

Lücke, Matthias; Stöhr, Tobias

Conference Paper

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

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Foreign Direct Investment I, No. D05-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Lücke, Matthias; Stöhr, Tobias (2015) : Heterogeneous Immigrants and Foreign Direct Investment: The Role of Language Skills, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Foreign Direct Investment I, No. D05-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/113191>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

This version: February 15, 2015

Abstract

We estimate a gravity model for bilateral FDI out-stocks from panel data for OECD reporting countries with bilateral and year fixed effects. With this demanding test, 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, in terms of size and statistical significance, for immigrants from third countries who speak the language(s) of the FDI host country. They thus are potential substitutes for bilateral migrants. A 1 percent increase in either immigrant group raises the FDI out-stock by 0.2 to 0.4 percent. Combined with various robustness checks, our findings suggest that immigrants facilitate outgoing FDI through their language skills, rather than through other characteristics like cultural familiarity. As most developing country residents have few language skills in common with rich country residents, developing country migrants in rich countries have a key role to play in facilitating FDI in their countries of origin.

Keywords: migration, FDI, foreign languages, institutions, globalization

JEL classification: F21, F22, O14

1 Motivation

One important feature of 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 trade and foreign direct investment, by reducing international communication costs. Over the past two decades, many empirical studies have identified a positive effect of immigrant stocks on foreign direct investment in the immigrants' country 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 with migrant stocks added as explanatory variables to reflect their possible impact on the cost of international economic transactions. One concern with this approach is that migration, trade, and FDI 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 effects of migrant stocks on outgoing bilateral 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 the underlying hypothesis because it addresses the impact of short-term, year-to-year variations in the number of immigrants on outgoing FDI, rather than the long-term impact that may be identified in cross-section analyses.

Second, we investigate how the impact of immigrants on outgoing bilateral FDI depends on the ease with which residents of the two countries can communicate, given their shared 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 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); recent work such as Egger and Lassmann (2014) presents robust proof of the direction from causality from language to trade.¹

Second, Melitz and Toubal (2014) develop a 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). They use global data to 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 comes rather close to our hypothesis that migrants may be more important the smaller the common spoken language variable.

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 on the part 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 a wide range of activities with an investment component, including bilateral trade (because of the fixed costs 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 as in Helpman et al. (2004). The firm's decision to undertake FDI (rather than export or produce only for the domestic market) is driven by the fixed costs of exporting vs. setting up production abroad. It is natural to think that the presence of immigrants reduces those fixed costs². On this basis, migrant stocks become part of a rigorously derived gravity 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

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

²In related work, Oldenski (2012) shows that communication-intensive industries are more likely to market their products abroad through FDI than through exports. This finding points to the importance of communication with customers, adjusting the firm's 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.

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 Metrics/Identification

Although 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), this approach is probably invalid 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 with a valid exclusion restriction at the global level. Furthermore, the bias does not occur at a country but at the country-pair level, because, 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 and country-level estimates are potentially severely biased.

Thus our analysis of the impact of migrant stocks on FDI requires 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 bilateral ties that affect all economic interactions - migration, FDI, and trade. Our bilateral fixed effects are time-invariant - which is a sensible assumption only over a relatively short horizon. In order to reduce the scope for reverse causality even further, we use lagged migrant stocks to explain FDI stocks. Thus our approach is to estimate the short term effects as precisely as possible.

When we choose our estimator, we need to 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 x_{it} are strictly exogenous.⁴

⁴These assumptions are discussed in the robustness section below. A Monte Carlo study is provided by Westerlund and Wilhelmsson (2011). Note that this is neither the negative binomial MLE nor panel qMLE.

Thus our gravity model for bilateral FDI out-stocks includes log migrant stocks, time-variant gravity variables such as GDP, GDP per capita, and free trade agreements, and time-invariant bilateral fixed effects and year fixed effects. 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 FE

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

Y_{it} and Y_{jt} contain contry-specific gravity variables that vary over time. 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

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

This approach also has the advantage that we neither have to truncate the dependent variable, which contains large amounts of zeros, nor make arbitrary non-linear transformations of the dependent variable, which have been shown to lead to severely biased and inconsistent results (cf. Silva and Tenreyro, 2006; Westerlund and Wilhelmsson, 2011).

