

# ACCEPTED VERSION

Benedikt Heid and Wenxi Lu

## Genetic distance, cultural differences, and the formation of regional trade agreements

Review of World Economics, 2021; 158:1-23

© Kiel Institute 2021

*This is a post-peer-review, pre-copyedit version of an article published in **Review of World Economics**, The final authenticated version is available online at:*

<http://dx.doi.org/10.1007/s10290-021-00410-9>

### PERMISSIONS

<https://www.springer.com/gp/open-access/publication-policies/self-archiving-policy>

#### Self-archiving for articles in subscription-based journals

Springer journals' [policy on preprint sharing](#).

By signing the Copyright Transfer Statement you still retain substantial rights, such as self-archiving:

*Author(s) are permitted to self-archive a pre-print and an author's **accepted manuscript** version of their Article.*

.....

*b. An Author's Accepted Manuscript (AAM) is the version accepted for publication in a journal following peer review but prior to copyediting and typesetting that can be made available under the following conditions:*

*(i) Author(s) retain the right to make an AAM of their Article available on their own personal, self-maintained website immediately on acceptance,*

*(ii) Author(s) retain the right to make an AAM of their Article available for public release on any of the following 12 months after first publication ("Embargo Period"): their employer's internal website; their institutional and/or funder repositories. AAMs may also be deposited in such repositories immediately on acceptance, provided that they are not made publicly available until after the Embargo Period.*

*An acknowledgement in the following form should be included, together with a link to the published version on the publisher's website: "This is a post-peer-review, pre-copyedit version of an article published in [insert journal title]. The final authenticated version is available online at: [http://dx.doi.org/\[insert DOI\]](http://dx.doi.org/[insert DOI])".*

When publishing an article in a subscription journal, without open access, authors sign the Copyright Transfer Statement (CTS) which also details Springer's self-archiving policy.

See Springer Nature [terms of reuse](#) for archived author accepted manuscripts (AAMs) of subscription articles.

**15 December 2023**

<http://hdl.handle.net/2440/140189>

# Genetic distance, cultural differences, and the formation of regional trade agreements\*

Benedikt Heid

Wenxi Lu

University of Adelaide and CESifo

Harbin Institute of Technology

January 22, 2021

**Abstract:** Genetic distance between countries' populations has been shown to proxy cross-country differences in cultures and preferences. In an unbalanced panel of 133 countries from 1970 to 2012, we find that higher genetic distance between two countries decreases their probability of having a trade agreement, even when controlling for geographic distance and other controls. The impact of cultural differences proxied by genetic distance is persistent over time and economically significant: While increasing the geographic distance between two countries by one percent decreases the probability of a regional trade agreement by 0.11 percentage points, increasing their genetic distance by one percent decreases the probability by 0.06 percentage points.

**JEL Classification Codes:** F13; F14; F15; Z10

**Keywords:** trade agreements, trade policy, trade negotiations, genetic distance, cultural differences

---

\* Contact information: Benedikt Heid, School of Economics, University of Adelaide, CESifo. 10 Pulteney St, SA 5000 Adelaide, Australia. Phone: +61 (0) 831 34930. Email: benedikt.heid@adelaide.edu.au; Wenxi Lu, Department of Economics and Trade, School of Economics and Management, Harbin Institute of Technology, Heilongjiang Province, 150000, China. Email: wenxi.lu@hit.edu.cn. Heid gratefully acknowledges financial support from the Australian Research Council (DP190103524) and Lu from the National Natural Science Foundation of China (91846301). We thank Ralph-Christopher Bayer and Laura Márquez-Ramos for useful comments. All remaining errors are our own.

