partner	11,999	8.33	8.40	1.56	4.68	11.4

Notes: For sources of variables see text. GDP in billion USD and GDP per capita in 1000 USD before taking logs.

Table 2: *Migrants, exports, and FDI - alternative estimators and fixed effects*

Panel A: Estimates without conditioning on unrestricted sample

Estimator	PPML		Panel Poisson with FE			
	(1) Exports	(2) FDI	(3) Exports	(4) FDI	(5) Exports	(6) FDI
Lagged log(bilateral migrants)	0.207*** (0.011)	0.269*** (0.019)	0.051* (0.027)	-0.059 (0.060)	0.003 (0.082)	-0.460* (0.237)
Lagged log(bilateral migrants) · common language scarcity					0.066 (0.107)	0.526* (0.292)
Gravity variables	✓	✓	✓	✓	✓	✓
Reporter & Partner FE	✓	✓				
Bilateral FE			✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Used observations	11,999	11,973	11,616	7,442	11,616	7,442

Panel B: Preferred panel Poisson fixed effects specifications on better quality migration data

Dependent variable	(7)	(8)	(9)	(10)	(11)	(12)
	Exports	FDI	Exports	FDI	Exports	FDI
Subsample	all	all	all	all	all available	
Lagged log(bilateral migrants)	0.068 (0.090)	-0.522** (0.214)	-0.050 (0.090)	-0.739*** (0.224)	0.102*** (0.030)	-0.040 (0.107)
Lagged log(bilateral migrants) · common language scarcity	0.038 (0.125)	0.799*** (0.312)	0.090 (0.104)	1.022*** (0.292)		
Lagged log(bilateral migrants) · Lang ^F					-0.035 (0.041)	0.238** (0.020)
Gravity variables	✓	✓	✓	✓	✓	✓
Bilateral FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓			✓	✓
Reporter-year and partner-year FE			✓	✓		
Non-zero observations	7,432	4,733	7,432	4,733	5,426	4,126
Number of bilateral non-zero observations	1,960	1,204	1,960	1,204	1,383	1,014

Panel C: Estimates using panel Poisson with fixed effects and language scarcity by ease of doing business

Dependent variable	(13)	(14)	(15)	(16)	(17)	(18)
	Exports	Exports	FDI	FDI	FDI	FDI
Subsample	Time to import		Property registration time			
	< p(50)	≥ p(50)	< p(50)	≥ p(50)	< p(50)	≥ p(50)
Lagged log(bilateral migrants)	0.007 (0.089)	0.704* (0.387)	-0.474** (0.215)	0.576 (0.643)	-0.514** (0.227)	-0.465 (0.481)
Lagged log(bilateral migrants) · common language scarcity	0.113 (0.119)	-0.731 (0.450)	0.700** (0.296)	-0.180 (0.779)	0.865** (0.343)	0.922 (0.650)
Gravity variables	✓	✓	✓	✓	✓	✓
Bilateral FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Non-zero observations	3,580	3,691	2,902	1,717	2,657	1,984
Number of bilateral non-zero observations	1,002	1,074	772	520	696	567

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Panel A uses any available data, Panel B and Panel C work with data that excludes large artificial jumps in the migration data that occurred in 2002-2004 due to definitorial changes, amnesties, etc.. These would otherwise create artificial within-bilateral-tie variation biasing the estimates. Number of non-zero observations and non-zero bilateral ties reported for the respective estimation. Gravity variables are reporter's and partner's respective log GDP and log GDP per capita, common legal system, common currency, regional trade agreement, log distance, contiguity, colony, common colonizer, common official language, common spoken language. Time invariant ones dropped automatically. The median of scarcity is 0.91. A scarcity value of 0.2 corresponds to its 8th percentile. Time to import and time to register property in panel C are measured in days. Columns 11 and 12 report estimates not based on the common spoken language variable, but the alternative Lang^F variable of Dow and Karunaratna (2006), which indicates differences in languages between countries. The index incorporates linguistic similarity as well as how frequently languages are spoken in reporter and partner country. R² could only be calculated for columns 1 and 2 (0.945, 0.882) and columns 9 and 10 (0.999, 0.996).