3 Data

Overview

Of the various data required for our analysis, the bottleneck is data on immigrant stocks. Global bilateral migrant stock matrices are available from the World Bank only every five or ten years. This would be insufficient for our analysis where we seek to eliminate unobserved heterogeneity at the country-pair level to the maximum extent possible. Therefore we use the OECD DIOC database which provides panel data of immigrants in OECD countries from any country in the world. Our final dataset includes all OECD countries except the Czech Republic, Estonia, Japan, Portugal, South Korea, and Turkey, which ei-

ther define migrants differently (nationality instead of country of birth) or do not provide annual migrant stock data at all. Data range from 2004 to 2011, with 5.9 data points per country pair on average.

We match our migrant data with FDI outstocks and outflows from OECD sources and UN Comtrade data on imports and exports. Language data are from the bilateral matrices and country-specific raw data provided by Melitz and Toubal (2014). 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 taken from the World Bank’s Doing Business Report.

The final dataset consists of 6763 bilateral observations (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 that are required to do business with another country, the more valuable is the presence of immigrants from that country. As a first test, we interact migrant stocks with the share of individuals in the two countries who have no spoken language in common.⁵

We calculate this share from the “common spoken language” variable in Melitz and Toubal (2014). Melitz and Toubal estimate the chance that two randomly chosen residents from two countries have a language in common. They compile a global database of all 42 languages which are spoken by at least 4% of the population in at least two countries. For each country i they report the share L_{li} of the population who speak language l . Then, for each country pair and language $l = 1, \dots, L$, they calculate the chance that two randomly chosen residents from the two countries 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 chosen residents from the two countries have a language in common: $\alpha_{ij} = \sum_l L_{li}L_{lj}$.⁶ 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

⁵ Here our reasoning follows Ginsburgh and Prieto-Rodriguez (2011).

⁶ 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.

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.

In our data, the scarcity variable is distributed between 0.01 and 1 and provides sufficient common support (Table 1). The average value of the common spoken language variable is 0.77 but the model value is 1. The median of 0.91 shows that in most bilateral ties there are few people who can communicate when randomly matched. As expected, the most important contributor to bilateral communication is English language skills. In our estimation we combine language scarcity with the bilateral migrant stock in an interaction term.

In a further step, we are interested in whether, for a given bilateral FDI out-stock, 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, it seems plausible that immigrants in the UK from other Portuguese-speaking countries such as Brazil or Portugal may facilitate UK FDI in Mozambique in a similar fashion to immigrants from Mozambique - provided that language skills, rather than other factors like familiarity with the local institutions and culture, are at the heart of the link between migrants and FDI.

To capture the effect of third-country immigrants in reporting country i who speak one of the languages of FDI host country j , we estimate their number based on the languages spoken in the immigrants’ countries of origin:

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

We introduce ”same language migrants” from third countries in our gravity equation in parallel with bilateral migrants and investigate their respective impacts on bilateral FDI.

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 without bilateral fixed effects (Columns 1 and 2); PPML with bilateral fixed effects (Columns 3 and 4); the panel Poisson estimator with bilateral and year fixed effects, but no gravity variables (Columns 5 and 6); panel Poisson with bilateral and year fixed effects and gravity variables (Columns 7 and 8).

[Table 2 about here.]

Under PPML without bilateral fixed effects, the bilateral migrant stock has a significant positive coefficient for both exports and FDI (Columns 1 and 2). However, this significant positive effect disappears once we introduce bilateral fixed effects (Columns 3 and 4) to address the omitted variable bias inherent in the first specification (cf. Section 2). With our preferred panel Poisson estimator, we find a positive significant coefficient for bilateral migrant stocks only for exports when we omit the gravity variables (Column 5). These results suggest that the positive association between migrant stocks and FDI (or exports) that is found in many conventionally estimated gravity models (comparable to Columns 1 and 2) is probably severely upwardly biased.

Next, we investigate our main hypothesis by allowing the impact of the bilateral migrant stock on FDI to vary according to the extent of common language skills across the two countries (our "common language scarcity" variable; Table 3). The coefficient of the interaction term between language scarcity and the bilateral migrant stock is significantly positive for FDI (Column 2), but not for exports (Column 1). Therefore, as hypothesized, bilateral migrants facilitate FDI the more, the scarcer the common language skills between the two countries. The effect remains of similar size when we eliminate countries with a particularly poor business climate from our sample on the grounds that FDI in these countries may be subject to other determinants (Columns 4 and 6).

