Language and Cultural Barriers in International Factor Movements*

Sebastian Otten[1,2]

[1] *Ruhr University Bochum*
[2] *RWI Essen*

This version: October 2013

PRELIMINARY DRAFT – PLEASE DO NOT QUOTE

Abstract

This paper analyzes the effect of language and cultural barriers on different types of international factor movements, i.e., international trade flows, cross-holdings of assets, and consolidated international banking claims. In addition to disentangling the impact of language and cultural dissimilarities I analyze the effect of English proficiency on international factor movements, an impact factor so far neglected in the literature. Using a gravity model the results show that linguistic and genetic distance have varying effects on the examined factor movements. While controlling for a host of other possible determinants, I find strong evidence that a higher linguistic distance between two countries reduces cross-border trade and investment holdings between these countries. The results for genetic distance, however, are mixed. While genetic differences significantly reduce bilateral trade, they have no effect on international investments. These findings are in line with the theoretical expectations and provide supportive evidence that language differences contribute to higher informational frictions across countries, thereby reducing bilateral trade and cross-border investment flows, respectively.

JEL Classifications: F14, F21, Z13

Keywords: trade, investment, genetic, language, gravity models

*The author is grateful to Thomas K. Bauer, Julia Bredtmann, Ingo Isphording and Christoph M. Schmidt for helpful comments and suggestions. I am also very thankful to Andrew K. Rose for providing some of the gravity data. All remaining errors are my own. – All correspondence to: Sebastian Otten, Ruhr University Bochum, 44780 Bochum, Germany, Email: sebastian.otten@rwi-essen.de.

1 Introduction

In a number of studies, economists have shown the importance of trust and information in overcoming informal barriers to international factor movements. For example, Guiso et al. (2009) find that lower bilateral trust leads to less trade between countries and less cross-border portfolio investments. Informal barriers may arise because of weak international legal institutions or inadequate information of opportunities for advantageous economic exchange (Rauch and Trindade, 2002). In particular, crossing cultural and language boundaries, as is often the case in international factor movements, reduces the level of trust between economic agents and makes obtaining information more costly by raising hurdles in communication.

Both economic theory and empirical evidence suggest that sharing an official or speaking a common language provide an important stimulus in international economic exchange. From international trade, for instance, it is known that having a common language increases trade flows by about 40%.[1] However, previous empirical literature leaves open the question whether differences between countries' native languages affect international factor mobility above and beyond the effect of sharing a common language. Furthermore, the possibility that economic agents might use a non-native language, a lingua franca, to communicate is disregarded in most empirical work.

Language and cultural barriers are also possible candidates to illuminate the "home bias puzzle" in cross-border portfolio investment. Unlike trade in goods, assets are "weightless". Therefore, there are no physical transportation costs, which might explain investors' preferences for their own country. Moreover, standard finance theory based on portfolio choice models suggests diversifying investment risks by holding a variety of assets from countries whose business cycles are uncorrelated with each other (Portes and Rey, 2005). The international financial integration over the last thirty years, which was fostered by, e.g., capital account liberalizations, electronic trading, increasing exchanges of information across borders, and falling transaction costs, should reduce the disproportionate high share of national assets in portfolios toward a higher share of foreign assets (Coeurdacier and Rey, 2013; Lane and Milesi-Ferretti, 2003). However, despite better financial integration, the observable market segmentation in country portfolios contradicts the assumed view of frictionless global financial markets. Rather, it points to the impact of informal barriers such as information asymmetries and cultural differences in international capital flows (Portes and Rey, 2005).[2] Knowledge about these barriers is important in understanding the

[1] In a meta-analysis based on 701 language effects collected from 81 academic articles, Egger and Lassmann (2012) find that a common language increases bilateral trade flows by 44% on average.

[2] This view is supported by Coeurdacier and Rey (2013), who state that in 2007, US investors still held more than 80% of the US equities compared to 92% in 1989 documented by French and Poterba (1991). In both cases, this is a much higher proportion than the share of US equities in the world market portfolio. For literature surveys on the phenomenon of home bias, see Lewis (1999) and more recently

"home bias puzzle" in investment decisions. However, empirical evidence on the effect of language and cultural differences on investors' preferences to invest in their home country (or in this regard close countries) is scarce.

In this paper, I evaluate the extent to which language and cultural barriers affect different types of international factor movements, i.e., international trade flows, cross-holdings of assets, and consolidated international banking claims by addressing three major issues. First, I investigate the effect of differences between countries' native languages on the three types of bilateral factor movements using a novel measure of linguistic distance. This measure allows for directly testing whether and to what extend dissimilarities between countries' native languages affect international factor mobility. Next, I examine the impact of cultural differences on bilateral economic exchange by using a measure of genetic distance as proposed and applied by Guiso et al. (2009). My particular interest lies in a comparison of the effects of genetic and linguistic distance, which might vary substantially over the aforementioned factor movements. Finally, I follow the idea that economic agents might use English, the lingua franca in international business, for communication. In doing so, I examine how the results of differences between countries' native languages change when including a measure for countries' linguistic distance toward English in the empirical models.

The impact of language and cultural differences is likely to be especially important in transactions where individuals have to coordinate with each other. This holds true for less transparent and standardized transactions in markets with non-trustworthy institutions and a difficult bureaucratic and commercial environment. Cross-border portfolio investments and consolidated international banking claims are relatively standardized transactions, generally taking place between financial intermediaries, e.g., investment banks, pension funds, mutual funds, etc. In contrast to these transactions, international trades of goods are less organized exchanges between mostly heterogeneous firms. Differences between native languages or insufficient proficiencies in English, the lingua franca in international business, induce communication barriers affecting both bilateral trade as well as international investment flows. Furthermore, a greater cultural distance could decrease the likelihood of finding an adequate trading partner due to larger differences in the supplied and demanded products. The opposite effect, however, might exist for international investments. A greater cultural distance between two countries should be associated with a lower correlation of their stock markets and thereby increase the diversification potential of the respective assets for the portfolios of foreign investors.

To test these hypotheses, the empirical analysis is based on three different data sets. The first data set comprises international bilateral trade flows from 1950 to 2006 between 180 exporter countries and 187 importer countries. The second database covers cross-

Coeurdacier and Rey (2013).

holdings of asset stocks between 62 source and 173 host countries for the years 2001 till 2003. To check the sensitivity of this relative small sample, I further use data of consolidated international banking claims covering 24 source and 182 host countries between 1983 and 2006.

In the empirical analysis, I apply a gravity model as first proposed by Tinbergen (1962) and since then applied in numerous empirical studies on factor mobility.[3] These models explain the intensity of bilateral relations by some elements of mass, e.g., economic size (often approximated by GDP) and geographic distance. I follow this approach, but measure distance not only in the geographical dimension, but further incorporate measures for cultural and language dissimilarities as described above.

This study attempts to quantify the effect of informal barriers on international factor movements by focusing on language and cultural barriers. It directly relates to two strands in the recent empirical literature which seek to explain the determinants of bilateral trade and international portfolio investment, respectively.

The first strand emphasizes the importance of language and cultural barriers in international trade. To incorporate language-related barriers into a gravity model, common empirical practice is to use an indicator variable that covers countries which share the same official language (see, e.g., Anderson and van Wincoop, 2004). While most studies follow this approach, Melitz (2008) goes beyond the consideration of official languages and develops two different measures. The first measure reflects the probability that two randomly chosen individuals from either country share a common language. The second measure indicates whether the same language is spoken by at least 20% of the populations in both countries. These measures share the shortcoming that they only consider countries that share the same language, but do not account for the dissimilarity between the countries' languages. To the best of my knowledge, there are only four studies that take differences between native languages into account. Hutchinson (2005) relies on the measure by Chiswick and Miller (1999) and his approach is therefore restricted to distances toward English. In contrast, Lohmann (2011) uses an index of shared linguistic features within language pairs for about 200 countries. Two recent studies by Isphording and Otten (2013) and Mélitz and Toubal (2012) directly focus on the differences between native languages using the same measure of linguistic distance applied in this paper.

Certainly, a range of dominant regional languages (e.g., English in the Western countries, Russian in Eastern Europe, French in Africa, and Spanish in Latin America) play a major role in international trade (Isphording and Otten, 2013). Especially, proficiency in English as the leading language in international business is essential for most trade relations. However, the only study I am aware of that addresses issues of the ability to communicate

[3]For recent applications in international trade, see, e.g., Melitz (2008), Head et al. (2010) and Aviat and Coeurdacier (2007) for an application in bilateral asset holdings. An overview of the estimation and interpretation of gravity models can be found in Head and Mayer (2013).

in English is the one by Ku and Zussman (2010). To measure the level of English proficiency in more than 100 countries, they use a country's mean score in the "Test of English as a Foreign Language" (TOEFL).

This paper also relates to a line of research that studies the role of cultural differences in trade flows. Felbermayr and Toubal (2010) draw on bilateral score data from the Eurovision Song Contest, a popular pan-European television show, to construct a measure of cultural proximity and employ this measure into the framework of a gravity model of trade. The studies by Guiso et al. (2009) and Giuliano et al. (2013) use a measure of genetic distance to reveal the effect of cultural differences on trade. Specifically, Guiso et al. (2009) examine the relationship between bilateral trust and bilateral trade, using genetic distance as an instrument for trust. In contrast, Giuliano et al. (2013) directly analyze the effect of genetic distance on trade. They find that the negative effect of genetic distance on bilateral trade flows decreases substantially, or even disappears, once they control for further geographical barriers other than the simple geographic distance between capital cities.

The second strand in the literature emphasizes the role of language and cultural barriers in cross-border portfolio investment. One set of studies focuses on the impact of information costs raising informational asymmetries and transaction costs between domestic and foreign investors (Ahearne et al., 2004; Daude and Fratzscher, 2008; Portes and Rey, 2005, amongst others). In this regard, e.g., Tesar and Werner (1995), Grinblatt and Keloharju (2001), Chan et al. (2005) and more recently Lane and Milesi-Ferretti (2008), Beugelsdijk and Frijns (2010), and Aggarwal et al. (2012) explicitly underline the influence of language barriers on investment decisions. For example, Grinblatt and Keloharju (2001) find that investors in Finland are more likely to trade stocks of firms that share the investor's language and cultural background. To incorporate language-related barriers, these studies again use a variable that indicates whether two countries or investors share the same official language. As mentioned above, this approach has the shortcoming that the estimated effect only relies on the difference of sharing or not sharing a common language, but does not account for the dissimilarity between the languages. The only study that takes differences between languages into account is the one by Guiso et al. (2009), who use a measure for linguistic common roots created by Fearon and Laitin (2003).[4] A second set of studies examines the relationship between "familiarity effects", i.e., cultural proximity and cross-border investments (e.g., Aggarwal et al., 2012; Beugelsdijk and Frijns, 2010; Chan et al., 2005; Grinblatt and Keloharju, 2001). The only study that uses genetic distances to uncover the effect of cultural differences on cross-border portfolio investment is again the one by Guiso et al. (2009).