Table 3: *Migrants, exports, and FDI - By share of population speaking English in partner country*

Share of English speakers in partner country	(1) < 0.1	(2) < 0.2	(3) < 0.3	(4) < 0.4	(5) < 0.5	(6) < 0.6	(7) < 0.7	(8) < 0.8	(9) < 0.9
<i>Dependent variable: FDI</i>									
Lagged log(bilateral migrants)	0.438* (0.237)	0.377* (0.194)	0.140 (0.095)	0.158* (0.084)	0.159* (0.083)	0.146* (0.081)	0.158** (0.079)	0.149* (0.078)	-0.087 (0.110)
Non-zero observations	1,898	2,250	2,928	3,302	3,559	3,818	3,882	3,967	4,358
Number of non-zero bilateral ties	526	616	770	854	914	975	989	1,015	1,114
<i>Dependent variable: Exports</i>									
Lagged log(bilateral migrants)	-0.068 (0.069)	0.002 (0.080)	0.049 (0.044)	0.062* (0.037)	0.063* (0.037)	0.071** (0.034)	0.073** (0.033)	0.074** (0.033)	0.090*** (0.031)
Non-zero observations	3,738	4,224	4,976	5,436	5,803	6,142	6,210	6,315	6,857
Number of non-zero bilateral ties	1,031	1,161	1,336	1,446	1,534	1,619	1,635	1,667	1,814

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 All estimations specified as in Table 2, columns 7 and 8 using bilateral fixed effects and year fixed effects. Number of non-zero observations and non-zero bilateral ties reported in each case.

Table 4: *Migrants, exports, and FDI - third-country migrants interacted with common language scarcity*

Dependent variable	(1) Exports	(2) FDI	(3) Exports	(4) FDI	(5) Exports	(6) FDI
Lagged log(bilateral migrants)	0.084 (0.091)	-0.528** (0.211)	0.033 (0.090)	-0.478** (0.212)	0.059 (0.090)	-0.475** (0.208)
Lagged log(bilateral migrants) · common language scarcity	0.044 (0.117)	0.785*** (0.301)	0.102 (0.118)	0.691** (0.297)	0.055 (0.118)	0.693** (0.284)
Lagged log(same language migrants)	-0.055 (0.050)	0.084 (0.219)				
Lagged log(same language migrants) with common language scarcity > 0.9			-0.164** (0.076)	0.388 (0.297)		
Lagged log(same language migrants) with common language scarcity > 0.95					-0.069 (0.082)	0.780*** (0.300)
Gravity variables	✓	✓	✓	✓	✓	✓
Bilateral FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Non-zero observations	7,432	4,733	7,432	4,733	7,432	4,733
Number of non-zero bilateral ties	1,960	1,204	1,960	1,204	1,960	1,204

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of non-zero observations and non-zero bilateral ties reported for the respective estimation. The scarcity of language ties is captured as $1 - \alpha_{ij}$ (1- common spoken language). Gravity variables are reporter's and partner's respective log GDP and log GDP per capita, common legal system, common currency, regional trade agreement, log distance, contiguity, colony, common colonizer, common official language, common spoken language.

Table 5: *Migrants, exports, and FDI - migrants from the partner country's neighbors*

Dependent variable	(1) Exports	(2) FDI	(3) Exports	(4) FDI	(5) Exports	(6) FDI
Lagged log(bilateral migrants)	0.072 (0.090)	-0.527** (0.216)	0.045 (0.100)	-0.401** (0.193)	0.069 (0.092)	-0.503** (0.211)
Lagged log(bilateral migrants) · common language scarcity	0.036 (0.119)	0.799*** (0.293)	0.069 (0.131)	0.628** (0.280)	0.041 (0.122)	0.754*** (0.286)
Lagged log (migrants from partner country's neighbors)	-0.004 (0.006)	0.017 (0.014)	-0.005 (0.008)	0.014 (0.012)	-0.002 (0.008)	-0.019 (0.030)
Lagged log (migrants from partner country's neighbors [csl-weighted])			0.040 (0.077)	-0.299* (0.179)		
Lagged log (migrants from partner country's neighbors [csl-weighted]) common language scarcity			-0.044 (0.083)	0.356* (0.188)	-0.004 (0.013)	0.061* (0.033)
Gravity variables	✓	✓	✓	✓	✓	✓
Bilateral FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Non-zero observations	7,432	4,733	5,868	3,945	5,868	3,945
Number of non-zero bilateral ties	1,960	1,204	1,510	990	1,510	990

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Number of non-zero observations and non-zero bilateral ties reported for the respective estimation. The scarcity of language ties is captured as $1 - \alpha_{ij}$ (1- common spoken language). Gravity variables are reporter's and partner's respective log GDP and log GDP per capita, common legal system, common currency, regional trade agreement, log distance, contiguity, colony, common colonizer, common official language, common spoken language. Time invariant ones dropped. The neighborhood of a country of origin is defined as adjacent countries. The note "csl-weighted" indicates that migrants from neighboring countries $m = 1, \dots, N$ are multiplied by the common language α_{im} they have with the respective adjacent country of origin.