## 1 Introduction

The number of regional trade agreements (RTAs) has steadily increased during the last decades: Whereas less than two percent of country pairs had an RTA in 1970, and four percent around the end of the Cold War in 1990, more than 16% had in 2012, see Figure 1. With the stalled multilateral negotiations of the Doha Round within the World Trade Organization, the pace of newly concluded RTAs further increased, and is likely to continue to do so in the near future. This trend highlights the increasing need for bilateral and regional trade negotiations between potential member countries. The negotiations to obtain these are typically time-consuming and do not always succeed. Understanding the determinants of successfully concluded agreements can help to identify drivers and potential pitfalls for future trade agreements. The formation of trade agreements may be difficult due to a lack of trust and communication difficulties arising from ethnic or cultural differences between potential members.<sup>1</sup> Differences in cultural norms and expectations about the behavior of the other party can lead to misunderstandings and negatively affect negotiations.<sup>2</sup> Individuals from different cultural backgrounds differ in their level of trust, differ in how they act when confronting social dilemmas such as, e.g., prisoner's dilemmas or contributing to public goods, and

---

<sup>1</sup> Knack & Keefer (1997) find that countries which are ethnically more homogeneous have higher levels of trust.

<sup>2</sup> Zou et al. (2009) show that individuals' behavior depends on what they perceive to be the consensus or "common sense" view within their culture; for similar arguments see also Roth et al. (1991). Henrich (2000) and Henrich et al. (2001) show that behavior in the ultimatum game depends on the culture of the experiment subjects.

have differing degrees of willingness to punish others when they free-ride.<sup>3</sup> Establishing trust, escaping the prisoner's dilemma of strategic trade policy and how to deal with free riders are well-known key problems of international trade negotiations.<sup>4</sup> More specifically, negotiation and bargaining styles differ across countries, and cultural differences are more pronounced in bargaining settings.<sup>5</sup> Trade negotiations are particularly affected by cultural differences as they involve infrequent, high stakes interactions between often changing high-level politicians or bureaucrats where establishing trust and a common understanding may be difficult. Cultural differences may also reflect different preferences for policy outcomes in the countries' populations, making it harder for negotiators to reach a consensus and hence successfully conclude a trade agreement.

These cultural differences and associated costs are difficult to measure, particularly at a bilateral level between a large set of countries. We propose to use Spolaore and Wacziarg's (2009) genetic distance, a measure of how genetically related populations are in terms of their last common ancestor, as a readily available proxy for

---

<sup>3</sup>Buchan et al. (2002) find that Japanese experiment subjects have a lower level of trust than their American counterparts. Gächter et al. (2010) and Herrmann et al. (2008) find significant differences in the willingness to punish non-cooperative players in experiments in different cultural backgrounds. These are not isolated findings: Cross-cultural differences in behavior in trust games are corroborated in a meta-analysis by Johnson and Mislin (2011).

<sup>4</sup>Brander (1986) is probably the first one to characterize trade negotiations as an attempt to escape the prisoner's dilemma of unilateral strategic trade policy.

<sup>5</sup>Roth et al. (1991) find that while subjects in different countries exhibit similar behavior in experimental markets, individual bargaining behavior varies considerably across countries. Gelfand et al. (2015) find that strategies which lead to successful negotiations in the United States are detrimental in Egypt. For a literature survey on cultural differences and negotiations, see Gelfand et al. (2012).

communication and negotiation costs arising from differences in culture and norms as a determinant of trade agreements. Anthropologic studies have shown that genetic distance can help to identify common cultural groups, in addition to geographic distance and shared language, two measures of cultural difference routinely used in the trade literature.<sup>6</sup> Similarly, Desmet et al. (2011) find that genetic distance correlates well with measures of cultural distances based on survey responses.

Figure 1 Insert Here

We use an (unbalanced) panel of 133 countries and 43 years and a battery of control variables to examine the role of genetic distance in establishing RTAs across countries. Our results show that genetic distance has a significant, negative and economically meaningful influence on the probability of forming an RTA. Importantly, we find that genetic distance helps to predict RTA formation even after controlling for a battery of other proxies of cultural differences such as geographic distance, linguistic distance, religious distance, colonial past, and differences in legal systems typically used in the literature. Genetic distance seems to correlate with some aspect of cultural differences that is not captured by other measures of cultural differences that would otherwise be omitted when explaining RTA formation.

---

<sup>6</sup>For example, cross-cultural differences such as norms around kinship correlate with human genetic diversity, see Jones (2003). For a general introduction to the relationship between human genetic and cultural diversity, see Stone and Lurquin (2007).