[4] For an overview of the approaches employed in the economic literature to measure the dissimilarity between languages, see Isphording and Otten (2011).

This paper makes three important contributions to the existing literature. First, I provide some first evidence on the effect of language and cultural differences on international trade flows. Using a novel measure of linguistic distance and incorporating genetic distance as a proxy for cultural differences, this is the first study that allows disentangling the impact of language and cultural dissimilarities on trade flows. Given the established correlation between geographic distance and language and cultural differences, the paper further sheds light on how an appropriate modeling of linguistic and genetic distance helps to solve the "distance puzzle" in international trade. Second, I extend the finance literature by providing first evidence on the effect of language and cultural dissimilarities on cross-border portfolio investment and international banking claims. Finally, this is the first comprehensive analysis of the effect of English proficiency on international factor movements, an impact factor so far neglected in the literature. The only notably exception I am aware of is the study by Ku and Zussman (2010). In contrast to their investigation, however, I use a measure of linguistic distance to model English proficiency. Thereby, I am able to extend the underlying sample to almost all of the world's countries. Furthermore, employing linguistic distance instead of TOEFL scores has the advantage of using an unbiased proxy for English proficiency. In addition, I extend their analysis by further examining financial flows, i.e., cross-holdings of asset stocks and international banking claims, providing first evidence in this line of research.

The results show that linguistic and genetic distance have varying effects on the examined factor movements. While controlling for a host of other possible determinants, I find strong evidence that a higher linguistic distance between two countries reduces cross-border trade and investment holdings between these countries. The size of this effect is substantial. For example, an increase in linguistic distance by one standard deviation decreases bilateral exports by 7.4%. This finding suggests that the further away a country's language is from the rest of the world's languages, the higher the information costs for trading partners and foreign investors and thus the smaller bilateral factor flows. In contrast to these findings, the results for genetic distance are mixed. Cultural differences have significant effects only on bilateral trade, while no significant effects are found for international investments. In particular, the results show that a greater genetic distance between two countries reduces the volume of bilateral trade and, if anything, increases the level of cross-border investments. When including the country's linguistic distance toward English in the models, the estimates indicate a significant negative impact of a higher linguistic distance toward English on international factor mobility.

Taken together, these findings are fully in line with the theoretical arguments outlined above and provide supportive evidence that language differences contribute to higher informational frictions across countries, thereby reducing bilateral trade and cross-border investment flows, respectively. As hypothesized above, the results for cultural differences are asymmetric for international trade and investment holdings.

The remainder of the paper is structured as follows. The next section briefly outlines the theoretical framework and specifies a gravity model including country-year fixed effects. Moreover, I formalize the identification strategy underlying the estimations of English proficiency on international factor mobility. Section 3 describes the bilateral trade and financial data, the measures of linguistic and genetic distance, and the extensive set of control variables. The detailed specifications used in the estimations of the different models and the respective empirical results are presented and discussed in Section 4. The concluding section summarizes and interprets the main results.

2 Empirical Model

2.1 Theoretical Background: The Structural Gravity Model

In estimating the effects of language and cultural barriers on international factor mobility, I follow common empirical practice and build on the gravity model of international trade, which has become the standard model for estimating the determinants of bilateral factor flows.[5] The gravity model can be derived from a number of theoretical frameworks, relying on different modeling assumptions (see Anderson and van Wincoop, 2003, 2004; Eaton and Kortum, 2002). The first theoretical foundations of the gravity model have been laid by Anderson (1979) and Bergstrand (1985).

The following equations build on this work and are in particular based on the canonical model by Anderson and van Wincoop (2003, 2004). In this setting, bilateral trade is given by

$$X_{ij}^k = \frac{E_j^k \, Y_i^k}{Y^k} \left(\frac{t_{ij}^k}{P_j^k \, \Pi_i^k} \right)^{1-\sigma_k} , \tag{1}$$

where X_{ij}^k is defined as nominal exports from country i into country j in product class k. E_j^k denotes the expenditure in country j for product class k, Y_i^k is the value of production in country i for product class k, and Y^k is world output in sector k. t_{ij}^k are the trade related costs between i and j, the parameter σ_k is the elasticity of substitution among brands, and P_j^k and Π_i^k denote inward and outward multilateral resistance, respectively, which are functions of trade barriers and the entire set of $\left(E_j^k, \, Y_i^k\right)$.[6] The main contribution of the

[5] Gravity equations have been applied not only to estimate the determinants of trade flows, see Anderson (2011) and Bergstrand and Egger (2011) for recent surveys, but also to migration, equity, and FDI flows. For a recent study on migration flows, see, e.g., Adsera and Pytlikova (2012) and Martin and Rey (2004) for the first application to international portfolio investment. Brenton et al. (1999) apply the gravity equation to FDI flows.

[6] See Anderson and van Wincoop (2004) for the derivation and a general discussion of this equation. Anderson and Yotov (2010b) refer to the second ratio of the right-hand side product as a measure of

"structural" gravity equation is that bilateral trade is determined by relative trade barriers, namely the bilateral trade barrier t_{ij}^k divided by multilateral resistance, which depend on the average trade barriers the exporter and importer face with all their trading partners (Anderson and van Wincoop, 2004).

Assuming (i) aggregation to one-sector economies[7], thereby omitting the superscripts k, (ii) expenditures to be equal to the value of production ($E_i = Y_i$), and (iii) symmetric bilateral trade costs ($t_{ij} = t_{ji}$), from which follows $P_i = \Pi_i$, Eq. (1) can be rewritten as:

$$X_{ijt} = \frac{Y_{it}\, Y_{jt}}{Y_t^W} \left(\frac{t_{ijt}}{P_{jt}\, \Pi_{it}} \right)^{1-\sigma} . \tag{2}$$

Since most of the recent gravity papers rely on panel data, these studies argue that the cross-section equation derived from Anderson and van Wincoop holds in all time units of the panel (Hornok, 2012). I follow this literature and denote time-varying components in Eq. (2) with the subscript t. Let Y_{it}, Y_{jt} represent nominal GDP in country i and j in year t and let Y_t^W define world GDP in year t.

The final step in the derivation of the "structural" gravity equation is to model the trade costs t_{ijt}. Here, Anderson and van Wincoop follow the literature and assume that trade costs are a log-linear function of a set of observable proxies for various types of trade costs along with some direct measurable components of trade barriers:

$$t_{ijt} = \prod_{m=1}^{M} \left(z_{ijt}^m \right)^{\gamma_m} . \tag{3}$$

Common examples for z_{ijt}^m from the literature are geographic distance between the capitals of country i and j, adjacency, and membership in a customs or monetary union. Substituting Eq. (3) into Eq. (2) and log-linearizing the resulting equation yields

$$ln\, X_{ijt} = ln\, Y_{it} + ln\, Y_{jt} - ln\, Y_t^W + \sum_{m=1}^{M} \lambda_m\, ln \left(z_{ijt}^m \right) - (1-\sigma) \left[ln\left(\Pi_{it}\right) + ln\left(P_{jt}\right) \right] , \tag{4}$$

where $\lambda_m = (1-\sigma)\, \gamma_m$. X_{ijt}, Y_{it}, Y_{jt}, and Y_t^W are observable, whereas the multilateral resistance terms are unobservable.

The "structural" gravity equation can be estimated using different approaches. Anderson and van Wincoop (2003) employ nonlinear least squares (NLS) after solving for the multilateral resistance terms, which leads to unbiased estimators if the error term is uncorrelated with the regressors. However, most of the subsequent studies control for the

"constructed home bias", which is a useful quantification with respect to the exploration of possible explanations of the "home bias puzzle" in research on international portfolio investment.

[7] Anderson and Yotov (2010a,b) show that the bias resulting from aggregation is large. However, for the sake of simplicity, I follow the conventional approach and use data on aggregated goods.

unobservable multilateral resistance terms and the observable production variables using country-specific fixed effects and apply ordinary least squares (OLS) (see, e.g., Anderson, 2011; Eaton and Kortum, 2002). The analysis here uses mainly directional (i.e., source and destination) country-year fixed effects to control not only for the multilateral resistance terms, but also for all country-specific time-varying factors. The gravity equation to be estimated becomes:

$$ln\, X_{ijt} = \boldsymbol{Z}^{'}_{ijt}\, \boldsymbol{\beta} + \sum_{i=2}^{I} \delta_{it}\, c^{Ex}_{it} + \sum_{j=2}^{J} \theta_{jt}\, c^{Im}_{jt} + \epsilon_{ijt}\;, \qquad (5)$$

where $\sum \delta_{it} c^{Ex}_{it}$ and $\sum \theta_{jt} c^{Im}_{jt}$ denote full sets of exporter-year and importer-year dummy variables, respectively, while $\boldsymbol{Z}_{ijt}$ is a vector of observable proxy variables for bilateral trade costs. The error term ϵ_{ijt} consists of two components. One component is a random error, which simply reflects measurement error associated with export flows. The second component, however, comprises all unobservable bilateral factors that are correlated with exports between country i and j and are not appropriately controlled for in the estimation.[8] In case these factors are correlated with the included regressors, the estimates will be biased due to omitted variables.[9]

2.2 Empirical Strategy

In the following analysis, different types of factor movements are examined by estimating the gravity model in Eq. (5) using three different dependent variables. A first set of equations is estimated employing the logarithm of bilateral annual exports. Following previous literature, export volume instead of total trade volume is used, because the direction of trade volumes matters for the parameter estimation. A second set of models is estimated by using the logarithm of cross-holdings of asset stocks as the dependent variable, while I investigate the logarithm of consolidated international banking claims in a third set of models.

The derived gravity model in Eq. (5), including the full sets of directional country-year fixed effects, can be estimated using OLS.[10] Due to the panel nature of the data, leading to

[8]To avoid bias accruing from the correlation between included determinants of bilateral trade and the unobservable bilateral components of the error term, Baltagi et al. (2003) and Baldwin and Taglioni (2006) suggest the additional inclusion of time-invariant country-pair fixed effects in the estimation model. The downside of this approach, however, is that time-invariant parameters would no longer be identifiable as the fixed effects net out all time-constant variation. Since the regressors of interest in this analysis are time-invariant, I do not follow this approach.

[9]Another source of bias leading to non-orthogonality of the error term are endogenous regressors resulting from reverse causality. Since the causal direction of the variables of interest in the following analysis is considered as unambiguous, issues regarding this source of endogeneity are not further discussed. For an overview of how the literature has addressed these problems, see Anderson and van Wincoop (2004).

[10]The standard OLS procedure has been criticized in the recent literature, because of dropping country-

a repeated presence of country-pairs, all bilateral observations of country i and country j may have common disturbance elements. For this reason, the standard errors are adjusted to be robust to clustering of residuals by country-pairs.[11]

To examine the impact of language and cultural barriers on international factor movements, the analysis proceeds in three steps. While the first two steps rely on the gravity model derived in Eq. (5), country-pair specific random effects instead of the two sets of country-year fixed effects are applied in the third step.

