DISCUSSION PAPER SERIES

IZA DP No. 9419

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

Matthias Lücke
Tobias Stöhr

October 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Matthias Lücke
Kiel Institute for the World Economy

Tobias Stöhr
Kiel Institute for the World Economy, DIW Berlin and IZA

Discussion Paper No. 9419
October 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9419
October 2015

ABSTRACT

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

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 a robust 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.

JEL Classification: F21, F22, O14

Keywords: migration, FDI, foreign languages, globalization

Corresponding author:

Tobias Stöhr
Kiel Institute for the World Economy
Kiellinie 66
24105 Kiel
Germany
E-mail: tobias.stoehr@ifw-kiel.de

* 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, and Rainer Thiele. 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).

Most of these studies are based on a gravity model 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. It addresses 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 leads to 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.

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]

Second, 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 4 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.

Third, 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 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.

Fourth, 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 those 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.

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

[2] In 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.

[3] Aubry 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.

2 Identification

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 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 lagged immigrant stocks to explain FDI stocks. Thus we estimate 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 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

[4]These 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.

Wilhelmsson (2011), we are interested in the simple gravity equation with bilateral fixed effects

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

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 lagged bilateral log migrant stock which vary over time as well as across country pairs. α_{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

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

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. Global bilateral migrant stock matrices are available from the World Bank only every five or ten years. This is insufficient for our analysis where we seek to eliminate unobserved heterogenenity at the country-pair level as much as possible. Therefore we use OECD panel data that provide good coverage of immigrants in OECD countries from all countries of the world.

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. It seems unlikely that combining data based on different definitions will affect our findings, given the relatively short time horizon of our panel analysis.

Our final dataset includes all OECD countries except Luxembourg and Turkey.[5] Data

[5]Luxembourg 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

range from 2004 to 2011, with an average of 4.24 data points per country pair.

We match our immigrant stock data with FDI outstocks and outflows from OECD sources and UN Comtrade data on exports and imports. 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 from Mayer and Zignago (2011), updated from the World Bank's World Development Indicators. Finally, country data on the extent of bureaucracy and other barriers to investment are from the World Bank's Doing Business Report.

The final dataset consists of 12,345 bilateral observations in 2,910 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.[6]

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 $l = 1, ..., L$, 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: $\alpha_{ij} = \sum_l L_{li}L_{lj}$.[7] 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

a single year, making panel estimation impossible.

[6] Here we follow Ginsburgh and Prieto-Rodriguez (2011) in our reasoning.

[7] 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.

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.

In our data, "common language scarcity" is distributed between 0 and 1 and provides sufficient common support (Figure 1) to make inferences on a cross-country basis. The average is 0.79 and the median 0.92. 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 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:

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

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 four different specifications: the PPML estimator (often used in the literature) without bilateral fixed effects (Columns 1 and 2); PPML with bilateral fixed effects (Columns 3 and 4); our preferred panel Poisson estimator with bilateral and year fixed effects, but no gravity variables (Columns 5 and 6); and the panel Poisson setup with bilateral and year fixed effects and gravity variables that will be used in the rest of the paper (Columns 7 and 8).

[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 (Columns 1 and 2). When we replace reporter and partner fixed effects with bilateral fixed effects (Columns 3 and 4) to address the omitted variable bias inherent in the first specification (cf. Section 2), we still observe a positive effect of the immigrant stock but the corresponding coefficient is substantially smaller, especially for FDI (Column 4 vs 2).

With our preferred panel Poisson estimator, which relies only on variation over time, we find a positive significant coefficient for bilateral immigrant stocks only for exports (Column 7), but not FDI (Column 8). 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 .

Next, we investigate our main hypothesis by allowing the impact of the bilateral immigrant stock on FDI to vary with our "common language scarcity" variable (Table 3). The coefficient of the corresponding interaction term is significantly positive for FDI (Column 2), but not for exports (Column 1). As hypothesized, bilateral immigrants facilitate FDI the more, the scarcer are the common language skills between the two countries. This effect remains of similar size when we eliminate countries with a particularly poor business climate from our sample because FDI in these countries may be subject to other determinants (Columns 4 and 6). By contrast, although we have earlier found that exports are facilitated by bilateral immigrants (Table 2, Column 7), the size of this effect does not significantly depend on common language scarcity (Table 3, Column 1).

[Table 3 about here.]