We would like to stress that our results should not and cannot be construed as to imply that countries should not engage in trade negotiations with countries with which they have a larger genetic distance, nor do we argue for a biological determinism of trade policy. Instead, insofar as genetic distance proxies cultural differences, our results highlight the potential usefulness of heightened awareness of possible misunderstandings which may arise during trade negotiations due to cultural differences.

We contribute to the literature which has documented the effect of genetic distance on economic outcomes. The seminal contribution is Spolaore and Wacziarg (2009) who show that genetic distance between countries can explain cross-country differences in income per capita. The larger these cultural differences, proxied by genetic distance, the more difficult the diffusion and adaptation of the frontier technology. We complement this literature on the influence of genetic distance on economic growth by focusing on the influence of genetic distance on economic policy, particularly RTA formation.

Our interpretation of genetic distance as a proxy for certain unobserved barriers to economic integration such as cultural heterogeneity is in line with a broader literature which links ethnic diversity measured by genetic distance and cultural heterogeneity.<sup>7</sup>

Desmet et al. (2011) document that genetic distance allows better predictions of similarity of individuals' survey responses on cultural values than those resting on only geographic and linguistic information. Using the example of Yugoslavia, they find that

---

<sup>7</sup>Ahlerup and Olsson (2012) provide an overview of this literature; see also Ashraf and Galor (2013).

genetic distance as a proxy for cultural heterogeneity predicts the disintegration of culturally diverse nation states, i.e., a particularly deep form of economic integration. We find that genetic distance also affects economic integration through RTAs. RTAs can create aggregate welfare gains when signatory parties act cooperatively. To establish an RTA and reap its welfare gains, signatories must overcome differences in norms and preferences as well as coordinate differences in socio-economic policies.

Guiso et al. (2009) also use genetic distance as a proxy for cultural difference. They show that respondents in the Eurobarometer survey trust individuals less with whom they have a higher genetic distance. This lower trust at the individual level is correlated with lower trade and portfolio investment between countries. Bove and Gokmen (2018) replicate Spolaore and Wacziarg (2009) and show that the impact of genetic distance on income differences between countries is stable over time. Melitz and Toubal (2019) show that somatic distance as well as genetic distance correlate with trade flows when controlling for measures of bilateral trust between countries. Our study finds that genetic distance has a stable and significant impact on RTA formation over more than four decades. Davies and Guillin (2014) use genetic distance as a proxy for communication barriers and find that US outbound services FDI is correlated between countries with low genetic distance. Leblang (2010) does not find a significant effect of genetic distance on bilateral FDI and portfolio investment in a single cross-section of countries.<sup>8</sup> Finally,

---

<sup>8</sup> FDI data are often missing for many country pairs, restricting Leblang's (2010) analysis to 28 FDI-receiving countries. Our sample comprises more countries and over 40 years.

Chaudhry and Ikram (2015) find that long-run GDP growth is correlated between countries with lower genetic distance.

Our paper also contributes to the literature on the determinants of RTAs, see, e.g., Magee (2003), Baier and Bergstrand (2004), Chen and Joshi (2010), and Egger et al. (2011).<sup>9</sup> With the exception of Martin et al. (2012), who use genetic distance in one specification for a cross-sectional regression for the year 2000, none of these papers studies the impact of genetic distance. Using panel data, we can analyze the impact of genetic distance while controlling for time-varying country-specific unobserved drivers of trade policy.

The remainder of the paper is organized as follows: Section 2 describes our data. Section 3 describes our empirical strategy and results. Section 4 discusses several robustness checks. Section 5 provides a discussion of our main findings. Section 6 concludes.

## 2 Data

***RTA and genetic distance:*** Our dependent variable is  $RTA_{ijt}$ , a binary variable which takes the value 1 if there is a customs union or free trade agreement between two

---

<sup>9</sup> All the cited papers use probit models in their analysis. Besides probit models, a plethora of methods have been used to analyze the determinants of RTAs: Egger and Larch (2008) use spatial econometric probit models and Márquez-Ramos et al. (2011) use ordered probit models to explain the drivers of different levels of trade integration between countries. Kohl and Brouwer (2014) use a clustering algorithm to identify “natural” trade integration blocs and estimate the impact of determinants of these blocs using a probit model.



countries, and 0 otherwise.<sup>10</sup> We use a panel from 1970 to 2012, purely driven by data availability of the variables included in our regressions.