Identifying Language and Cultural Barriers

In the first step, interest is solely directed toward identifying the effect of bilateral language and cultural differences on international factor mobility. As the analysis of these factors lies outside of the standard gravity model, Eq. (3) is augmented by variables measuring bilateral language and cultural differences. Since it is difficult to incorporate these factors in an estimation model, I use observable measures for linguistic and genetic distance, respectively, as proxy variables. The bilateral trade costs[12] are then given by

$$t_{ijt} = exp\left[\gamma\, LD_{ij} + \kappa\, GD_{ij} + \boldsymbol{Z}'_{ijt}\, \boldsymbol{\beta}\right] \,. \tag{6}$$

The set of dyadic variables in Eq. (6) can be divided into two groups. The first group includes the variables of primary interest, LD_{ij} and GD_{ij}, where the former denotes the linguistic and the latter the genetic distance between country i and j. The second group, represented by the vector $\boldsymbol{Z}_{ijt}$, comprises a set of observable control variables typically used in gravity regressions. The controls include (unless otherwise specified) the logarithm of the geographic distance between i and j, an indicator variable that equals one if two countries share a border, an indicator that equals one if both countries share the same legal origin, and an indicator that equals one if two countries were once, or are still, in a colonial relationship. Furthermore, two indicator variables are added, where the first equals one if both countries belong to a common regional trade arrangement (RTA) in year t and the second equals one if both countries are GATT/WTO participants in year t.

pairs with zero trade volumes (see, e.g., Helpman et al., 2008; Santos Silva and Tenreyro, 2006). There are a number of possible methods to incorporate zeros in the gravity model. For an overview, see Anderson (2011) and Head and Mayer (2013). However, I do not use these methods in the following analysis. Instead, I follow the conventional method described in Section 2.1 and take the natural logarithm of annual exports as the dependent variable, thereby dropping observations where exports are recorded as zero. To verify that the results obtained in the analysis are robust to this specification, I conduct a first series of robustness checks in which $ln\,(X_{ijt} + 1)$ serves as dependent variable. In a second series of robustness checks, the models are estimated using Poisson Pseudo-Maximum Likelihood (PPMLE) instead of OLS (see, e.g., Santos Silva and Tenreyro, 2006). The results are available upon request.

[11] To check the robustness of the results, I also reestimate all models using two-way clustering by country and year employing the method of Cameron et al. (2011). The results are available upon request.

[12] In conformity with the standard gravity model, the term "trade costs" is used to refer to the costs of bilateral factor movements, although different factor movements are considered (i.e., trade, portfolio investments, and banking claims).

The final control variable included in the vector $\boldsymbol{Z}_{ijt}$ is an indicator that equals one if the two countries share a common currency.

The essential idea is that the variation in the country-pair dimension of the bilateral linguistic and genetic distance identifies the effect of language and cultural dissimilarities on factor mobility. However, simultaneity between the variables of interest and the dependent variable or omitted variables that are correlated with both the dependent variable and one of the variables of interest would cause endogeneity and bias the parameter estimates. Since linguistic and genetic distance are used as proxies for language and cultural differences, reverse causality is not considered as a source of endogeneity, as it is unlikely that these measures are influenced by international factor movements.[13]

The main concern in this analysis is the possibility of omitted variables. However, to bias the estimates, a possible confounder has to be correlated with the measure of linguistic or genetic distance. I argue that the existence of such a variable is unlikely.[14] Nonetheless, to address this possible bias, the empirical strategy is to include a number of variables in Eq. (6) that affect factor mobility, thereby controlling for as many theoretical causes of bilateral factor movements as possible. Since the variables of interest are time-invariant, it is not possible to include country-pair fixed effects in Eq. (6).

The primary interest is how an increase in the linguistic and genetic distance affects bilateral factor mobility. Therefore, the baseline specifications are as follows: First, I estimate Eq. (5) using Eq. (6) without the genetic distance variable GD_{ij} to analyze the effect of linguistic distance on bilateral factor movements. Second, I estimate the same specification, but use genetic distance rather than linguistic distance as a regressor. Finally, a horse race between linguistic and genetic distance is conducted by including both variables in Eq. (5).

Identifying the Effect of English Proficiency

In the second and third step, the analysis turns to the investigation of the impact of English proficiency on international factor mobility. To analyze this effect, Eq. (6) has to be augmented by variables that capture English proficiency. Since it is difficult to measure a country's average English skills, each country's linguistic distance toward English (henceforth LDE) is used as a proxy for the average English ability of the population living in the respective country.[15] However, in the presence of directional country-year fixed effects, as in Eq. (5), regressors that vary in the same dimensions as these fixed effects

[13]For a description of these measures and the underlying data, see Section 3.1.

[14]Spolaore and Wacziarg (2009, 2012) argue in the same direction by pointing out that genetic distance acts as an exogenous regressor in their models of development diffusion and technology adoption, respectively.

[15]Isphording and Otten (2011) provide empirical evidence that the linguistic distance between individuals' mother tongue and their second language is a strong predictor of their language proficiency.

cannot be identified, since the fixed effects net out all variation from the data except for the bilateral variation. This precludes, for example, the estimation of separate exporter and importer country-specific effects. A solution to get around this problem is to create a new variable that varies bilaterally and include it in the estimation. One possibility to transform two monadic variables describing country i and j, respectively, into a dyadic variable is to simply multiply both variables, e.g., one can create a dyadic area variable by multiplying the area of country i times the area of country j. The created bilateral-varying variable is identifiable even when directional country fixed effects are included in the estimation model. However, the identification solely rests on functional form assumptions and is therefore a sort of constructed identification. A second possibility is to use the logarithm of the average, sum, or difference of country-specific variables. In these cases the created bilateral variable is identified because of the change in the functional form. Without taking the logarithm, the new variable is only a linear transformation of the monadic variables and cannot be identified (Head and Mayer, 2013). Since the identification assumptions that have to be made to generate the dyadic variables are strong, there are reasons to be cautious in accepting the parameter estimates based on these methods. In the next steps of the analysis, I therefore do not use the described approaches to examine the effect of a country's LDE.

In the second step, I investigate the effect of a country's LDE in relation to the bilateral linguistic distance on international factor movements. In order to do so, notice that there are two types of countries: (i) countries with a higher LDE than the respective bilateral linguistic distance and (ii) countries with a lower LDE than the respective bilateral linguistic distance. In the following, two different identification assumptions are considered to construct separate indicator variables. Formally, the first indicator variable is defined as:

$$LDE_{ij}^{(1)} = \begin{cases} 1 & \text{if } (LDE_i > LD_{ij}) \wedge (LDE_j > LD_{ij}) \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where LDE_i and LDE_j denote the linguistic distance toward English of country i and j, respectively. Under the first identification assumption, the indicator variable captures the effect of both trading partners facing a higher LDE than their bilateral linguistic distance. The second indicator variable is defined as below:

$$LDE_{ij}^{(2)} = \begin{cases} 1 & \text{if } (LDE_i > LD_{ij}) \vee (LDE_j > LD_{ij}) \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where $LDE_{ij}^{(2)}$ captures the effect of at least one trading partner facing a higher LDE than their bilateral linguistic distance.

The identification assumptions above determine the exact formulation of the indicator variables. However, it is reasonable to assume that bilateral factor mobility is unaffected by the LDE if both countries in a pair share a common language, i.e., if the bilateral linguistic distance is zero. Therefore, both indicator variables are multiplied with a trigger variable denoting one if two countries in a pair do not share a common language and zero otherwise.

The aim of using these indicator variables is to estimate the effect of a high LDE in relation to the bilateral linguistic distance on bilateral factor mobility while controlling for the bilateral linguistic distance and each country's LDE by employing directional country fixed effects. The essential idea is that if the LDE is higher than the bilateral linguistic distance, the importance of the bilateral linguistic distance in the communication process increases, because the hurdles in the acquisition of English language skills are higher compared to the hurdles in the acquisition of the trading partner's native language. Therefore, the probability that the trading partners communicate in one of the native languages increases.

The estimation of specifications including these indicator variables implicitly assume that a country's LDE and the bilateral linguistic distance affect bilateral factor movements above and beyond the inclusion of the direct effects. To test these hypotheses, Eq. (6) is extended by separately adding an interaction term of the trigger variable and each indicator variable. The resulting specifications are as follows: First, I estimate Eq. (5) using Eq. (6) including the interaction term $TR_{ij} \times LDE_{ij}^{(1)}$. Second, I estimate the same specification, but use $TR_{ij} \times LDE_{ij}^{(2)}$ as a regressor.

Identifiability and the Random Effects Estimator

In the third step of the analysis, I attempt to explore the direct effect of the LDE on international factor movements. Considering the gravity model in Eq. (5), which includes the full sets of directional country-year fixed effects, reliable estimates for the effect of a country's LDE on international factor mobility are not estimable. For instance, it is not possible to estimate separate country-specific effects without relying on strong identification assumptions, because the fixed effects net out all variation from the data except for the bilateral variation. In the following, a random effects approach is therefore presented as an alternative method to account for the unobservable multilateral resistance terms in the gravity equation (see Eq. (4)).

The focus of this analysis step is on estimating the effects of both the country-specific LDE and the bilateral linguistic distance on international factor mobility. To identify the impact of the country-specific LDE, Eq. (6) is augmented by the LDE of country i, LDE_i, and the LDE of country j, LDE_j. The trade cost function is then given by:

$$t_{ijt} = exp\left[\eta\, LDE_i + \varphi\, LDE_j + \gamma\, LD_{ij} + \kappa\, GD_{ij} + \boldsymbol{Z}^{'}_{ijt}\,\boldsymbol{\beta}\right] \ . \tag{9}$$

Substituting Eq. (9) into the "structural" gravity equation (Eq. (2)) and log-linearizing the resulting equation yields:

$$\begin{aligned} ln\, X_{ijt} = {} & ln\, Y_{it} + ln\, Y_{jt} - ln\, Y_t^W + \eta\, LDE_i + \varphi\, LDE_j + \gamma\, LD_{ij} \\ & + \kappa\, GD_{ij} + \boldsymbol{Z}^{'}_{ijt}\,\boldsymbol{\beta} - (1-\sigma)\left[ln\,(\Pi_{it}) + ln\,(P_{jt})\right] \ . \end{aligned} \tag{10}$$

To estimate the "structural" gravity equation using random effects (GLS), the population of country i and j in year t is added to Eq. (10) to control for further country-specific characteristics. World GDP and other common time effects are captured by a full set of year dummy variables, which are also added to Eq. (10). The gravity equation to be estimated becomes:

$$\begin{aligned} ln\, X_{ijt} = {} & \phi\, ln\, POP_{it} + \nu\, ln\, POP_{jt} + \vartheta\, ln\, Y_{it} + \rho\, ln\, Y_{jt} + \eta\, LDE_i + \varphi\, LDE_j \\ & + \gamma\, LD_{ij} + \kappa\, GD_{ij} + \boldsymbol{Z}^{'}_{ijt}\,\boldsymbol{\beta} + \sum_{t=2}^{T} \tau_t\, T_t + \zeta_{ij} + \epsilon_{ijt} \ , \end{aligned} \tag{11}$$

where Y_{it} and Y_{jt} represent GDP per capita in country i and j in year t, and ζ_{ij} denotes a directional country-pair specific random effect, assuming $E\,(\zeta_{ij}) = 0$.[16]

Under the identification assumption that the error term ϵ_{ijt} and the bilateral random effect ζ_{ij} are uncorrelated with the variables of interest, the random effects estimator delivers unbiased parameter estimates. The main concern in this regard is again the possibility of omitted variables. However, as discussed above, to bias the estimates, a possible confounder has to be correlated with the measure of linguistic or genetic distance. I argue that the existence of such a variable is unlikely. Furthermore, the random effects estimator has two advantages over the fixed effects estimator in this analysis. First, it allows identifying the effect of variables that only vary in the same dimension as the bilateral fixed effects and are as such not identifiable in the fixed effects model. Second, the estimates do not suffer from potential over-specification due to the high number of fixed effects, which implies that much less data information will be used for the estimation of the parameters of interest (Matyas et al., 2013).