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, FDI implies an emphasis on internalizing a wide range of business processes within the firm. Headquarter employees contribute their specialized skills to the affiliate, including through expatriate placements. Affiliates employ host country residents who often speak the firm's international working language. Against this background, bilateral immigrants in the FDI country of origin, unless they have specialized skills relevant to the firm, may have little to contribute beyond their language skills to reducing international transaction costs for the firm. It seems plausible, furthermore, that immigrants'

language skills come into play only when residents of the FDI origin and host countries find it difficult to communicate due to a scarcity of common languages.

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 3, Column 2). This result is implausible and 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 2). Within this range, the composite effect of immigrants on FDI is predicted to be positive: given the baseline coefficient estimates (Table 3, Column 2), the composite effect of the immigrant stock on FDI exceeds 0 if common language scarcity exceeds 0.65.

[Figure 2 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; no other migration-related covariates are included. The hypothesized relationship between immigrants and FDI is nil up to the respective threshold and linear thereafter.[8] Figure 3 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 (Table 3). 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.

[Figure 3 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 0.92 (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

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

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

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, it turns out that the coefficient estimates become erratic when all four migrant-related variables are included. With an average of only 5 data points for each country pair and large numbers of zeros, our panel data apparently 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. Thus for country pairs with few common language skills, third-country immigrants with the "right" languages may act as substitutes for bilateral immigrants.[9] This substitutive is also statistically significant using a function with a kink such as the one discussed earlier to model the scarcity threshold.

[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 to 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 if common language skills between FDI origin and host country are also scarce, immigrants from neighboring countries have a small but significant positive effect on bilateral outward FDI (Table 5, Columns 4 and 6). Once again, the coefficients for bilateral immigrants are almost unchanged from our basic results (Table 3).

[Table 5 about here.]

[9] The 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.

So far, we have assumed a double-log-linear relationship between immigrants and outward FDI. We now explore whethere 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 3). 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 to distinguish between high-skilled (tertiary-educated) and other migrants (Table 6, column 5 and 6). As these data are only available for the year 2000 we assume that the skill ratio observed in 2000 is representative of our whole period of observation (2004 to 2011). Our regression results are therefore highly tentative (Table 6, Columns 5 and 6). We find this assumption still more plausible than to apply the total skill ratio of all emigrants by country of origin to every individual country of destination. With this caveat, 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 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 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. In order to check the robustness of our results, we have run placebo regressions that rearrange the chronological order of events in our presumed causal chain. Reassuringly, when we replace the lags of the immigrant stock and interaction 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.

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). We 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 4). Therefore, Card-style shift-share instrumental variables lack instrumental relevance for this study. There is no also evidence of significantly more bilateral immigration based on common spoken languages in the short run (Figure 5). Furthermore, the results are robust to the exclusion of longstanding bilateral migration ties, based on per capita bilateral immigrant stocks in 1960.

We have tested extensively whether the immigrant stock - language scarcity interaction might reflect any other correlated factor. In Table 7, we report exemplary regression results 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). For none of these variables are we able to establish a pattern similar to that for common language scarcity. Furthermore, we show in Table 8 that including these additional interaction terms does not substantially affect our baseline estimates.

[Tables 7 and 8 about here.]

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.[10]

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 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.

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 find a positive effect on bilateral FDI outstocks through this group.

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 facil-

[10]For 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.

itate 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

Aubry, A., Kugler, M., Rapoport, H., 2012. Migration, FDI and the Margins of Trade. 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.

Mayer, T., Zignago, S., 2011. Notes on CEPII's distance measures: the GeoDist database. CEPII Working Paper (2011-25).

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.

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.

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.

Wooldridge, J. M., May 1999. Distribution-free estimation of some nonlinear panel data models. Journal of Econometrics 90 (1), 77–97.

7 Tables

Table 1: *Descriptive statistics for key variables*

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
FDI outstock in mill. USD	7,861	3,641	11.7	20,460	0	595,139
Exports in mill. USD	7,861	2,524	101.2	11,038	0	280,710
Log bilateral immigrant stock	7,861	6.49	6.58	3.34	0	16.28
Scarcity of bilateral language ties $(1 - \alpha_{ij})$	7,861	0.79	0.92	0.26	0	1
$(1 - \alpha_{ij})$ x Log bilateral immigrant stock	7,861	5.00	4.91	3.04	0	14.18
Log GDP of reporter	7,861	6.20	5.99	1.48	2.53	9.62
Log GDP of partner	7,861	10.50	10.61	0.50	9.22	11.50
Log GDP per capita of reporter	7,861	3.94	3.94	2.27	-2.25	8.90
Log GDP per capita of partner	7,861	8.44	8.48	1.52	4.97	11.40