We use the genetic distance measure between populations of countries from Spolaore and Wacziarg (2018) who extend the data in Spolaore and Wacziarg (2009) to a larger sample of countries.<sup>11</sup> Genetic distance measures rely on the fact that during human evolution, random variations in the form of genes (so-called alleles) occur over time. Geneticists use the difference in the frequency of alleles to measure genetic distance between populations. It is important to stress that these measures only focus on random drift variation in genes, i.e., neutral variations which do not give any discernible advantage for evolutionary selection. Geneticists can use these variations to calculate the proximate time elapsed since two populations became separated and hence the number of genealogical steps one must take to reach the last common ancestor population. Spolaore and Wacziarg (2018) use  $F_{ST}$ , a measure of genetic distance, for 267 ethnic groups by Pemberton et al. (2013).  $F_{ST}$  is a normalized difference in allele frequencies in two populations: The larger  $F_{ST}$ , the more different the distribution of alleles, and hence the more generations one has to go back in time to reach the last common ancestor, and hence the larger the genetic distance. As countries typically are

---

<sup>10</sup> We use Mario Larch's Regional Trade Agreements Database from Egger and Larch (2008) in its updated version `rta_20170310.dta` which can be accessed at <http://www.ewf.uni-bayreuth.de/en/research/RTA-data/index.html>.

<sup>11</sup> The data are available at <https://sites.tufts.edu/enricospolaoire/category/personal-webpage/>. Spolaore and Wacziarg (2009) use the original genetic distance data from Cavalli-Sforza et al. (1994) which covers only 42 populations.

populated by multiple ethnic groups, Spolaore and Wacziarg (2018) combine the genetic distances with country-level ethnic data from Alesina et al. (2003) to measure the genetic distances between countries, weighted by the ethnic composition of countries' populations. When country  $i$  consists of  $K$  ethnic groups and country  $j$  consists of  $M$  ethnic groups, genetic distance between  $i$  and  $j$  is calculated as:

$$Genetic\ Distance_{ij} = \sum_{k=1}^K \sum_{l=1}^L (s_{ik} \times s_{jl} \times d_{kl}), \quad (1)$$

where  $s_{ik}$  is the share of ethnic group  $k$  in country  $i$ ,  $s_{jl}$  is the share of ethnic group  $l$  in country  $j$  and  $d_{kl}$  is the  $F_{ST}$  genetic distance between ethnic groups  $k$  and  $l$ . It can be interpreted as the expected genetic distance between two individuals picked at random from countries  $i$  and  $j$  and therefore is a measure of the average genetic distance between two countries.

We now discuss our control variables in five groups, in the order we introduce them in the regressions.

***Geographic distance and common border:*** Standard gravity-type regressors have been shown to be important drivers of RTA formation. Particularly, we include  $Ln(Geographic\ Distance)_{ij}$ , the log distance between country  $i$  and  $j$ , and an indicator variable  $Common\ Border_{ij}$  which is one if country  $i$  shares a common border with country  $j$ , and zero otherwise. Both variables are from Centre d'Études Prospectives et d'Informations Internationales (CEPII), see Mayer and Zignago (2011).

***Other cultural difference proxies:*** Spolaore and Wacziarg (2016a) show that genetic distance is correlated with measures of linguistic and religious distance between countries. One reason for this may be that genetic distance captures differences in language and religion due to differences in the composition of countries' populations which are not captured by simple country-pair dummy variables like common language typically used in empirical international trade. Melitz and Toubal (2014) show that sharing a common native language is associated with higher trust between populations and leads to higher trade between countries. We include both *Common Official Language*<sub>ij</sub>, a dummy which is one if the two countries share an official language, and zero otherwise, and *Common Native Language*<sub>ij</sub>, a dummy indicating whether at least nine percent of the populations of the two countries speak a common language. Both variables are from Mayer and Zignago (2011). We follow Spolaore and Wacziarg (2016a) and use yet another measure for the closeness of countries in terms of language: They construct *Ln(Linguistic Distance)*<sub>ij</sub>, a measure of (log) linguistic distance using classifications of languages into language trees which count the number of common nodes in such a language tree. For example, both French and Italian are part of the Indo-European - Italic - Romance - Italo-Western branch of languages, i.e., they share four common nodes. Similar to genetic distance, these linguistic distances are weighted with the respective population share of a language in a given country. We also include *Ln(Religious Distance)*<sub>ij</sub>, a similarly constructed measure for religious distance also provided by Spolaore and Wacziarg (2016a). For