Table 6: *Diminishing returns and high skilled migrants*

Dependent variable	(1) Exports	(2) FDI	(3) Exports	(4) FDI	(5) Exports	(6) FDI	(7) Exports	(8) FDI
Subsample	$< p(75)$	mig stock	w/o fav. destination		w/ mig skill data		only best data	
Lagged log(bilateral migrants)	0.406** (0.171)	-1.267** (0.579)	0.071 (0.097)	-0.603*** (0.232)			0.275*** (0.099)	-0.238 (0.271)
Lagged log(bilateral migrants) · common language scarcity	-0.383* (0.216)	1.463** (0.681)	0.028 (0.126)	0.767** (0.312)			-0.221* (0.126)	0.681* (0.352)
Lagged log(low and medium skilled bilateral migrants)					0.073 (0.060)	0.026 (0.210)		
Lagged log(high skilled bilateral migrants)					-0.167 (0.250)	-1.055** (0.536)		
Lagged log(high skilled bilateral migrants) · common language scarcity					0.480 (0.327)	1.486** (0.757)		
Gravity variables	✓	✓	✓	✓	✓	✓	✓	✓
Bilateral FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Non-zero observations	4,771	2,468	6,933	4,292	5,430	3,731	5,829	3,817
Number of non-zero bilateral	1,311	658	1,835	1,095	1,376	903	1,571	984

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Number of non-zero observations and non-zero bilateral ties reported for the respective estimation. Gravity variables are reporter's and partner's respective log GDP and log GDP per capita, common legal system, common currency, regional trade agreement, log distance, contiguity, colony, common colonizer, common official language, common spoken language. In columns 1 and 2 the bilateral ties with the 25% largest migration stocks are dropped. In columns 3 and 4 the favorite destination per country of origin is dropped. In column 5 and 6 the sample size dropped because of missing data on skills of bilateral migrants. In column 7 and 8 only countries which provide migration data from population registers or residence permits are included. Countries which use surveys to estimate immigrant numbers are excluded.

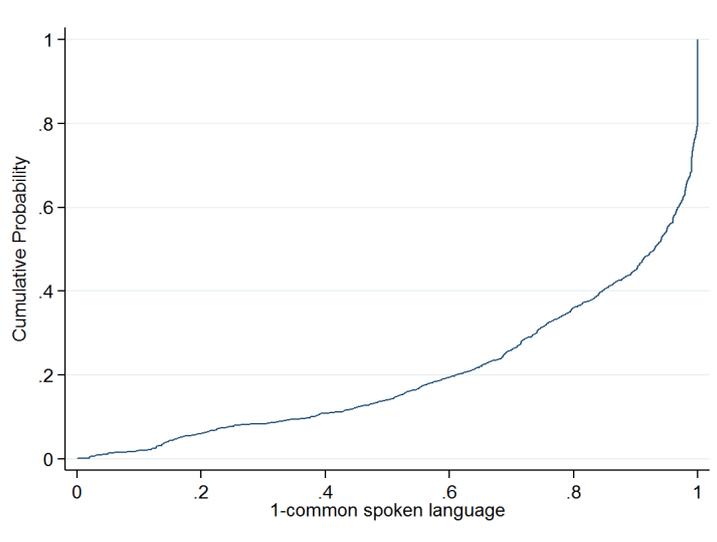


Figure 1: Cumulative density function of language scarcity ($1 - \alpha_{ij}$)

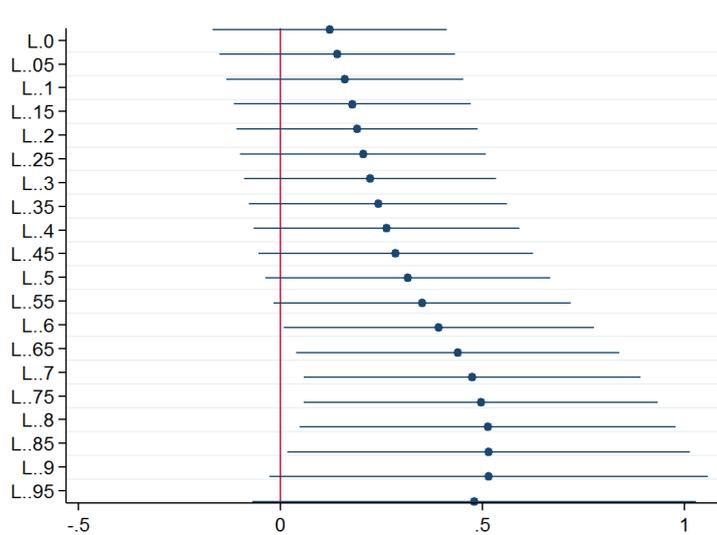


Figure 2: Point estimates and 95% confidence intervals for regressions that use $((1 - \alpha_{ij}) - p)/(1 - p) \cdot \text{lagged } \log(\text{migrants})$ above a threshold $p \in [0, 1)$ as the only migration related covariate