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 using poisson pseudo maximum likelihood (ppml)

	(1)	(2)	(3)	(4)
Estimator	PPML			
Dependent variable	Exports	FDI	Exports	FDI
Lagged log(bilateral migrants)	0.198***	0.285***	0.172***	0.206***
	(0.013)	(0.022)	(0.026)	(0.036)
Gravity variables	√	√	√	√
Reporter & Partner FE	√	√		
Bilateral FE			√	√
Year FE	√	√	√	√
R^2	0.942	0.882	0.951	0.721
Observations	7,861	7,861	7,861	7,861
Used observations	7,861	7,827	7,861	5,838

Panel B: Estimates using panel poisson with fixed effects

	(5)	(6)	(7)	(8)
Dependent variable	Exports	FDI	Exports	FDI
Lagged log(bilateral migrants)	0.203***	0.086	0.095***	-0.064
	(0.047)	(0.106)	(0.029)	(0.107)
Gravity variables			√	√
Reporter & Partner FE				
Bilateral FE	√	√	√	√
Year FE	√	√	√	√
R^2	n/a	n/a	n/a	n/a
Observations	12,345	12,345	12,345	12,345
Non-zero observations	7,432	4,733	7,432	4,733
Number of bilateral non-zero observations	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. 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 in panel 2.

Table 3: *Migrants, exports, and FDI - the role of common language scarcity*

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Exports	FDI	Exports	FDI	Exports	FDI
Subsample	all		$< p(75)$ t_2_import		$< p(75)$ t_2_regproperty	
Lagged log(bilateral migrants)	0.068	-0.522**	0.024	-0.538**	0.019	-0.498**
	(0.090)	(0.214)	(0.089)	(0.221)	(0.094)	(0.218)
Lagged log(bilateral migrants) ·	0.038	0.799***	0.097	0.841***	0.097	0.866***
common language scarcity	(0.118)	(0.291)	(0.119)	(0.309)	(0.125)	(0.312)
Gravity variables	√	√	√	√	√	√
Bilateral FE	√	√	√	√	√	√
Year FE	√	√	√	√	√	√
Non-zero observations	7,432	4,733	5,270	3,790	5,434	3,663
Number of non-zero bilateral ties	1,960	1,204	1,438	975	1,489	967

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. t_2_import and t_2_regproperty are the time to import in days and the time to register property in days in the partner country taken from the latest "Doing Business Report".

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

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Exports	FDI	Exports	FDI	Exports	FDI
Lagged log(bilateral migrants)	0.084	-0.528**	0.033	-0.478**	0.059	-0.475**
	(0.091)	(0.211)	(0.090)	(0.212)	(0.090)	(0.208)
Lagged log(bilateral migrants) ·	0.044	0.785***	0.102	0.691**	0.055	0.693**
common language scarcity	(0.117)	(0.301)	(0.118)	(0.297)	(0.118)	(0.284)
Lagged log(same language migrants)	-0.055	0.084				
	(0.050)	(0.219)				
Lagged log(same language migrants)			-0.164**	0.388		
with common language scarcity > 0.9			(0.076)	(0.297)		
Lagged log(same language migrants)					-0.069	0.780***
with common language scarcity > 0.95					(0.082)	(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*

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Exports	FDI	Exports	FDI	Exports	FDI
Lagged log(bilateral migrants)	0.072	-0.527**	0.045	-0.401**	0.069	-0.503**
	(0.090)	(0.216)	(0.100)	(0.193)	(0.092)	(0.211)
Lagged log(bilateral migrants) ·	0.036	0.799***	0.069	0.628**	0.041	0.754***
common language scarcity	(0.119)	(0.293)	(0.131)	(0.280)	(0.122)	(0.286)
Lagged log (migrants from	-0.004	0.017	-0.005	0.014	-0.002	-0.019
partner country's neighbors)	(0.006)	(0.014)	(0.008)	(0.012)	(0.008)	(0.030)
Lagged log (migrants from partner country's			0.040	-0.299*		
neighbors [csl-weighted])			(0.077)	(0.179)		
Lagged log (migrants from partner country's			-0.044	0.356*	-0.004	0.061*
neighbors [csl-weighted]) common language scarcity			(0.083)	(0.188)	(0.013)	(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 $m = 1, ..., N$ are multiplied by the common language they α_{im} they have with the respective adjacent country of origin.