[Table 3 about here.]

The impact of bilateral migrant stocks on FDI is sizable: for a country pair with average common language scarcity at 0.77 (cf. 1), the two coefficients

sum to 0.22 for the full sample and 0.41 for the sample without the worst-performing partner countries in terms of the time it takes to register property. Thus, approximately, a 1 percent increase in the bilateral migrant stock implies a 0.2 to 0.4 percent increase in the FDI outstock. When residents in the two countries share almost no language skills (common language scarcity = 1), the corresponding increase in FDI lies between 0.4 and 0.7 percent.

Next, we investigate how immigrants from third countries who speak the same language(s) as FDI host country residents affect bilateral FDI (our "same language migrants" - cf. Section 2). It would be natural to add "same language migrants" to our gravity equation 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 6 data points for each country pair and large numbers of zeros, our panel data apparently do not permit reliable coefficient estimates for so many regressors that vary over time and across country pairs. Therefore, we include "same language migrants" in our gravity equation alternatively (i) on its own (Table 4); (ii) interacted with common language scarcity (Table 5); and (iii) with bilateral and third-country migrants added up (Table 6).

When we include "same language migrants" on its own (Table 4), we find a significant positive effect of these third-country immigrants on bilateral FDI, on top of the broadly unchanged effect for bilateral migrants. The size of the effect is broadly similar to bilateral migrants: Approximately, a 1 percent increase in the number of "same language migrants" implies a 0.4 to 0.5 percent increase in the bilateral FDI outstock.

When we interact "same language migrants" with "common language scarcity" (Table 5), we find a significant positive effect whose size varies substantially across the three samples. For a country pair with average common language scarcity, a 1 percent increase in "same language migrants" implies an increase in FDI of between 0.4 and 0.8 percent. At the same time, the coefficients for the interaction of bilateral migrants with common language scarcity are now smaller than before, at lower levels of statistical significance. Note that the size of the FDI effect is for same-language migrants is not comparable to the size of the interaction term for bilateral migrants because the baseline category differs. Given the limitations of our data we conclude that both effects are important.

[Table 5 about here.]

This broad conclusion is confirmed by our final set of results for third-country, same-language migrants. We add bilateral and same language migrants together and include this new variable both on its own and interacted with common language scarcity (Table 6). The impact of these variables on FDI is very similar in size to the impact of bilateral migrants alone. For a country pair with average common language scarcity, a 1 percent increase in "combined" migrants raises FDI by between 0.2 and 0.4 percent.

It is remarkable that we find a robust FDI effect exactly for those third-country migrants that have the "right" language skills to communicate with FDI host country residents. It is nevertheless conceivable that factors other than language may be crucial for migrants' ability to facilitate FDI. We test this hypothesis by including the number of migrants from the FDI host country's neighbors in our gravity equation (Table 7). Typically, geographic proximity is a good proxy for a cultural proximity that might facilitate FDI instead of or beyond language skills. It turns out that there is no significant effect from "neighborhood" migrants on bilateral FDI - very much in contrast to third-country, same-language migrants (cf. Table 4). The coefficients for bilateral migrants are almost unchanged from our basic results (Table 3).

[Table 7 about here.]

Finally, we use the limited available data on bilateral migrant stocks broken down by skill levels in 2000 to distinguish between high-skilled (tertiary-educated) and other migrants (Table 8). As we have to apply the skill ratio in 2000 to the whole period of observation, these results are highly tentative. They suggest that the presence of high-skilled migrants does much more to facilitate bilateral FDI than the presence of low-to-medium-skilled migrants. Without proper annual data on migrant stocks by skill levels, it is impossible to corroborate this hypothesis.

[Table 8 about here.]

5 Robustness

The exogeneity of immigration conditional on the covariates and fixed effects is an important requirement for a causal interpretation of our results. Changes in FDI might be correlated with changes in migration stocks due to some underlying factor or due to reverse causality. In order to check the robustness of our results, we ran placebo regressions that rearrange the chronological order of events in our

presumed causal chain. Reassuringly, replacing the lags of the migrant stocks and the interactions with leads in order to explain FDI yields no significant effects. Immigration is thus not caused by FDI in the countries of birth of migrants. Furthermore, migration is not significantly correlated with FDI in-stocks. Thus, immigrants do not migrate to OECD countries in tandem with, following, or leading investment from their origins on a large scale.