[16] A country-pair random effect is used, since this is a more general model compared to the inclusion of separate random effects for each country i and j.

3 Data

3.1 Measuring Linguistic and Genetic Distance

As the data on linguistic and genetic distance are not commonly used in the economic literature, they are in the following described in some detail. Linguistic distance is the dissimilarity of languages in a multitude of dimensions, such as vocabulary, grammar, pronunciation, scripture, and phonetic inventories. This multidimensionality of linguistic distance makes it difficult to find an appropriate empirical operationalization to be used in applied economic studies. Isphording and Otten (2011) give an overview of the different approaches used in the economic literature to measure the linguistic distance between two languages and compare their advantages and disadvantages. In this analysis, the so-called Levenshtein distance is used to measure linguistic distance.

The Levenshtein Distance

As described by Isphording and Otten (2011, 2013), it is possible to derive an operationalization of linguistic distance without strong identification assumptions that underlie previous approaches. The so-called *Automatic Similarity Judgment Program* (ASJP) developed by the German Max Planck Institute for Evolutionary Anthropology[17] aims at developing an automatic procedure to evaluate the phonetic similarity between all of the world's languages and offers a convenient way of deriving a continuous measure of linguistic differences that is purely descriptive in nature. The basic idea is the automatic comparison of the pronunciation of words having the same meaning across languages. The average similarity across a specific set of words is then taken as a measure for the linguistic distance between the languages (Bakker et al., 2009).

This distance can be interpreted as an approximation of the number of cognates between languages. The linguistic term cognates denotes common ancestries of words. A higher number of cognates indicates closer common ancestries. Although restricting its computation to differences in pronunciation, a lower Levenshtein distance therefore also indicates a higher probability of sharing other language characteristics such as grammar (see Serva, 2011). The language acquisition of second language learners is crucially affected by such differences in pronunciation and phonetic inventories, as these determine the difficulty in discriminating between different words and sounds.[18]

The algorithm calculating the distance between words relies on a specific phonetic alphabet, the ASJPcode. The ASJPcode uses the characters within the standard ASCII[19]

[17]Further information can be found at http://www.eva.mpg.de.

[18]For a recent overview of the linguistic literature on language background and language acquisition, see Llach (2010).

[19]American Standard Code for Information Interchange, keyboard-character-encoding scheme.

alphabet to represent common sounds of human communication. The ASJPcode consists of 41 different symbols representing 7 vowels and 34 consonants. Words are then analyzed as to how many sounds have to be substituted, added, or removed to transfer the one word in one language into the same word in a different language (Holman et al., 2011). The words used in this approach are taken from the so-called 40-item Swadesh list, a list including 40 words that are common in nearly all the world's languages, including parts of the human body or expressions for common things of the environment. The Swadesh list is deductively derived by Swadesh (1952), its items are believed to be universally and culture independently included in all world's languages.[20]

The ASJP program judges each word pair across languages according to their similarity in pronunciation. For example, to transfer the phonetic transcription of the English word *you*, transcribed as *yu*, into the transcription of the respective German word *du*, one simply has to substitute the first consonant. But to transfer *maunt3n*, which is the transcription of *mountain*, into *bErk*, which is the transcription of the German *Berg*, one has to remove or substitute each of the 7 consonants and vowels, respectively.

The following formalization of the computation follows Petroni and Serva (2010). To normalize the distance according to the word length, the resulting number of changes is divided by the word length of the longer word. Denoting this normalized distance between item i of language α and β as $D_i(\alpha, \beta)$, the calculation of the normalized linguistic distance (LDN) is computed as the average across all $i = 1, ..., M$ distances between synonyms of the same item:

$$LDN(\alpha, \beta) = \frac{1}{M} \sum_i D(\alpha_i, \beta_i). \tag{12}$$

To additionally account for potential similarities in phonetic inventories which might lead to a similarity by chance, a global distance between languages is defined as the average Levenshtein distance of words with different meanings:

$$\Gamma(\alpha, \beta) = \frac{1}{M(M-1)} \sum_{i \neq j} D(\alpha_i, \beta_j). \tag{13}$$

The final normalized and divided linguistic distance (LDND) is then defined as the quotient of the normalized linguistic distance and the global distance between α and β:

$$LDND(\alpha, \beta) = \frac{LDN(\alpha, \beta)}{\Gamma(\alpha, \beta)}. \tag{14}$$

The resulting continuous measure can broadly be interpreted as a percentage measure of dissimilarity between languages, with lower numbers indicating a closer relation. In a few cases, the resulting numbers are bigger than 100%, indicating a dissimilarity that

[20] A list of the 40 words is available upon request.

exceeds a potentially incidental similarity between languages that would be expected due to similar phonetic inventories.[21]

The Levenshtein distance is computed for every country-pair in the data sets. In mono-lingual countries, the respective native language is assigned to the country. In multi-lingual countries, the most prevalent native language is assigned, which was identified using a multitude of sources, including CIA's World Factbook, encyclopedias, and Internet resources.[22]

Table 1 lists the closest and furthest languages toward English and the closest and furthest language-pairs worldwide. The results show that the measurement via the normalized and divided Levenshtein distance is in line with an intuitive guessing about language dissimilarities.

The Genetic Distance

For an exogenous measure of the cultural differences between two countries, it is important that the measure does not reflect bilateral relationships. For instance, religious differences are rooted in past history. However, this history is relatively recent, therefore a measure of religious dissimilarities could reflect past relationships between countries (Guiso et al., 2009). To measure deeper cultural differences between two countries, the genetic distance between the dominant population groups in these countries is used.

The data on genetic differences was originally gathered by Cavalli-Sforza et al. (1994) for 42 subpopulations. Spolaore and Wacziarg (2009) extended this data to genetic differences between 180 countries.[23] In order to do so, Spolaore and Wacziarg combine the frequencies of gene manifestations in populations sampled by Cavalli-Sforza et al. (1994) and the ethnicity composition of countries compiled by Alesina et al. (2003) to derive a measure of the genetic distance between countries.

The genetic distance is measured as the difference in allele frequencies. Alleles are the specific manifestation of a gene, which might differ between individuals. The genetic distance measure as defined by Cavalli-Sforza et al. (1994) is related to the inverse probability that groups of alleles are the same for two populations. Hence, the lower the

[21] The ASJP algorithm allows including or excluding loan words from different languages, e.g., the predominance of former Latin words in many of the European languages. While it makes sense to exclude these loan words in the analysis of the long-term development of languages, these words are included in this analysis, as they lead to certain similarities of languages that might ease the later language transfer in the acquisition process. The necessary software to compute the distance matrix is available at http://www.eva.mpg.de. The complete distance matrix used in this analysis is available upon request.

[22] For example, English is used as the native language in the United Kingdom, because it is a mono-lingual country and English is the national language. In a multi-lingual country such as Canada, English instead of French is employed, because English is the most prevalent native language. A comprehensive index of assigned languages with further explanations is available upon request.

[23] For a more detailed description of the genetic distance data and the construction of the measure, see Spolaore and Wacziarg (2009). The data set is available at http://www.anderson.ucla.edu/faculty_pages/romain.wacziarg.

common frequency of alleles in two populations, the longer these populations have been separated.

Changes in genes, hence the emergence of new alleles, happen randomly at an almost constant time.[24] This constant rate of change over time makes it a reasonable proxy for the time populations spent separated, making the genetic distance an "excellent summary statistic capturing divergence in the whole set of implicit beliefs, customs, habits, biases, conventions, etc. that are transmitted across generations—biologically and/or culturally—with high persistence." (Spolaore and Wacziarg, 2009, p. 471).

Using this measure of genetic distance as a proxy for cultural differences and assuming a reasonable correlation between the measured genetic distance and any unobservable cultural differences should allow the identification of the effect of cultural dissimilarities between two countries on international factor mobility.

3.2 Data on Trade Flows, Portfolio Investment, and Banking Claims

The analysis further relies on bilateral trade data from the International Monetary Fund's Direction of Trade Statistics (DOTS), which reports annual aggregate import and export flows. FOB exports and CIF imports are measured in millions of nominal US dollars.[25] The sample covers the 1950–2006 period for 180 exporter and 187 importer countries. For some country-pairs, the trade data are unavailable, thus around 530,000 non-zero observations of exports are used in the baseline regression. The bilateral trade data are merged with the bilateral linguistic and genetic distance data at the country-pair level. Since both the linguistic and the genetic distance data are symmetric, the direction of trade does not have to be considered.

The data on bilateral asset holdings is provided by the International Monetary Fund in its Coordinated Portfolio Investment Survey (CPIS), which reports year-end data on stocks of cross-border assets measured in millions of nominal US dollars.[26] The sample covers the years 2001–2003 and contains information on portfolio investments of 62 source in 173 host countries, leading to 8,596 non-zero observations of asset stocks used in the baseline regression. As the bilateral trade data, this data set is merged with the bilateral linguistic and genetic distance data at the country-pair level.

The source of the second financial data set is the Consolidated Banking Statistics published by the Bank of International Settlements (BIS), which reports data on bilateral

[24] As evolutionary pressure might direct the random change into certain directions, the genetic distance measure focuses on neutral genes, which are not prone to evolutionary pressure.

[25] The terms FOB and CIF are abbreviations for "Free on Board" and "Cost including Insurance and Freight", respectively, and describe the method used by countries to report the value of imports and exports. For further details on the data set, see http://www.imf.org/external/data.htm.

[26] Further details on the data and the data set itself are available at http://cpis.imf.org.

consolidated international banking claims measured in millions of nominal US dollars on a quarterly basis.[27] The sample covers annual data for the years 1983–2006 and contains information on banking claims of 24 source and 182 host countries, leading to 34,838 non-zero observations of international banking claims used in the baseline regression. Again, this data set is merged with the bilateral linguistic and genetic distance data at the country-pair level.