Table 6: *Diminishing returns and high skilled migrants*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Exports	FDI	Exports	FDI	Exports	FDI	Exports	FDI
Subsample	$< p(75)$ mig stock		w/o fav. destination		w/ mig skill data		only best data	
Lagged log(bilateral migrants)	0.406**	-1.267**	0.071	-0.603***			0.275***	-0.238
	(0.171)	(0.579)	(0.097)	(0.232)			(0.099)	(0.271)
Lagged log(bilateral migrants) ·	-0.383*	1.463**	0.028	0.767**			-0.221*	0.681*
common language scarcity	(0.216)	(0.681)	(0.126)	(0.312)			(0.126)	(0.352)
Lagged log(low and					0.073	0.026		
medium skilled bilateral migrants)					(0.060)	(0.210)		
Lagged log(high skilled bilateral					-0.167	-1.055**		
migrants)					(0.250)	(0.536)		
Lagged log(high skilled bilateral					0.480	1.486**		
migrants) · common language scarcity					(0.327)	(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.

Table 7: *Robustness check - other gravity variables*

	(1)	(2)	(3)	(4)	(5)
Dependent variable	FDI	FDI	FDI	FDI	FDI
Interaction	Baseline	dist	col	lp1	colony
Lagged log(bilateral migrants)	-0.066	0.011	-0.025	-0.033	-0.042
	(0.107)	(0.110)	(0.106)	(0.127)	(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. 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.

Table 8: *Robustness check - other gravity variables and language scarcity interaction*

	(1)	(2)	(3)	(4)	(5)
Dependent variable	FDI	FDI	FDI	FDI	FDI
Interaction	Baseline	dist	col	lp1	colony
Lagged log(bilateral migrants)	-0.522**	-0.458**	-0.485**	-0.606***	-0.498**
	(0.214)	(0.196)	(0.223)	(0.232)	(0.212)
Lagged log(bilateral migrants) · common language scarcity	0.799***	0.914***	0.752**	0.848***	0.779***
	(0.291)	(0.275)	(0.303)	(0.288)	(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. 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.

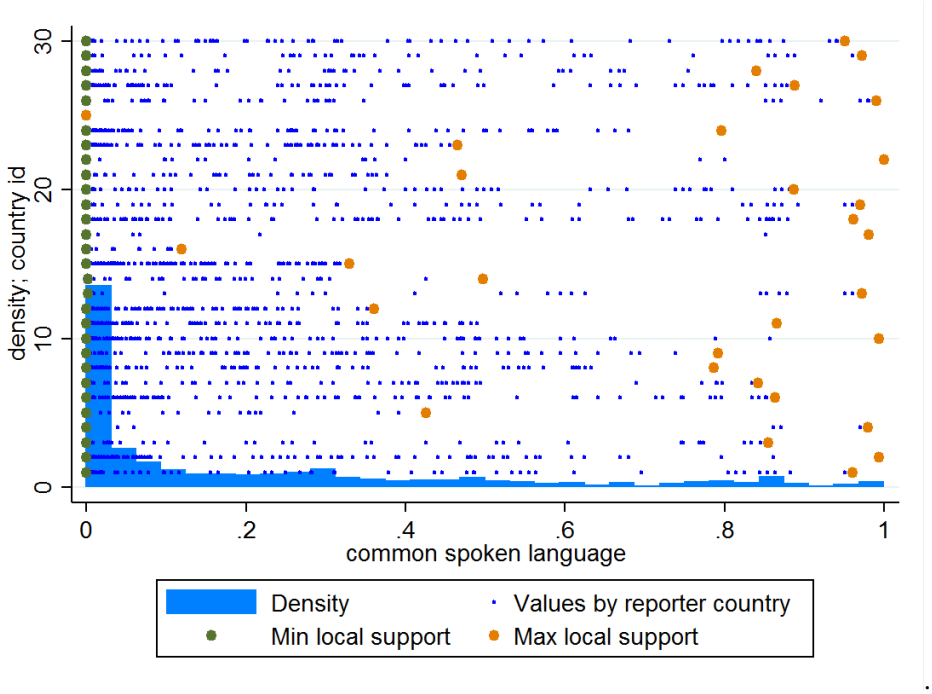

Figure 1: Distribution of common spoken language variable and support by reporting country (”local”)