Furthermore, a systematic bias might arise from past bilateral migrants who might cause current migration due to network or similar effects (Munshi, 2003). However, additional tests show that the growth of the log immigrant stock is not higher when there is a larger migrant presence, as a network effect would imply. Also, there is no evidence of significantly more bilateral immigration based on common spoken languages.

Our results hold when excluding country-pairs with common borders. The coefficient on the scarcity interaction term is (insignificantly) larger in the full sample. This suggests that, as one could expect, contiguous countries whose citizens cannot communicate easily benefit more from migrants with regard to FDI. We furthermore tested whether the scarcity interaction might reflect any other correlated factor. Table 9 reports exemplary results for several gravity and language-related variables. Neither for the distance between countries, common official languages, language proximity (a proxy of how much overlap two languages have, yielding e.g., a high value for Dutch and English), nor past colonial ties can a similar pattern to the scarcity interaction be found.

[Table 9 about here.]

The estimates are furthermore robust to excluding all zero bilateral migrant stocks, the smallest quarter of bilateral migrant stocks, or the highest quarter. 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.⁷

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

⁷For 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.

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 this context.

This conclusion receives strong support from a further 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 similar positive effect on bilateral FDI as bilateral immigrants - in terms of size as well as statistical significance. Quantitatively, we find that a 1 percent increase in either bilateral immigrants or third-country immigrants with the "right" language skills implies a 0.2 to 0.4 percent increase in the FDI out-stock, for a country pair with an average level of common language skills.

To further corroborate the key role of language skills, we check for a similar effect for immigrants in the reporting country from the FDI host country's neighbors. Geographic proximity is usually a good proxy for cultural familiarity, which might help to facilitate FDI. However we find no effect on bilateral FDI for immigrants from countries that border on the FDI host country.

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

- Aubry, A., Kugler, M., Rapoport, H., 2012. Migration, FDI and the Margins of Trade. mimeo.
URL http://econ.biu.ac.il/files/economics/seminars/amandine_aubry.pdf
- Egger, P. H., Lassmann, A., Aug. 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., Mar. 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).
URL http://mpra.ub.uni-muenchen.de/36347/2/MPRA_paper_36347.pdf
- 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.
- 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., Mar. 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 | Std. Dev. | Min | Max |
|---|-------|-------|-----------|-------|---------|
| FDI outstock in mill. USD | 6,763 | 2,341 | 11,025 | 0.00 | 257,780 |
| Exports in mill. USD | 6,763 | 1,562 | 6,702 | 0.00 | 126,682 |
| Log bilateral immigrant stock | 6,763 | 6.49 | 3.06 | 0.00 | 14.2 |
| Scarcity of bilateral language ties ($1 - \alpha_{ij}$) | 6,763 | 0.77 | 0.28 | 0.01 | 1 |
| $(1 - \alpha_{ij}) \times$ Log bilateral immigrant stock | 6,763 | 4.80 | 2.80 | 0.00 | 12.9 |
| Uniqueness of language tie | 6,723 | 0.15 | 0.24 | 0.00 | 0.99 |
| Uniqueness \times Log bilateral immigrant stock | 6,723 | 1.23 | 2.20 | 0.00 | 12.5 |
| Log GDP of reporter | 6,763 | 5.71 | 1.25 | 2.49 | 8.11 |
| Log GDP of partner | 6,763 | 10.51 | 0.46 | 9.08 | 11.1 |
| Log GDP per capita of reporter | 6,763 | 3.77 | 2.33 | -2.27 | 8.90 |
| Log GDP per capita of partner | 6,763 | 8.38 | 1.54 | 4.80 | 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 using poisson pseudo maximum likelihood (ppml)

| Estimator Dependent variable | (1) | (2) | (3) | (4) |
|---------------------------------|---------------------|---------------------|---------------------------|-------------------------------|
| | Exports | FDI | Exports | FDI |
| Lagged log(bilateral migrants) | 0.210*** (0.019) | 0.287*** (0.029) | 0.053 (-) [†] | 0.103 (0.102) [†] |
| Gravity variables | ✓ | ✓ | ✓ | ✓ |
| Reporter & Partner FE | ✓ | ✓ | | |
| Bilateral FE | | | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| R ² | 0.945 | 0.904 | 0.996 | 0.978 |