The data on the control variables are obtained from the CEPII "square" gravity data set compiled by Head et al. (2010).[28] This database contains a set of standard gravity variables which is used in all estimations, such as the geographical distance[29], colonial ties, and information on regional trade agreements.

4 Empirical Findings

4.1 The Effect of Linguistic and Genetic Distance on Trade

This section analyzes the effect of linguistic and genetic distance on bilateral export flows. Linguistic and genetic distance are important determinants in explaining bilateral exports by approximating language and cultural barriers.

In order to examine systematically how linguistic and genetic distance affect export flows, Table 2 reports results for six OLS regressions employing exporter-year and importer-year fixed effects. The country-year specific fixed effects capture the effects of all monadic variables (e.g., GDP per capita, population, and multilateral resistance terms) and remove these variables from the specifications. In each column, standard errors of estimates are robust to heteroskedasticity and correlation of error terms within country-pairs. The analysis starts by estimating bilateral exports on linguistic distance and a host of control variables using the gravity model in Eq. (5) (column (1)). The results show that the coefficient of the linguistic distance is negative and significant at the 0.1% level. The estimated coefficient of -0.078 implies that a one standard deviation increase in the linguistic distance lowers a country's exports by $exp\,(-0.078) - 1 \approx 7.5\%$, which is a sizeable impact. In column (3), linguistic distance is substituted with genetic distance. The effect is similar, but weaker. A one standard deviation increase in the genetic distance lowers bilateral exports by 4.9%. The results are robust in magnitude and in significance when both factors are simultaneously included in the regression (column (5)). This is not

[27] Further details on the data and the data itself are available at http://www.bis.org/statistics/consstats.htm. I use a data set provided by Rose and Spiegel (2009).

[28] For further details on the variables and the data set, see http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8.

[29] In all estimations, the population-weighted great circle distance between large cities of the two countries is used.

surprising, given the relative low correlation between these two variables, and shows that both capture different aspects in the determination of trade flows.[30]

The effect of linguistic and genetic distance is identified by assuming a linear relationship between these variables and bilateral exports. However, imposing wrong functional forms may lead to misleading results for trade barrier estimates (Anderson and van Wincoop, 2004). Eaton and Kortum (2002) assume heterogeneous trade barriers for six different geographic distance intervals. They generalize the treatment of geographical distance with a flexible functional form using a spline approach, which is likely to be more robust to specification errors. To avoid imposing a functional form on the relationship between exports and linguistic and genetic distance, respectively, I follow Eaton and Kortum (2002) and rerun the models using indicator variables. To capture the effect of linguistic distance, a set of five indicator variables is employed. The first indicator comprises all country-pairs speaking the same language, i.e., their linguistic distance is zero. The remaining four indicators turn on for the first up to the fourth quartile of the positive values of the linguistic distance distribution. The same approach is used for the genetic distance. However, instead of employing four indicators for the non-zeros values, the distribution is separated only into two groups. The three dummies indicate the effect on exports for country-pairs with zero, low, and high genetic distance, respectively. The results are shown in columns (2), (4), and (6). The estimated coefficients have the same signs and similar levels of significance. Since the reference group is comprised of country-pairs that speak the same language and have genetically the same dominant populations, respectively, the effects become stronger in magnitude.[31]

The coefficients of the traditional gravity control variables are in line with estimates from the literature. These effects all bear the expected signs, are large in size, economically substantial, and statistically highly significant. For instance, geographical distance reduces trade with an estimated elasticity of about 1.3%. The other dyadic control variables – shared border, shared legal origin, colonial relationship, common regional trade arrangement, both countries GATT/WTO members, and common currency – promote bilateral exports as expected.

Throughout the rest of the analysis of bilateral exports, both linguistic and genetic distance are simultaneously included in the estimation models.

Table 3 reports the results showing whether a country's LDE and the bilateral linguistic distance affect bilateral factor movements above and beyond the inclusion of the direct effects. In doing so, the indicator variables defined in Eq. (7) and (8) are added separately to the baseline regressions from Table 2 (column (5) and (6)). Both indicator variables enter the regressions with a positive sign, but are statistically insignificant except for

[30]The correlation coefficient amounts to 0.175.

[31]To check the robustness of these specifications, the estimations were tested with different sets of indicator variables. The results are broadly comparable to the results presented in Table 2.

column (3). The coefficient of $LDE^{(2)}$ reveals that if at least one trading partner in a pair faces a higher LDE than the bilateral linguistic distance, exports between these countries increase by 6.1%. The coefficients for the linguistic and genetic distance variables are robust in magnitude and in significance when the indicator variables are included in the regressions. As expected, the coefficients of the gravity control variables remain also unchanged.

The results described above present OLS estimates employing exporter-year and importer-year fixed effects, representing the standard way gravity models have been estimated in the literature. The downside of using this specification is that all monadic effects are captured by the fixed effects. Under the stronger assumption that the error term and the random effect are uncorrelated with the variables of interest, the random effects estimator delivers unbiased parameter estimates.

Table 4 presents results obtained by estimating the random effects model in Eq. (11). Due to the unavailability of GDP per capita data for some countries, the sample is reduced to 495,286 country-pair observations. For a proper comparison, the analysis starts by re-estimating the baseline regressions from Table 2 (column (5)) using the smaller sample. The five subsequent columns employ the random effects model. Throughout all estimations, year-specific fixed effects are added to take common time effects into account. While column (2) shows the results without further controls, a comprehensive set of export and import continent dummy variables is added to the regression model in column (3). In Column (4), the continent dummy variables are replaced by sets of world region fixed effects to control for heterogeneous effects within the continents. The final two specifications add the indicator variables defined in Eq. (7) and (8) to the model in column (4).

The coefficients of the income and population variables reveal the expected results. Throughout all specifications, they are positively related to bilateral exports and statistically significant. The parameters of particular interest are η and φ, representing the effect of the exporter's and importer's LDE. The magnitudes of these effects are substantial. A one standard deviation increase in the exporter LDE reduces bilateral exports by between 17.5 and 22.7%, depending on the specification. A similar effect shows up for the importer LDE. The results for the bilateral linguistic and genetic distance are robust in sign and significance, but larger in magnitude.

The results reported in Tables 2-4 demonstrate that linguistic and genetic distance play an import role in explaining international export flows. I hypothesized above that the effect of linguistic distance in the determination of international investments, i.e., cross-border asset stocks and international banking claims, should be comparable to the impact on bilateral exports. However, the role of cultural difference on international investment decisions is theoretically ambiguous. These conjectures are analyzed in the next section.

4.2 The Effect of Linguistic and Genetic Distance on Investments

This section analyzes the effect of linguistic and genetic distance on international investments. In order to examine different types of investments, two data sets are considered. First, the gravity models displayed in Tables 2-4 are re-estimated using data on cross-border asset stocks. Second, data on consolidated international banking claims are employed to rerun the gravity equations shown in Tables 2-4.

Consistent with prior expectations, Tables 5 and 6 shows that linguistic distance enters with negative sign and is highly significant. The estimated coefficients are larger compared to the estimates in Table 2. As expected, the results for the genetic distance are ambiguous. The specifications applying the linear measure of genetic distance report insignificant coefficients (column (3) and (5)). In contrast, employing the splines approach yields negative significant results, but only for country-pairs with a low genetic distance. This result might indicate that investors prefer culturally familiar countries for their investments. However, if they invest in culturally unfamiliar countries, they prefer countries with a relative high difference due to the lower correlation of the stock markets and the higher diversification potential for the portfolios of foreign investors.

Tables 7 and 8 show the results of whether a country's LDE and the bilateral linguistic distance affect bilateral factor movements above and beyond the inclusion of the direct effects. In doing so, the indicator variables defined in Eq. (7) and (8) are added separately to the baseline regressions from Tables 5 and 6, respectively. Both indicator variables are statistically insignificant in all regressions. The coefficients of the variables of interest as well as the gravity control variables remain unchanged as compared to the prior results.

Turing to the results of the random effects model, Tables 9 and 10 reveal that both the exporter's and importer's LDE enter with the expected negative signs and are statistically significant at the 0.1% level. The sizes of the estimated coefficient are substantial. The results of the bilateral linguistic distance are robust in sign and significance, whereas the estimated coefficients of the genetic distance are ambiguous.

5 Conclusion

Empirical examinations of international factor movements, e.g., bilateral trade, financial and migration flows, have a long tradition in the economic literature. Starting with the analysis by Tinbergen (1962), the gravity model plays a key role in this literature. Hence, numerous studies have provided important information about the impact of potential enhancers and barriers on international factor mobility such as membership in the GATT/WTO (e.g., Rose, 2004), colonial linkages (e.g., Head et al., 2010), and geographical distance (e.g., Portes and Rey, 2005).

This paper is motivated by the impression that the relationship between language and cultural differences and factor movements is underinvestigated in the economic literature. The standard measures of language and cultural dissimilarities employed in gravity models (such as indicators for sharing a common language, geographic distance, and colonial ties) are insufficient in picking up the entire influence of these impact factors on bilateral factor movements. Moreover, so far unconsidered communication channels, i.e., English as the lingua franca in international business relations, appear to have a major impact on bilateral flows.

In this paper, I show that language and cultural differences have a significant negative impact on bilateral factor mobility, even after controlling for a host of standard gravity variables. While these effects are economically important and highly significant for international trade flows, only language differences affect financial flows. This finding might be interpreted as a weak support for the diversification motive in international investment transactions, which indicates that the stock markets of culturally unfamiliar countries reveal a lower correlation to the stock markets of foreign investors.

In the preferred specification, I control for a large number of covariates commonly used in gravity equations and account for exporter- and importer-year fixed effects. The results show that an increase in the linguistic distance by one standard deviation decreases bilateral exports by 7.4% and both cross-border asset stocks and international banking claims by about 12%.

Although the findings on the isolated effect of a country's linguistic distance toward English (LDE) suggest highly significant and economically important effects across all specifications and datasets, they should be interpret with some caution. To deliver unbiased estimates of the true effect, the employed random effects model rests on strong identification assumptions. Other empirical methods such as the approach suggested by Baier and Bergstrand (2009) and the analysis of measures of LDE that additionally vary over time may shed some further light on the relation between international factor mobility and English proficiency. Only future research will be able to give a comprehensive answer.

References

Adsera, A. and M. Pytlikova (2012). The role of language in shaping international migration. CReAM Discussion Paper No. 06/12.

Aggarwal, R., C. Kearney, and B. Lucey (2012). Gravity and culture in foreign portfolio investment. *Journal of Banking & Finance 36*(2), 525–538.

Ahearne, A. G., W. L. Griever, and F. E. Warnock (2004). Information costs and home bias: an analysis of US holdings of foreign equities. *Journal of International Economics 62*(2), 313–336.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth 8*(2), 155–194.

Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review 69*(1), 106–116.