example, Christianity, Islam, and Judaism can be classified as “Near-Eastern Monotheistic Religions”. These religious distance measures are then weighted according to the share of a religion within a given country.<sup>12</sup> To control for the effects of both linguistic and religious similarity, we control for both linguistic and religious distance.

***Economic and legal differences, past and present:*** Indicators of colonial history are routinely used in the literature on RTA formation. We use *Colonial Relationship* $p_{ij}$ , a dummy indicating whether country  $i$  and  $j$  ever were in a colonial relationship, and *Common Coloniser* $p_{ij}$ , a dummy indicating whether both countries ever had a common coloniser, both from Mayer and Zignago (2011).

Differences in legal origins of countries reduce the amount of trade between countries, see, e.g., Felbermayr and Toubal (2010). Trade agreements may therefore be particularly important for countries with different legal systems to overcome these additional trade costs, increasing the likelihood of an RTA. At the same time, trade negotiations may be particularly difficult between countries with different legal systems.<sup>13</sup> Baier and Bergstrand (2004) do not find evidence that common legal origin matters for RTA formation using a cross section of 54 countries in 1996, but this may be due to their

---

<sup>12</sup> For details on the calculation of these measures, see Spolaore and Wacziarg (2016a). They also show that genetic distance is correlated with a cultural difference measure based on question-specific distances from the World Valued Survey (WVS) for 98 questions. Contrary to genetic distance which is available for 180 countries, this measure is only available for 74 countries. To maintain a large sample, we do not include it in our regressions.

<sup>13</sup> During the stalled negotiations for a potential trade agreement between the European Union and the United States, a commonly repeated argument was that differences in legal philosophies in consumer protection law (precautionary principle in the EU versus risk assessment and cost-benefit principles in the US) made an agreement difficult to reach, see Bergkamp and Kogan (2013).

smaller sample size. It could be that our genetic distance measure picks up the variation of differences in legal systems and leads us to erroneously attribute their effect to genetic distance. We therefore use the classification of JuriGlobe and define a binary variable  $Common\ Legal\ Origin_{ij}$  which is one if countries  $i$  and  $j$  share the same origin of their legal systems, and zero otherwise.<sup>14</sup> Following Egger et al. (2011), we also include  $DIF\ GDP\ Per\ Capita_{ijt} = \ln(\text{abs}(GDP\ Per\ Capita_{it} - GDP\ Per\ Capita_{jt}))$ , the log absolute difference in GDP per capita to proxy endowment differences such as the difference in the capital-labor ratio which is highly correlated with GDP per capita. This measure controls for Heckscher-Ohlin-type arguments which may influence the formation of trade agreements between countries with different endowments. Population and GDP data are from the World Development Indicators from the World Bank.

***Differences in political systems and relations:*** Bergstrand et al. (2016) stress the importance of political factors for RTA formation. We therefore include the absolute difference in the political freedom between countries  $i$  and  $j$  at time  $t$ ,  $DIF\ Polity_{ijt} = \text{abs}(Polity_{it} - Polity_{jt})$ , using the political freedom index by Marshall et al. (2016). We use the other indices available in this database to construct similar regressors: We also include measures of the absolute difference in political regimes (using both democracy and autocracy scores,  $DIF\ Democracy_{ijt}$  and  $DIF\ Auto_{ijt}$ , respectively), the absolute

---

<sup>14</sup> Data are available at <https://juri-globe.ca/en/allcategories-en-gb/3350-category-en-gb/index-of-states-and-their-corresponding-legal-and-constitutional-systems>. Legal systems are categorized as either civil law, common law, Muslim law, customary, or a mixture of these categories. We treat mixed legal systems as a separate category.

difference in party competition in parliament ( $DIF\ Parcomp_{ijt}$ ), the absolute difference in regulation of political participation ( $DIF\ Parreg_{ijt}$ ), and the absolute difference in political competition in government ( $DIF\ Polcomp_{ijt}$ ).