Online Appendix

Table A1: *Robustness check - other gravity variables*

Dependent variable	(1)	(2)	(3)	(4)	(5)
Interaction	FDI	FDI	FDI	FDI	FDI
	Baseline	dist	col	lp1	colony
Lagged log(bilateral migrants)	-0.066 (0.107)	0.011 (0.110)	-0.025 (0.106)	-0.033 (0.127)	-0.042 (0.108)
Distance · lagged log(bilateral migrants)		-0.000 (0.000)			
Common official language · lagged log(bilateral migrants)			-0.439 (0.295)		
Language proximity · lagged log(bilateral migrants)				-0.015 (0.043)	
Colony · lagged log(bilateral migrants)					-0.303 (0.290)
Gravity variables	✓	✓	✓	✓	✓
Bilateral FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Non-zero observations	4,733	4,733	4,733	4,733	4,733
Non-zero bilateral ties	1,204	1,204	1,204	1,204	1,204

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of non-zero observations and non-zero bilateral ties reported for the respective estimation. Dist indicates distance. Col stands for common official language. lp1 is language proximity from the Melitz and Toubal dataset. Remaining specification as in the Table 2, column 8.

Table A2: *Robustness check - other gravity variables and language scarcity interaction*

Dependent variable	(1)	(2)	(3)	(4)	(5)
Interaction	FDI Baseline	FDI dist	FDI col	FDI lp1	FDI colony
Lagged log(bilateral migrants)	-0.522** (0.214)	-0.458** (0.196)	-0.485** (0.223)	-0.606*** (0.232)	-0.498** (0.212)
Lagged log(bilateral migrants) · common language scarcity	0.799*** (0.291)	0.914*** (0.275)	0.752** (0.303)	0.848*** (0.288)	0.779*** (0.286)
Distance · lagged log(bilateral migrants)		-0.000** (0.000)			
Common official language · lagged log(bilateral migrants)			-0.107 (0.309)		
Language proximity · lagged log(bilateral migrants)				0.027 (0.039)	
Colony · lagged log(bilateral migrants)					-0.158 (0.270)
Gravity variables	✓	✓	✓	✓	✓
Bilateral FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Non-zero observations	4,733	4,733	4,733	4,733	4,733
Non-zero bilateral ties	1,204	1,204	1,204	1,204	1,204

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of non-zero observations and non-zero bilateral ties reported for the respective estimation. Remaining specification as in the Table 2, column 8.

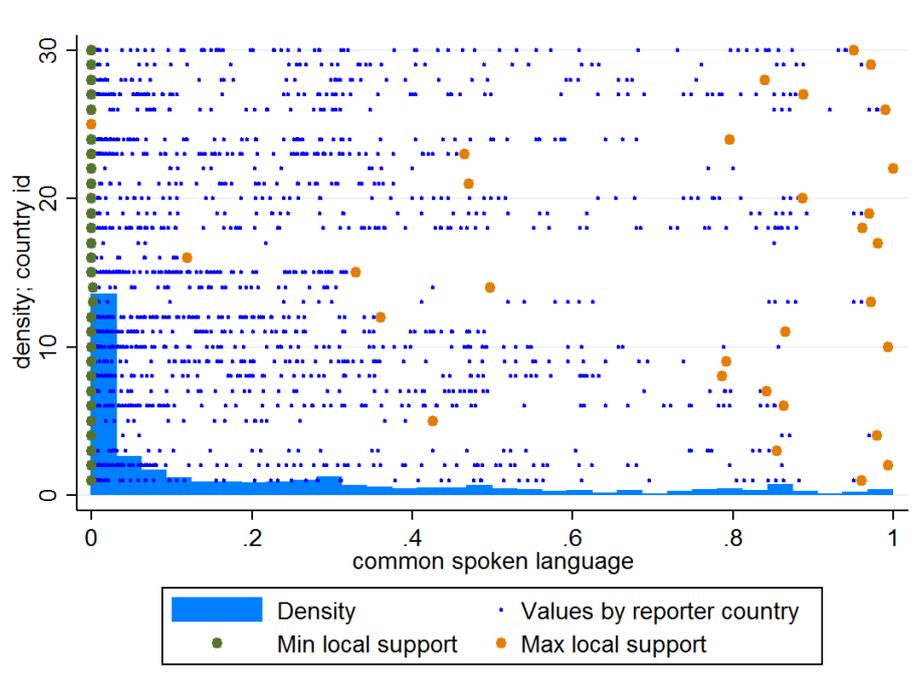


Figure A1: Distribution of common spoken language variable and support by reporting country ("local")