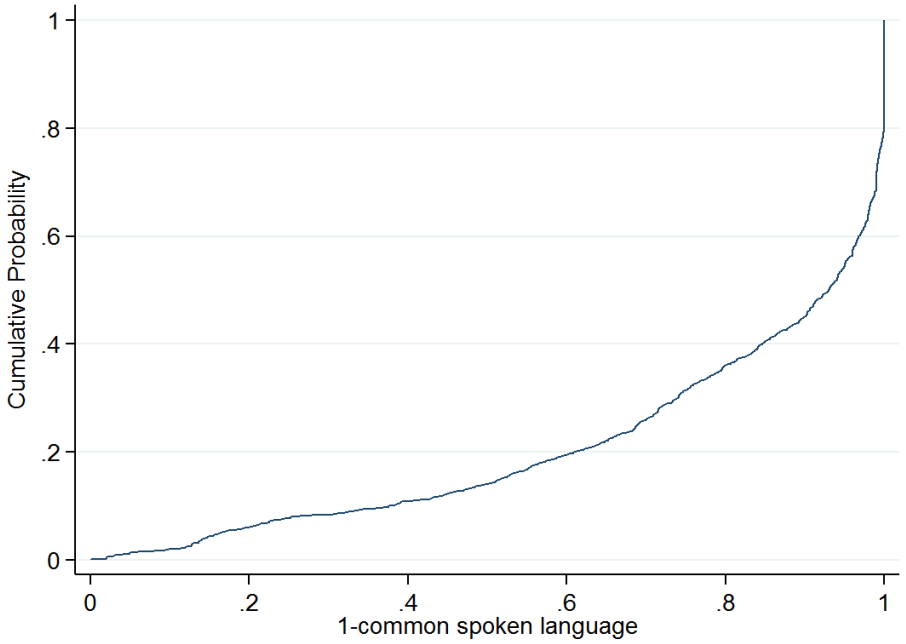

Figure 2: Cumulative density function of language scarcity $(1 - \alpha_{ij})$

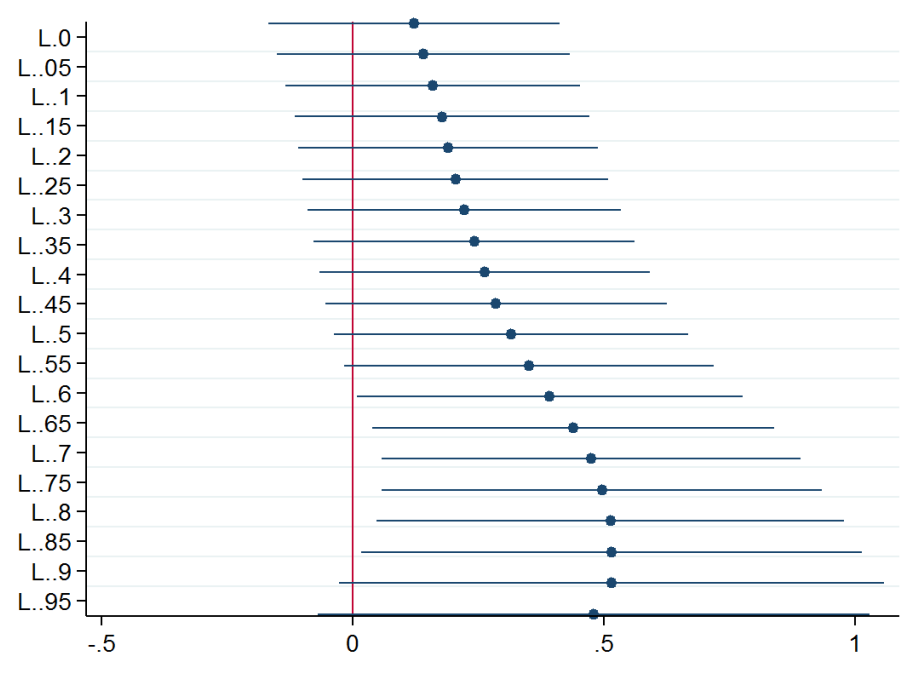

Figure 3: Point estimates and 95% confidence intervals for regressions that use $((1 - \alpha_{ij}) - p)/(1 - p) \cdot lagged\ log(migrants)$ above a threshold $p \in [0, 1)$ as the only migration related covariate