Panel B: Estimates using panel poisson with fixed effects

| Dependent variable | (5) | (6) | (7) | (8) |
|---|--------------------|------------------|------------------|------------------|
| | Exports | FDI | Exports | FDI |
| Lagged log(bilateral migrants) | 0.267** (0.105) | 0.241 (0.168) | 0.053 (0.047) | 0.103 (0.152) |
| Gravity variables | | | ✓ | ✓ |
| Reporter & Partner FE | | | | |
| Bilateral FE | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| R ² | n/a | n/a | n/a | n/a |
| Observations | 8,187 | 5,283 | 8,187 | 5,283 |
| Non-zero observations | 4,274 | 3,728 | 4,274 | 3,728 |
| Number of bilateral non-zero observations | 1,042 | 563 | 1,042 | 563 |

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. †: SEs are only successfully calculated most of the time and can be shown to be too small.

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

| Dependent variable Subsample | (1) | (2) | (3) | (4) | (5) | (6) |
|--|-------------------|--------------------|----------------------|--------------------|----------------------|------------------------|
| | Exports all | FDI | Exports < $p(75)$ | FDI t_2_import | Exports < $p(75)$ | FDI t_2_regproperty |
| Lagged log(bilateral migrants) | -0.025 (0.142) | -0.518 (0.346) | -0.135 (0.143) | -0.506 (0.352) | -0.098 (0.155) | -0.463 (0.374) |
| Lagged log(bilateral migrants) · common language scarcity | 0.105 (0.187) | 0.959** (0.428) | 0.306 (0.187) | 0.959** (0.437) | 0.165 (0.205) | 1.128** (0.467) |
| Gravity variables | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bilateral FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-zero observations | 4,083 | 2,269 | 2,926 | 1,947 | 2,977 | 1,826 |
| Number of non-zero bilateral ties | 1,042 | 563 | 769 | 477 | 791 | 467 |

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 with similar language skills*

| Dependent variable Subsample | (1) | (2) | (3) | (4) | (5) | (6) |
|--|-------------------|--------------------|----------------------|--------------------|----------------------|------------------------|
| | Exports all | FDI | Exports < $p(75)$ | FDI t_2_import | Exports < $p(75)$ | FDI t_2_regproperty |
| Lagged log(bilateral migrants) | -0.024 (0.139) | -0.523 (0.345) | -0.131 (0.141) | -0.523 (0.354) | -0.086 (0.151) | -0.474 (0.374) |
| Lagged log(bilateral migrants) · common language scarcity | 0.125 (0.185) | 0.920** (0.424) | 0.314* (0.187) | 0.902** (0.434) | 0.180 (0.202) | 1.087** (0.459) |
| Lagged log (same language migrants) | -0.087 (0.058) | 0.372* (0.210) | -0.041 (0.052) | 0.524* (0.297) | -0.092 (0.063) | 0.372* (0.220) |
| Gravity variables | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bilateral FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-zero observations | 4,083 | 2,269 | 2,926 | 1,947 | 2,977 | 1,826 |
| Number of non-zero bilateral ties | 1,042 | 563 | 769 | 477 | 791 | 467 |

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

| Dependent variable Subsample | (1) | (2) | (3) | (4) | (5) | (6) |
|--|--------------------|--------------------|----------------------|---------------------|----------------------|------------------------|
| | Exports all | FDI | Exports < $p(75)$ | FDI t_2_import | Exports < $p(75)$ | FDI t_2_regproperty |
| Lagged log(bilateral migrants) | -0.057 (0.139) | -0.433 (0.346) | -0.147 (0.142) | -0.362 (0.358) | -0.119 (0.152) | -0.392 (0.372) |
| Lagged log(bilateral migrants) · common language scarcity | 0.183 (0.184) | 0.715* (0.420) | 0.347* (0.190) | 0.494 (0.432) | 0.240 (0.200) | 0.869* (0.451) |
| Lagged log(same language migrants) · common language scarcity | -0.143* (0.076) | 0.503** (0.243) | -0.083 (0.080) | 0.980*** (0.334) | -0.161* (0.085) | 0.504** (0.238) |
| Gravity variables | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bilateral FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-zero observations | 4,083 | 2,269 | 2,926 | 1,947 | 2,977 | 1,826 |
| Number of non-zero bilateral ties | 1,042 | 563 | 769 | 477 | 791 | 467 |