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

Anderson, J. E. and E. van Wincoop (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review 93*(1), 170–192.

Anderson, J. E. and E. van Wincoop (2004). Trade costs. *Journal of Economic Literature 42*(3), 691–751.

Anderson, J. E. and Y. V. Yotov (2010a). Specialization: Pro-and Anti-Globalizing, 1990-2002. NBER Working Paper No. 16301.

Anderson, J. E. and Y. V. Yotov (2010b). The Changing Incidence of Geography. *American Economic Review 100*(5), 2157–2186.

Aviat, A. and N. Coeurdacier (2007). The geography of trade in goods and asset holdings. *Journal of International Economics 71*(1), 22–51.

Baier, S. L. and J. H. Bergstrand (2009). Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation. *Journal of International Economics 77*(1), 77–85.

Bakker, D., A. Müller, V. Velupillai, S. Wichmann, C. H. Brown, P. Brown, D. Egorov, R. Mailhammer, A. Grant, and E. W. Holman (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology 13*(1), 169–181.

Baldwin, R. and D. Taglioni (2006). Gravity for dummies and dummies for gravity equations. NBER Working Paper No. 12516.

Baltagi, B. H., P. Egger, and M. Pfaffermayr (2003). A generalized design for bilateral trade flow models. *Economics Letters 80*(3), 391–397.

Bergstrand, J. H. (1985). The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence. *Review of Economics and Statistics 67*(3), 474–481.

Bergstrand, J. H. and P. H. Egger (2011). Gravity Equations and Economic Frictions in the World Economy. In D. Bernhofen, R. Falvey, D. Greenaway, and U. Kreickemeier (Eds.), *Palgrave Handbook of International Trade*, pp. 532–570. London, UK: Palgrave Macmillan.

Beugelsdijk, S. and B. Frijns (2010). A cultural explanation of the foreign bias in international asset allocation. *Journal of Banking & Finance 34*(9), 2121–2131.

Brenton, P., F. Di Mauro, and M. Lücke (1999). Economic Integration and FDI: An Empirical Analysis of Foreign Investment in the EU and in Central and Eastern Europe. *Empirica 26*(2), 95–121.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza (1994). *The history and geography of human genes.* Princeton, NJ: Princeton University Press.

Chan, K., V. Covrig, and L. K. Ng (2005). What Determines the Domestic Bias and Foreign Bias? Evidence from Mutual Fund Equity Allocations Worldwide. *The Journal of Finance 60*(3), 1495–1534.

Chiswick, B. R. and P. W. Miller (1999). English language fluency among immigrants in the United States. In S. W. Polachek (Ed.), *Research in Labor Economics*, Volume 17, pp. 151–200. Oxford, UK: JAI Press.

Coeurdacier, N. and H. Rey (2013). Home Bias in Open Economy Financial Macroeconomics. *Journal of Economic Literature 51*(1), 63–115.

Daude, C. and M. Fratzscher (2008). The pecking order of cross-border investment. *Journal of International Economics 74*(1), 94–119.

Eaton, J. and S. Kortum (2002). Technology, Geography, and Trade. *Econometrica 70*(5), 1741–1779.

Egger, P. H. and A. Lassmann (2012). The language effect in international trade: A meta-analysis. *Economics Letters 116*(2), 221–224.

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, Insurgency, and Civil War. *The American Political Science Review 97*(1), 75–90.

Felbermayr, G. J. and F. Toubal (2010). Cultural proximity and trade. *European Economic Review 54*(2), 279–293.

French, K. R. and J. M. Poterba (1991). Investor Diversification and International Equity Markets. *American Economic Review 81*(2), 222–226.

Giuliano, P., A. Spilimbergo, and G. Tonon (2013). Genetic distance, transportation costs, and trade. *Journal of Economic Geography 1001*(April 2006), 1–39.

Grinblatt, M. and M. Keloharju (2001). How Distance, Language, and Culture Influence Stockholdings and Trades. *The Journal of Finance 56*(3), 1053–1073.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural biases in economic exchange? *Quarterly Journal of Economics 124*(3), 1095–1131.

Head, K. and T. Mayer (2013). Gravity Equations: Workhorse, Toolkit, and Cookbook. In *Handbook of International Economics, Vol. 4*, pp. forthcoming. Elsevier.

Head, K., T. Mayer, and J. Ries (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics 81*(1), 1–14.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics 123*(2), 441–487.

Holman, E. W., C. H. Brown, S. Wichmann, A. Müller, V. Velupillai, H. Hammarström, S. Sauppe, H. Jung, D. Bakker, P. Brown, O. Belyaev, M. Urban, R. Mailhammer, J.-M. List, and D. Egorov (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology 52*(6), 841–875.

Hornok, C. (2012). Gravity or Dummies? The Limits of Identification in Gravity Estimations. Department of Economics Working Papers.

Hutchinson, W. K. (2005). "Linguistic Distance" as a Determinant of Bilateral Trade. *Southern Economic Journal 72*(1), 1–15.

Isphording, I. E. and S. Otten (2011). Linguistic Distance and the Language Fluency of Immigrants. Ruhr Economic Papers #274.

Isphording, I. E. and S. Otten (2013). The Costs of Babylon – Linguistic Distance in Applied Economics. *Review of International Economics, 21*(2), 354–369.

Ku, H. and A. Zussman (2010). Lingua franca: The role of English in international trade. *Journal of Economic Behavior & Organization 75*(2), 250–260.

Lane, P. R. and G. M. Milesi-Ferretti (2003). International financial integration. *IMF Staff Papers 50*(S1), 82–113.

Lane, P. R. and G. M. Milesi-Ferretti (2008). International Investment Patterns. *Review of Economics and Statistics 90*(3), 538–549.

Lewis, K. K. (1999). Trying to Explain Home Bias in Equities and Consumption. *Journal of Economic Literature 37*(2), 571–608.

Llach, A. M. (2010). An overview of variables affecting lexical transfer in writing: A review study. *International Journal of Linguistics 2*(1), E2.

Lohmann, J. (2011). Do language barriers affect trade? *Economics Letters 110*(2), 159–162.

Martin, P. and H. Rey (2004). Financial super-markets: size matters for asset trade. *Journal of International Economics 64*(2), 335–361.

Matyas, L., C. Hornok, and D. Pus (2013). The Formulation and Estimation of Random Effects Panel Data Models of Trade. MPRA Paper No. 36789.

Melitz, J. (2008). Language and foreign trade. *European Economic Review 52*(4), 667–699.

Mélitz, J. and F. Toubal (2012). Native language, spoken language, translation and trade. CEPR Discussion Paper No. 8994.

Petroni, F. and M. Serva (2010). Measures of lexical distance between languages. *Physica A: Statistical Mechanics and its Applications 389*(11), 2280–2283.

Portes, R. and H. Rey (2005). The determinants of cross-border equity flows. *Journal of International Economics 65*(2), 269–296.

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

Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review 94*(1), 98–114.

Rose, A. K. and M. M. Spiegel (2009). Noneconomic Engagement and International Exchange: The Case of Environmental Treaties. *Journal of Money, Credit and Banking 41*(2-3), 337–363.

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

Serva, M. (2011). Phylogeny and geometry of languages from normalized levenshtein distance. Ithaca, NY: Cornell University Library.

Spolaore, E. and R. Wacziarg (2009). The Diffusion of Development. *The Quarterly Journal of Economics 124*(2), 469–529.

Spolaore, E. and R. Wacziarg (2012). Long-Term Barriers to the International Diffusion of Innovations. In J. Frankel and C. Pissarides (Eds.), *NBER International Seminar on Macroeconomics 2011*, pp. 11–46. Chicago, IL: University of Chicago Press.

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society 96*(4), 452–463.

Tesar, L. L. and I. M. Werner (1995). Home bias and high turnover. *Journal of International Money and Finance 14*(4), 467–492.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. New York, NY: The Twentieth Century Fund.

Tables

Table 1: Closest and Furthest Language Pairs with Respect to the Levenshtein Distance

Closest		Furthest	
Language	Distance	Language	Distance
Distance to English			
Afrikaans	62.08	Vietnamese	104.06
Dutch	63.22	Turkmen	103.84
Norwegian	64.12	Hakka (China)	103.10
Swedish	64.40	Cambodian	103.00
Frisian	69.49	Finnish	102.27
Language pairs worldwide			
Croatian Slovenian	28.36	Laotien Korean	106.35
Jamaican Creole Vincentian Creole	30.88	Fijian Shona (Zimbabwe)	106.75
Egyptian Arabic Iraqi Arabic	31.44	Vietnamese Western Punjabi (Pakistan)	107.33

Notes: – The table shows the five closest and furthest languages toward English and the three closest and furthest language-pairs worldwide according to the normalized and divided Levenshtein distance. – Only languages spoken within samples are listed. – Geographic origin of language in parentheses.

Table 2: Effect of Linguistic and Genetic Distance on Bilateral Exports

	Linguistic Distance		Genetic Distance		Baseline Model	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Geo. Distance	−1.268***	−1.274***	−1.279***	−1.274***	−1.250***	−1.257***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.023)
Shared Border	0.445***	0.444***	0.441***	0.409***	0.430***	0.409***
	(0.092)	(0.092)	(0.092)	(0.092)	(0.091)	(0.092)
Shared Legal	0.466***	0.486***	0.499***	0.498***	0.466***	0.488***
	(0.027)	(0.027)	(0.026)	(0.026)	(0.027)	(0.027)
Colonial Relationship	1.227***	1.261***	1.298***	1.289***	1.236***	1.263***
	(0.097)	(0.099)	(0.098)	(0.098)	(0.097)	(0.099)
RTA	0.619***	0.614***	0.615***	0.605***	0.616***	0.604***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Both GATT/WTO	0.336***	0.342***	0.355***	0.345***	0.337***	0.337***
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
Shared Currency	0.837***	0.821***	0.788***	0.808***	0.799***	0.804***
	(0.095)	(0.095)	(0.095)	(0.095)	(0.096)	(0.095)
Linguistic Distance	−0.078***	–	–	–	−0.077***	–
	(0.016)				(0.016)	
Linguistic Distance (Ref. Same Lang.)						
Linguistic Distance Spline 2	–	−0.118	–	–	–	−0.110
		(0.075)				(0.076)
Linguistic Distance Spline 3	–	−0.211**	–	–	–	−0.185*
		(0.081)				(0.082)
Linguistic Distance Spline 4	–	−0.245**	–	–	–	−0.214*
		(0.083)				(0.084)
Linguistic Distance Spline 5	–	−0.293***	–	–	–	−0.260**
		(0.085)				(0.086)
Genetic Distance	–	–	−0.050**	–	−0.049**	–
			(0.017)		(0.017)	
Genetic Distance (Ref. Same Pop.)						
Genetic Distance Spline 2	–	–	–	−0.175***	–	−0.136*
				(0.053)		(0.054)
Genetic Distance Spline 3	–	–	–	−0.205***	–	−0.167**
				(0.055)		(0.056)
Exporter-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Importer-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.712	0.712	0.712	0.712	0.712	0.712
Observations	529,562	529,562	529,562	529,562	529,562	529,562