A history of military conflicts can motivate countries to deepen trade integration between them, an argument particularly applied to the European integration process, see Martin et al. (2012). On the other hand, military conflicts lead to lower levels of trust between countries, negatively affecting trade, see Guiso et al. (2009). At the same time, countries with lower genetic distance have a higher likelihood to engage in wars as they share similar preferences and compete for similar rival goods, see Spolaore and Wacziarg (2016b), leading to a potential omitted variable bias. We measure the experience of conflict and military related events by three variables: The total duration of wars between the two countries after the end of World War II, measured in days, ( $War\ Duration_{ij}$ )<sup>15</sup>, the existence of a military alliance between the two countries ( $Military\ Alliance\ Relationship_{ij}$ ) from the Correlates of War project, see Gibler (2009) and Maoz et al. (2019), and the bilateral correlation in UN votes ( $UN\ Vote\ Correlation_{ij}$ ) from Voeten et al. (2009) as a measure of implicit political alliance.

---

<sup>15</sup>Data are from Kreutz (2010) and contain information about armed conflicts between 1946 and 2005. ( $War\ Duration_{ij} = (War\ End\ Date)_{ij} - (War\ Start\ Date)_{ij}$ ) is the number of days of war between country  $i$  and  $j$  after 1945. We focus on wars after World War II as it marks the beginning of the current international order and because we focus on RTA formation between 1970 and 2012.

### 3 Empirical specification and results

We follow Chen and Joshi (2010) and estimate a linear probability model of RTA formation. Linear probability models are preferable to limited dependent variable models as they are easier to interpret and do not suffer from downward biased coefficient estimates in the presence of uncorrelated unobserved heterogeneity, see Mood (2010). They also allow us to control for time-varying unobserved variables for each origin and destination country by including time-varying country-specific dummy variables. We estimate the following model using the Stata package *reghdfe* by Guimarães and Portugal (2010):

$$RTA_{ijt} = \beta_1 \ln(\text{Genetic Distance})_{ij} + \mathbf{x}'_{ijt} \boldsymbol{\beta} + \mu_{it} + \eta_{jt} + \varepsilon_{ijt} \quad (2)$$

$\mathbf{x}_{ijt}$  includes bilateral control variables which may be correlated with genetic distance.  $\mu_{jt}$  and  $\eta_{jt}$  represent country-year fixed effects that control for unobserved country-level determinants of RTAs that vary over time, effectively controlling for overall changes in countries' trade policy as well as country-specific business cycle effects which may trigger RTA negotiations.<sup>16</sup> The country-year fixed effects also control for the interdependence of trade policy decisions as a country's willingness to sign an RTA with another country depends on the number of RTAs it has already signed with other countries. Baier et al. (2014) measure this interdependence of trade policy using so-called "multilateral FTA

---

<sup>16</sup> Note that country-year fixed effects automatically control for year fixed effects, i.e., across-the-board differences in RTA formation across years which affect all countries in a similar way.

terms” which measure the number of RTAs country  $i$  has signed with another country  $k \neq j$  at time  $t$ . We capture these terms by the  $\mu_{it}$  and  $\eta_{jt}$  fixed effects.<sup>17</sup> We start our sample in 1970 to avoid a perfect separation problem due to insufficient variation in the data.<sup>18</sup> Given the dyadic nature of the data set, we expect correlation in the error term between all observations involving country  $i$  or  $j$ , as a country’s general attitude towards trade policy and RTAs and other country-specific unobserved factors may drive the overall willingness of a country to sign RTAs with all bilateral partners. This is corroborated by the large degree of correlation for a given exporter  $i$  and a given importer  $j$  of trade flows, see, e.g., Spolaore and Wacziarg (2009) and Egger and Tarlea (2015). As customary in the literature, in our dataset, every country pair appears twice so that  $RTA_{ijt} = RTA_{jit}$ ,  $\forall i, j$ , i.e.,  $RTA_{ijt}$  is not directional, and neither are the regressors we use. This differs from gravity models of trade flows, where exports from  $i$  to  $j$  are not necessarily identical to exports from  $j$  to  $i$ . If neither regressors nor dependent variable have directional variation, then  $\eta_i = \mu_j$ ,  $\forall i, j$ .<sup>19</sup> This immediately implies that  $\varepsilon_{ijt} = \varepsilon_{jit}$   $\forall i, j$ , i.e. the error term is (perfectly) correlated within each