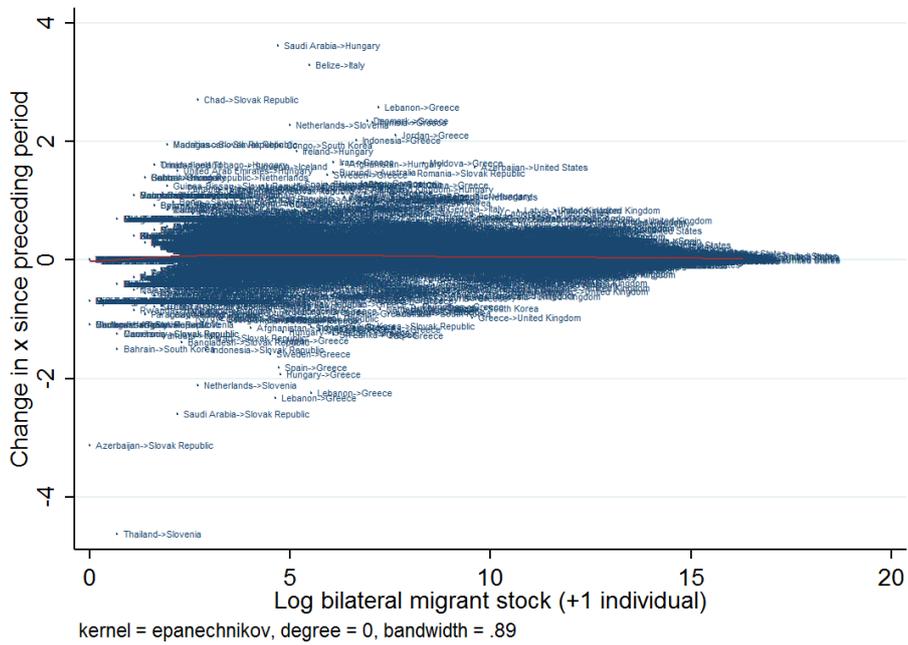


Figure A2: Correlation of initial migration stock and period-on-period change in migration stock

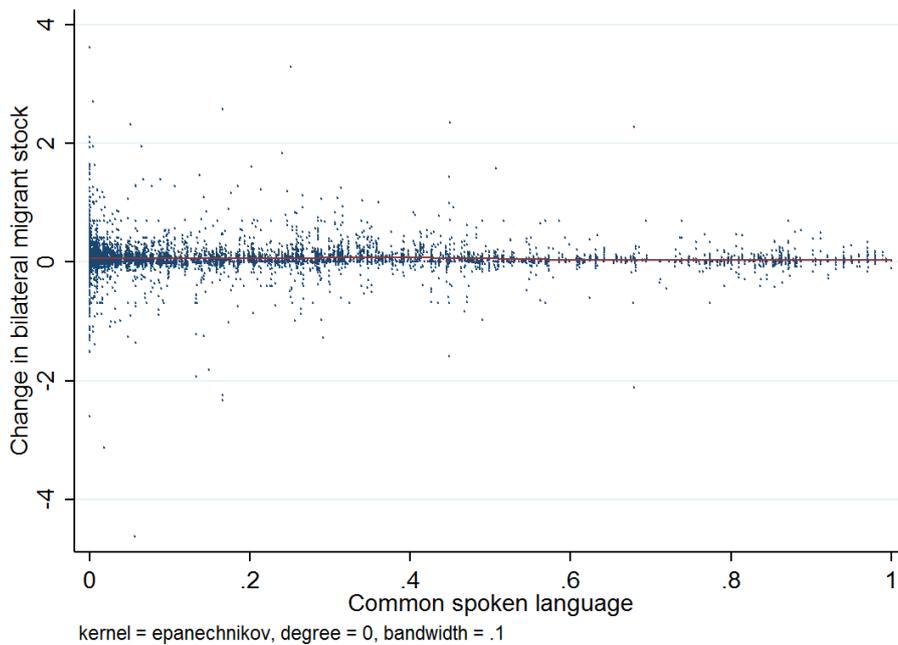


Figure A3: Correlation of common spoken language and period-on-period change in migration stock