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral export flows.*

Table 3: Effect of the Bilateral LDE Indicator on Bilateral Exports

	Both LDE > LD		**One LDE > LD**	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Geo. Distance	−1.250***	−1.257***	−1.252***	−1.258***
	(0.023)	(0.023)	(0.023)	(0.023)
Shared Border	0.426***	0.408***	0.428***	0.409***
	(0.091)	(0.092)	(0.091)	(0.092)
Shared Legal	0.464***	0.486***	0.461***	0.484***
	(0.027)	(0.027)	(0.027)	(0.027)
Colonial Relationship	1.241***	1.264***	1.242***	1.265***
	(0.097)	(0.099)	(0.098)	(0.099)
RTA	0.615***	0.604***	0.613***	0.604***
	(0.038)	(0.038)	(0.038)	(0.038)
Both GATT/WTO	0.335***	0.337***	0.335***	0.336***
	(0.047)	(0.047)	(0.047)	(0.047)
Shared Currency	0.797***	0.804***	0.796***	0.804***
	(0.096)	(0.095)	(0.096)	(0.095)
Linguistic Distance	−0.073***	–	−0.074***	–
	(0.016)		(0.016)	
Linguistic Distance (Ref. Same Lang.)				
Linguistic Distance Spline 2	–	−0.117	–	−0.121
		(0.077)		(0.076)
Linguistic Distance Spline 3	–	−0.184*	–	−0.189*
		(0.082)		(0.082)
Linguistic Distance Spline 4	–	−0.209*	–	−0.211*
		(0.086)		(0.085)
Linguistic Distance Spline 5	–	−0.252**	–	−0.244**
		(0.088)		(0.088)
Genetic Distance	−0.047**	–	−0.046**	–
	(0.017)		(0.017)	
Genetic Distance (Ref. Same Pop.)				
Genetic Distance Spline 2	–	−0.135*	–	−0.135*
		(0.054)		(0.054)
Genetic Distance Spline 3	–	−0.167**	–	−0.167**
		(0.056)		(0.056)
$LDE^{(1)}$	0.046	0.020	–	–
	(0.036)	(0.045)		
$LDE^{(2)}$	–	–	0.059*	0.033
			(0.027)	(0.034)
Exporter-Year FE	Yes	Yes	Yes	Yes
Importer-Year FE	Yes	Yes	Yes	Yes
R^2	0.712	0.712	0.712	0.712
Observations	529,562	529,562	529,562	529,562

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral export flows.*

Table 4: Effect of the Linguistic Distance toward English on Bilateral Exports

	Baseline	RE-Model			RE-Model incl. LDE	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Pop, origin	–	0.950***	0.926***	0.851***	0.850***	0.850***
		(0.010)	(0.010)	(0.012)	(0.012)	(0.012)
ln Pop, dest	–	0.821***	0.816***	0.839***	0.838***	0.838***
		(0.010)	(0.010)	(0.012)	(0.012)	(0.012)
ln GDP/pc, origin	–	0.945***	0.911***	0.892***	0.892***	0.892***
		(0.013)	(0.014)	(0.015)	(0.015)	(0.015)
ln GDP/pc, dest	–	0.658***	0.650***	0.646***	0.647***	0.647***
		(0.011)	(0.012)	(0.013)	(0.013)	(0.013)
ln Geo. Distance	−1.271***	−1.189***	−1.150***	−1.264***	−1.262***	−1.264***
	(0.023)	(0.022)	(0.027)	(0.027)	(0.027)	(0.027)
Shared Border	0.414***	0.890***	0.985***	0.733***	0.726***	0.730***
	(0.092)	(0.092)	(0.092)	(0.092)	(0.092)	(0.092)
Shared Legal	0.477***	0.347***	0.477***	0.525***	0.522***	0.518***
	(0.027)	(0.036)	(0.035)	(0.035)	(0.035)	(0.035)
Colonial Relationship	1.183***	1.761***	1.514***	1.537***	1.544***	1.543***
	(0.096)	(0.112)	(0.115)	(0.116)	(0.116)	(0.116)
RTA	0.595***	0.393***	0.393***	0.392***	0.392***	0.392***
	(0.038)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Both GATT/WTO	0.292***	0.142***	0.155***	0.154***	0.154***	0.154***
	(0.049)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)
Shared Currency	0.747***	0.523***	0.550***	0.541***	0.541***	0.541***
	(0.103)	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)
Linguistic Distance	−0.071***	−0.122***	−0.180***	−0.175***	−0.165***	−0.170***
	(0.016)	(0.018)	(0.019)	(0.019)	(0.020)	(0.019)
Genetic Distance	−0.055**	−0.095***	−0.019	−0.091***	−0.089***	−0.088***
	(0.018)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
LDE, origin	–	−0.192***	−0.219***	−0.242***	−0.248***	−0.258***
		(0.017)	(0.017)	(0.022)	(0.022)	(0.023)
LDE, dest	–	−0.242***	−0.277***	−0.227***	−0.233***	−0.243***
		(0.016)	(0.017)	(0.022)	(0.022)	(0.022)
$LDE^{(1)}$	–	–	–	–	0.086†	–
					(0.045)	
$LDE^{(2)}$	–	–	–	–	–	0.087*
						(0.034)
Exporter-Year FE	Yes	No	No	No	No	No
Importer-Year FE	Yes	No	No	No	No	No
Year FE	No	Yes	Yes	Yes	Yes	Yes
Continent FE	No	No	Yes	No	No	No
World Region FE	No	No	No	Yes	Yes	Yes
R^2	0.718	0.587	0.593	0.613	0.613	0.613
Observations	495,286	495,286	495,286	495,286	495,286	495,286

*Notes: – Significant at: ***0.1% level; **1% level; *5% level; †10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral export flows.*

Table 5: Effect of Linguistic and Genetic Distance on Cross-Border Asset Stocks

	Linguistic Distance		**Genetic Distance**		**Baseline Model**	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Geo. Distance	−1.007***	−1.030***	−1.085***	−1.079***	−1.024***	−1.054***
	(0.067)	(0.067)	(0.065)	(0.066)	(0.070)	(0.070)
Shared Border	0.225	0.218	0.210	0.145	0.218	0.142
	(0.187)	(0.188)	(0.185)	(0.182)	(0.187)	(0.185)
Shared Legal	0.335***	0.350***	0.424***	0.410***	0.335***	0.348***
	(0.086)	(0.087)	(0.080)	(0.080)	(0.086)	(0.086)
Colonial Relationship	0.375*	0.425*	0.468**	0.445*	0.371*	0.419*
	(0.179)	(0.183)	(0.177)	(0.173)	(0.178)	(0.179)
RTA	0.496***	0.486***	0.478***	0.486***	0.496***	0.489***
	(0.117)	(0.117)	(0.116)	(0.116)	(0.117)	(0.116)
Shared Currency	0.620***	0.613***	0.588***	0.615***	0.619***	0.630***
	(0.156)	(0.153)	(0.152)	(0.154)	(0.156)	(0.155)
Linguistic Distance	−0.125**	–	–	–	−0.128**	–
	(0.040)				(0.040)	
Linguistic Distance (Ref. Same Lang.)						
Linguistic Distance Spline 2	–	−0.173	–	–	–	−0.086
		(0.186)				(0.187)
Linguistic Distance Spline 3	–	−0.395*	–	–	–	−0.314†
		(0.189)				(0.190)
Linguistic Distance Spline 4	–	−0.372†	–	–	–	−0.278
		(0.205)				(0.207)
Linguistic Distance Spline 5	–	−0.446*	–	–	–	−0.379†
		(0.220)				(0.225)
Genetic Distance	–	–	0.050	–	0.069	–
			(0.092)		(0.094)	
Genetic Distance (Ref. Same Pop.)						
Genetic Distance Spline 2	–	–	–	−0.367***	–	−0.337**
				(0.111)		(0.112)
Genetic Distance Spline 3	–	–	–	−0.120	–	−0.060
				(0.149)		(0.157)
Origin-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.766	0.766	0.765	0.766	0.766	0.767
Observations	8,596	8,596	8,596	8,596	8,596	8,596

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral cross-border asset stocks.*

Table 6: Effect of Linguistic and Genetic Distance on Bilateral Banking Claims

	Linguistic Distance		Genetic Distance		Baseline Model	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Geo. Distance	−0.987***	−0.992***	−1.037***	−1.032***	−1.009***	−1.006***
	(0.077)	(0.076)	(0.070)	(0.078)	(0.071)	(0.078)
Shared Border	0.013	0.017	0.005	−0.034	−0.003	−0.015
	(0.206)	(0.207)	(0.200)	(0.204)	(0.204)	(0.208)
Shared Legal	0.147*	0.165**	0.215***	0.217***	0.144*	0.169**
	(0.059)	(0.057)	(0.056)	(0.057)	(0.059)	(0.058)
Colonial Relationship	1.032***	1.052***	1.135***	1.122***	1.031***	1.048***
	(0.117)	(0.116)	(0.114)	(0.115)	(0.116)	(0.116)
RTA	−0.099	−0.087	−0.128	−0.123	−0.105	−0.091
	(0.112)	(0.112)	(0.110)	(0.110)	(0.112)	(0.112)
Shared Currency	−0.165	−0.169	−0.160	−0.149	−0.171	−0.168
	(0.162)	(0.163)	(0.153)	(0.157)	(0.163)	(0.165)
Linguistic Distance	−0.123***	–	–	–	−0.127***	–
	(0.033)				(0.033)	
Linguistic Distance (Ref. Same Lang.)						
Linguistic Distance Spline 2	–	−0.458**	–	–	–	−0.437**
		(0.145)				(0.147)
Linguistic Distance Spline 3	–	−0.604***	–	–	–	−0.567***
		(0.155)				(0.158)
Linguistic Distance Spline 4	–	−0.534**	–	–	–	−0.505**
		(0.173)				(0.178)
Linguistic Distance Spline 5	–	−0.404*	–	–	–	−0.386*
		(0.172)				(0.177)
Genetic Distance	–	–	0.080	–	0.107	–
			(0.099)		(0.098)	
Genetic Distance (Ref. Same Pop.)						
Genetic Distance Spline 2	–	–	–	−0.175†	–	−0.100
				(0.093)		(0.095)
Genetic Distance Spline 3	–	–	–	0.031	–	0.055
				(0.141)		(0.145)
Origin-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.815	0.815	0.814	0.814	0.815	0.816
Observations	34,838	34,838	34,838	34,838	34,838	34,838

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral consolidated international banking claims.*