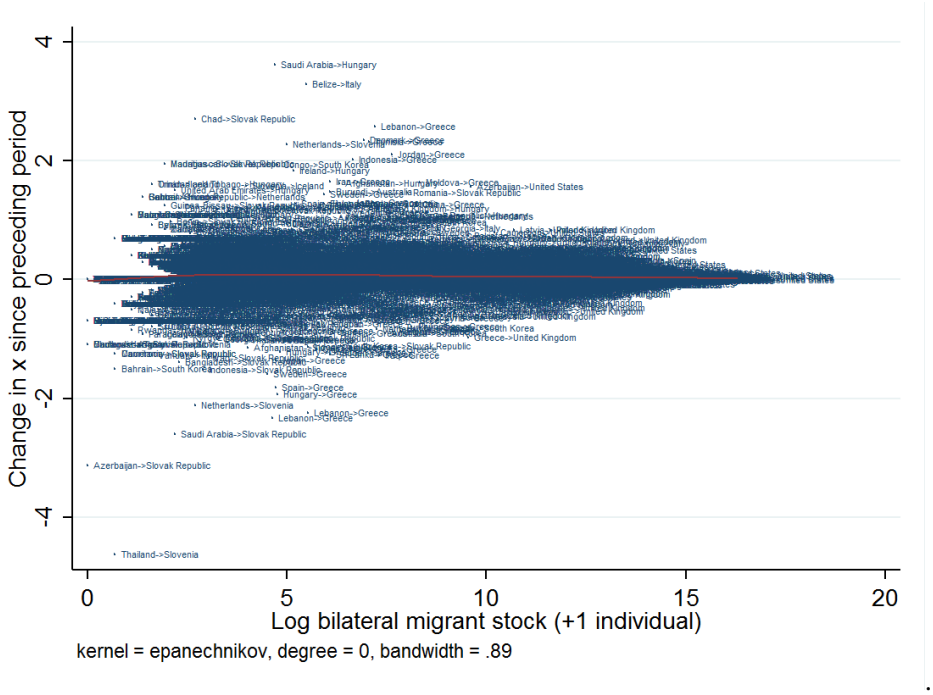

Figure 4: Correlation of initial migration stock and period-on-period change in migration stock

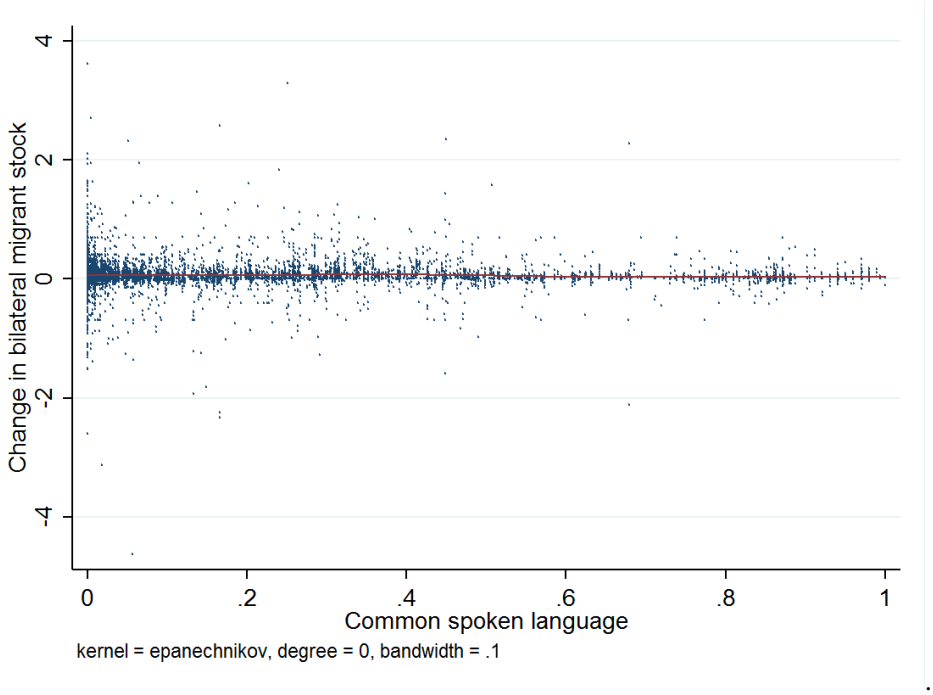

Figure 5: Correlation of common spoken language and period-on-period change in migration stock