---

<sup>17</sup> Baier et al. (2014) approximate these multilateral resistance terms by GDP-weighted averages of bilateral distances with trade partners. These terms also control for a country’s remoteness, i.e., for its average trade costs across all its trade partners, similar to the approximation proposed by Baier and Bergstrand (2009) in a trade gravity context. Our fixed effects control for these terms, circumventing the need to construct proxy indices.

<sup>18</sup> For earlier years, our regressors and country-year fixed effects perfectly separate the dependent variable, so maximum likelihood estimates of logit or probit models do not exist and using a linear probability model does not make sense. For a discussion of perfect separation, see, e.g., Mansournia et al. (2018).

<sup>19</sup> This is well-known in the gravity literature, see, e.g., Head and Mayer (2014), p. 140: In a bilateral gravity equation of symmetric bilateral trade flows regressed on symmetric trade cost measures, estimated importer and exporter dummies are identical. This also applies in our setting. Including origin and destination-specific dummies or country-specific dummies delivers numerically identical coefficients.



country-pair. The literature on RTA determinants has neglected this correlation so far and hence overstates the precision of estimated coefficients.<sup>20</sup> We use two-way clustered standard errors by Cameron et al. (2011) to take into account this particular structure of the data.<sup>21</sup> Note that two-way clustering is strictly more general than one-way clustering at the country-pair level. Using the latter would lead to too small standard errors in the presence of two-way clustering, see Cameron et al. (2011).

Table 1 Inserts Here

Table 1 reports the estimates of Equation (2). In column (1), we include (log) genetic distance as well as country-year fixed effects, but no controls. Genetic distance has a significant negative impact on RTA formation. As genetic distance is highly correlated with standard regressors used in the literature, we explore whether this result holds up. In column (2), we only include (log) geographic distance, and, confirming the literature, we find a significant negative effect of geographic distance on RTA formation. The effect is of similar magnitude as the effect of genetic distance. In column (3), we include both distance measures simultaneously. Both genetic and geographic distance have a

---

<sup>20</sup> Baier and Bergstrand (2004) discuss correlation of errors across countries within an RTA (e.g., across EU member countries) but do not consider the more general case of correlation of a given country's trade policy across all its potential partner countries we consider. The correlation within an RTA of Baier and Bergstrand (2004) is modelled on the value of the dependent variable, introducing endogeneity bias in the calculation of the standard errors. Our approach avoids this.

<sup>21</sup> The variance-covariance estimator by Cameron, Gelbach and Miller (2011) assumes  $E(\varepsilon_{ijgh}\varepsilon_{lmg'h'}|x_{ijgh}, x_{lmg'h'}) = 0$  unless  $g = g'$  or  $h = h'$  where  $ij$  and  $lm$  refer to two country pairs (i.e., observations in the data) where we now indicate explicitly the two groups (i.e., clusters), in our application the first and the second country in a country pair, by  $g$  and  $h$ . If  $g = g'$  or  $h = h'$ , i.e., within an origin or destination country,  $\varepsilon_{ij} = \varepsilon_{ji} \forall i, j$ , then  $E(\varepsilon_{ijgh}\varepsilon_{lmg'h'}|x_{ijgh}, x_{jig'h'}) = E(\varepsilon_{ijgh}\varepsilon_{ijgh}|x_{ijgh}, x_{ijgh})$ , and hence the estimator allows for arbitrary correlation between  $\varepsilon_{ijgh}$  and  $\varepsilon_{jig'h'}$ , including perfect correlation.