Table 7: Effect of the Bilateral LDE Indicator on Cross-Border Asset Stocks

	Both LDE > LD		One LDE > LD	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Geo. Distance	−1.023***	−1.053***	−1.024***	−1.050***
	(0.070)	(0.070)	(0.070)	(0.070)
Shared Border	0.201	0.140	0.218	0.152
	(0.187)	(0.185)	(0.187)	(0.185)
Shared Legal	0.329***	0.346***	0.335***	0.363***
	(0.087)	(0.088)	(0.088)	(0.088)
Colonial Relationship	0.385*	0.420*	0.372*	0.414*
	(0.182)	(0.179)	(0.179)	(0.178)
RTA	0.497***	0.489***	0.496***	0.488***
	(0.117)	(0.116)	(0.117)	(0.116)
Shared Currency	0.617***	0.630***	0.619***	0.636***
	(0.156)	(0.155)	(0.156)	(0.154)
Linguistic Distance	−0.121**	–	−0.127**	–
	(0.041)		(0.041)	
Linguistic Distance (Ref. Same Lang.)				
Linguistic Distance Spline 2	–	−0.096	–	−0.049
		(0.194)		(0.188)
Linguistic Distance Spline 3	–	−0.316†	–	−0.307
		(0.189)		(0.189)
Linguistic Distance Spline 4	–	−0.278	–	−0.291
		(0.207)		(0.207)
Linguistic Distance Spline 5	–	−0.376†	–	−0.436†
		(0.226)		(0.229)
Genetic Distance	0.084	–	0.069	–
	(0.097)		(0.095)	
Genetic Distance (Ref. Same Pop.)				
Genetic Distance Spline 2	–	−0.335**	–	−0.344**
		(0.112)		(0.112)
Genetic Distance Spline 3	–	−0.058	–	−0.059
		(0.158)		(0.157)
$LDE^{(1)}$	0.107	0.023	–	–
	(0.124)	(0.144)		
$LDE^{(2)}$	–	–	0.003	−0.107
			(0.082)	(0.094)
Origin-Year FE	Yes	Yes	Yes	Yes
Destination-Year FE	Yes	Yes	Yes	Yes
R^2	0.766	0.767	0.766	0.767
Observations	8,596	8,596	8,596	8,596

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral cross-border asset stocks.*

Table 8: Effect of the Bilateral LDE Indicator on Bilateral Banking Claims

	Both LDE > LD		One LDE > LD	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Geo. Distance	−1.008***	−1.006***	−1.009***	−1.004***
	(0.071)	(0.078)	(0.071)	(0.078)
Shared Border	0.009	−0.017	0.004	−0.012
	(0.206)	(0.210)	(0.204)	(0.208)
Shared Legal	0.149*	0.167**	0.155**	0.173**
	(0.059)	(0.060)	(0.059)	(0.058)
Colonial Relationship	1.021***	1.049***	1.017***	1.044***
	(0.117)	(0.117)	(0.117)	(0.116)
RTA	−0.105	−0.091	−0.102	−0.090
	(0.112)	(0.112)	(0.112)	(0.112)
Shared Currency	−0.171	−0.168	−0.171	−0.166
	(0.163)	(0.165)	(0.162)	(0.164)
Linguistic Distance	−0.130***	–	−0.131***	–
	(0.033)		(0.033)	
Linguistic Distance (Ref. Same Lang.)				
Linguistic Distance Spline 2	–	−0.439**	–	−0.433**
		(0.149)		(0.147)
Linguistic Distance Spline 3	–	−0.566***	–	−0.575***
		(0.159)		(0.158)
Linguistic Distance Spline 4	–	−0.503**	–	−0.516**
		(0.179)		(0.179)
Linguistic Distance Spline 5	–	−0.384*	–	−0.411*
		(0.180)		(0.185)
Genetic Distance	0.100	–	0.102	–
	(0.098)		(0.098)	
Genetic Distance (Ref. Same Pop.)				
Genetic Distance Spline 2	–	−0.100	–	−0.099
		(0.094)		(0.094)
Genetic Distance Spline 3	–	0.056	–	0.059
		(0.145)		(0.145)
$LDE^{(1)}$	−0.052	0.009	–	–
	(0.088)	(0.100)		
$LDE^{(2)}$	–	–	−0.081	−0.031
			(0.056)	(0.068)
Origin-Year FE	Yes	Yes	Yes	Yes
Destination-Year FE	Yes	Yes	Yes	Yes
R^2	0.815	0.816	0.815	0.816
Observations	34,838	34,838	34,838	34,838

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral consolidated international banking claims.*

Table 9: Effect of the Linguistic Distance toward English on Cross-Border Asset Stocks

	Baseline	RE-Model			RE-Model incl. LDE	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Pop, origin	–	0.397***	0.414***	0.388***	0.386***	0.386***
		(0.025)	(0.024)	(0.029)	(0.029)	(0.029)
ln Pop, dest	–	0.704***	0.714***	0.690***	0.689***	0.689***
		(0.026)	(0.026)	(0.030)	(0.031)	(0.030)
ln GDP/pc, origin	–	1.597***	1.763***	1.639***	1.644***	1.649***
		(0.052)	(0.059)	(0.091)	(0.091)	(0.091)
ln GDP/pc, dest	–	1.160***	1.157***	1.229***	1.231***	1.239***
		(0.038)	(0.042)	(0.052)	(0.052)	(0.052)
ln Geo. Distance	−1.049***	−0.629***	−0.698***	−0.915***	−0.913***	−0.922***
	(0.072)	(0.057)	(0.068)	(0.069)	(0.069)	(0.069)
Shared Border	0.173	0.631**	0.639**	0.279	0.245	0.254
	(0.188)	(0.201)	(0.201)	(0.214)	(0.213)	(0.212)
Shared Legal	0.331***	0.379***	0.332**	0.349***	0.335**	0.320**
	(0.087)	(0.110)	(0.110)	(0.106)	(0.107)	(0.107)
Colonial Relationship	0.394*	0.583**	0.609**	0.777***	0.805***	0.809***
	(0.181)	(0.209)	(0.211)	(0.214)	(0.215)	(0.215)
RTA	0.485***	0.160†	0.144†	0.187*	0.188*	0.186*
	(0.124)	(0.082)	(0.082)	(0.085)	(0.085)	(0.084)
Shared Currency	0.694***	2.559***	2.634***	1.956***	1.944***	1.939***
	(0.160)	(0.142)	(0.146)	(0.162)	(0.163)	(0.162)
Linguistic Distance	−0.132**	−0.116**	−0.155**	−0.091†	−0.078	−0.083†
	(0.042)	(0.044)	(0.048)	(0.049)	(0.050)	(0.049)
Genetic Distance	0.081	0.227***	0.154*	−0.006	0.013	0.000
	(0.092)	(0.061)	(0.061)	(0.071)	(0.073)	(0.071)
LDE, origin	–	−0.369***	−0.522***	−0.788***	−0.804***	−0.834***
		(0.039)	(0.047)	(0.062)	(0.063)	(0.066)
LDE, dest	–	−0.378***	−0.285***	−0.329***	−0.340***	−0.364***
		(0.044)	(0.050)	(0.068)	(0.069)	(0.071)
$LDE^{(1)}$	–	–	–	–	0.211	–
					(0.146)	
$LDE^{(2)}$	–	–	–	–	–	0.217*
						(0.100)
Origin-Year FE	Yes	No	No	No	No	No
Destination-Year FE	Yes	No	No	No	No	No
Year FE	No	Yes	Yes	Yes	Yes	Yes
Continent FE	No	No	Yes	No	No	No
World Region FE	No	No	No	Yes	Yes	Yes
R^2	0.767	0.580	0.601	0.640	0.641	0.641
Observations	8,277	8,277	8,277	8,277	8,277	8,277

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral cross-border asset stocks.*

Table 10: Effect of the Linguistic Distance toward English on Bilateral Banking Claims

	Baseline	**RE-Model**			**RE-Model incl. LDE**	
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
ln Pop, origin	–	0.578***	0.603***	0.587***	0.583***	0.588***
		(0.027)	(0.031)	(0.032)	(0.032)	(0.032)
ln Pop, dest	–	0.569***	0.578***	0.573***	0.573***	0.573***
		(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
ln GDP/pc, origin	–	1.224***	1.181***	0.927***	0.928***	0.927***
		(0.063)	(0.069)	(0.097)	(0.098)	(0.097)
ln GDP/pc, dest	–	0.766***	0.759***	0.753***	0.753***	0.753***
		(0.034)	(0.037)	(0.039)	(0.039)	(0.039)
ln Geo. Distance	−1.043***	−0.568***	−0.586***	−0.784***	−0.785***	−0.783***
	(0.066)	(0.049)	(0.065)	(0.064)	(0.064)	(0.064)
Shared Border	−0.036	0.837***	0.828***	0.247	0.221	0.251
	(0.200)	(0.212)	(0.219)	(0.206)	(0.209)	(0.205)
Shared Legal	0.149*	0.261***	0.311***	0.265***	0.252***	0.269***
	(0.059)	(0.074)	(0.076)	(0.072)	(0.073)	(0.072)
Colonial Relationship	1.066***	1.032***	0.918***	1.156***	1.174***	1.151***
	(0.106)	(0.128)	(0.137)	(0.131)	(0.133)	(0.131)
RTA	−0.150	−0.031	−0.037	−0.039	−0.039	−0.038
	(0.111)	(0.060)	(0.061)	(0.061)	(0.061)	(0.061)
Shared Currency	−0.143	1.272***	1.280***	1.058***	1.056***	1.059***
	(0.163)	(0.155)	(0.155)	(0.151)	(0.151)	(0.151)
Linguistic Distance	−0.115***	−0.162***	−0.156***	−0.102*	−0.094*	−0.104*
	(0.033)	(0.041)	(0.043)	(0.041)	(0.042)	(0.041)
Genetic Distance	0.077	−0.065	0.017	−0.115*	−0.110*	−0.115*
	(0.091)	(0.043)	(0.046)	(0.054)	(0.054)	(0.054)
LDE, origin	–	−0.028	−0.121***	−0.227***	−0.238***	−0.218***
		(0.028)	(0.034)	(0.042)	(0.043)	(0.045)
LDE, dest	–	−0.427***	−0.394***	−0.454***	−0.462***	−0.448***
		(0.047)	(0.052)	(0.067)	(0.068)	(0.068)
$LDE^{(1)}$	–	–	–	–	0.122	–
					(0.114)	
$LDE^{(2)}$	–	–	–	–	–	−0.036
						(0.066)
Origin-Year FE	Yes	No	No	No	No	No
Destination-Year FE	Yes	No	No	No	No	No
Year FE	No	Yes	Yes	Yes	Yes	Yes
Continent FE	No	No	Yes	No	No	No
World Region FE	No	No	No	Yes	Yes	Yes
R^2	0.818	0.561	0.571	0.630	0.630	0.630
Observations	33,984	33,984	33,984	33,984	33,984	33,984

*Notes: – Significant at: *** 0.1% level; ** 1% level; * 5% level; † 10% level. – Cluster-robust standard errors are reported in parentheses. – The dependent variable is the logarithm of bilateral consolidated international banking claims.*