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 6: *Migrants, exports, and FDI - bilateral and third-country migrants combined*

| Dependent variable Subsample | (1) | (2) | (3) | (4) | (5) | (6) |
|--|-------------------|--------------------|----------------------|--------------------|----------------------|------------------------|
| | Exports all | FDI | Exports < $p(75)$ | FDI t_2_import | Exports < $p(75)$ | FDI t_2_regproperty |
| Lagged log(bilateral + same language migrants) | -0.024 (0.148) | -0.544 (0.360) | -0.138 (0.149) | -0.532 (0.367) | -0.101 (0.161) | -0.484 (0.391) |
| Lagged log(bilateral + same language migrants) · common language scarcity | 0.104 (0.196) | 1.007** (0.442) | 0.317 (0.194) | 1.009** (0.451) | 0.167 (0.215) | 1.178** (0.479) |
| Gravity variables | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bilateral FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-zero observations | 4,083 | 2,269 | 2,926 | 1,947 | 2,977 | 1,826 |
| Number of non-zero bilateral ties | 1,042 | 563 | 769 | 477 | 791 | 467 |

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

| Dependent variable Subsample | (1) | (2) | (3) | (4) | (5) | (6) |
|--|-------------------|--------------------|-------------------|--------------------|-------------------|--------------------|
| | Exports | FDI | Exports | FDI | Exports | FDI |
| | all | | < p(75) | t_2_import | < p(75) | t_2_regproperty |
| Lagged log(bilateral migrants) | -0.025 (0.142) | -0.522 (0.347) | -0.137 (0.144) | -0.510 (0.353) | -0.094 (0.155) | -0.465 (0.375) |
| Lagged log(bilateral migrants) · common language scarcity | 0.106 (0.186) | 0.957** (0.428) | 0.302 (0.187) | 0.955** (0.436) | 0.171 (0.202) | 1.121** (0.464) |
| Lagged log (migrants from partner country's neighbors) | -0.011 (0.047) | 0.048 (0.161) | 0.030 (0.048) | 0.064 (0.205) | -0.035 (0.055) | 0.042 (0.227) |
| Gravity variables | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bilateral FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-zero observations | 4,083 | 2,269 | 2,926 | 1,947 | 2,977 | 1,826 |
| Number of non-zero bilateral ties | 1,042 | 563 | 769 | 477 | 791 | 467 |

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 8: *The role of high skilled migrants*

| Dependent variable Subsample | (1) | (2) |
|---|-------------------|-------------------------|
| | Exports | FDI |
| | all countries | with migrant skill data |
| Lagged log(low/medium skilled bilateral migrants) | -0.140 (0.101) | 0.008 (0.221) |
| Lagged log(high skilled bilateral migrants) | 0.386 (0.379) | -1.288* (0.781) |
| Lagged log(high skilled bilateral migrants) · common language scarcity | 0.258 (0.460) | 2.644** (1.111) |
| Gravity variables | ✓ | ✓ |
| Bilateral FE | ✓ | ✓ |
| Year FE | ✓ | ✓ |
| Non-zero observations | 2,829 | 1,746 |
| Number of non-zero bilateral | 713 | 420 |

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. 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". Note that the baseline category now includes countries with unskilled migrants and that the sample size dropped because of missing data on skills of bilateral migrants. By chance, these are many of the countries which belonged to either of the two excluded subgroups but not the other. Hence, the excluded subgroups now overlap.

Table 9: *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) | -0.518 (0.346) | 0.011 (0.184) | 0.130 (0.152) | 0.140 (0.166) | 0.156 (0.160) |
| Common language scarcity · lagged log(bilateral migrants) | 0.959** (0.428) | | | | |
| Distance · lagged log(bilateral migrants) | | 0.000 (0.000) | | | |
| Common official language · lagged log(bilateral migrants) | | | -0.922 (0.763) | | |
| Language proximity · lagged log(bilateral migrants) | | | | -0.017 (0.068) | |
| Colony · lagged log(bilateral migrants) | | | | | -0.807 (0.544) |
| Gravity variables | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bilateral FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-zero observations | 2,269 | 2,269 | 2,269 | 2,269 | 2,269 |
| Non-zero bilateral ties | 563 | 563 | 563 | 563 | 563 |

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.