significant and negative impact on RTA formation: If genetic distance between two countries increases by one percent, the probability of an RTA between them decreases by 0.09 percentage points<sup>22</sup>, whereas the same increase in geographic distance decreases the probability of an RTA by 0.14 percentage points. Hence genetic distance has a dampening effect on RTA formation of about two thirds the magnitude of the effect of geographic distance. To gauge its economic significance, we calculate the elasticity of the probability of RTA formation with respect to genetic distance, evaluated at the mean of  $RTA_{ijt}$ . The mean of  $RTA_{ijt}$  across all years, i.e., the unconditional probability of a country-pair-year-combination having an RTA is 0.07, i.e., seven percent of all country-pairs across all years have an RTA. Then, the elasticity of RTA formation with respect to genetic distance,  $(\Delta y/y)/(\Delta x/x)$ , is -1.3, i.e., the probability of an RTA decreases by  $\Delta y/y = (0.09/100)/0.07=0.013$ , i.e., 1.3 percent if genetic distance increases by one percent ( $\Delta x/x$ ). This is an economically sizeable effect: Doubling the genetic distance of a country pair reduces the probability of an RTA by  $50 \times 1.3 = 65$  percent, or by about two thirds.

In column (4), we introduce the common border dummy as well as the set of other cultural difference proxies. Sharing a common border increases the probability of RTA formation. Interestingly, we find an increase in the probability of having an RTA for

---

<sup>22</sup> The dependent variable is in levels and the regressor is in logarithms, i.e., if genetic distance increases by one percent, the probability for an RTA increases by  $\beta_1/100$  units, i.e.,  $\frac{\beta_1}{100} \times 100 = \beta_1 = -0.090$  percentage points, see Wooldridge (2002), page 656.

countries with a higher language distance. This may appear surprising, but note that this effect is conditional on controlling for sharing a common native or official language, i.e., the coefficient measures the additional effect of linguistic distance. Sharing a common language has the expected positive effect, though it is not significant. Religious distance also does not have a significant effect on RTA formation. Importantly, the estimated effect of genetic distance remains stable and significant.

In column (5), we introduce the set of variables which indicates past and present economic and legal differences. Sharing a common coloniser significantly increases the probability of RTA formation, and so does a common legal origin. Differences in the economic development between countries as measured by their GDP per capita difference reduce the probability of RTA formation. The effect of genetic distance remains negative and significant.

Finally, in column (6), we introduce the set of variables which measure differences in political systems and relations. We find that particularly military alliances and similar UN voting behavior significantly increase the probability of RTA formation. We also find significant effects of the difference in countries' democracy scores as well as political competition on RTA formation. Importantly, genetic distance still has a significant and negative effect on RTA formation. In this strictest specification, its impact on RTA formation remains about half the magnitude of geographic distance.

Summing up, genetic distance reduces the probability of RTA formation in a large panel of countries, even when controlling for a wide variety of variables typically used in the

literature. It therefore seems to be a simple and readily available catch-all proxy for coordination costs arising from cultural differences which have a negative impact on RTA formation.<sup>23</sup>

Having established the effect of genetic distance in a panel setting, we turn to the question of whether the impact of genetic distance on RTA formation is stable over time. Geographic distance has been shown to have a stable negative impact on trade flows over time, see Disdier and Head (2008).<sup>24</sup> As trade flows and RTA formation are driven by common factors, it seems natural to explore whether genetic and geographic distance have a constant effect on RTA formation over time or whether there are trends in their effects. To do so, we estimate a series of cross-sectional regressions separately for every year  $t$  in our sample using the following specification:

$$RTA_{ij} = \beta_1 \ln(\text{Genetic Distance})_{ij} + \mathbf{x}'_{ij} \boldsymbol{\beta} + \mu_i + \eta_j + \varepsilon_{ij} \quad (3)$$

Table 2 Inserts Here

$\mu_i$  and  $\eta_j$  are country fixed effects for the paired countries. Table 2 presents the regression results for selected years using the set of regressors used in column (6) in Table 1. We use the same sample as in Table 1, but now restricted to individual years. The number of observations increases over time because some countries do not exist in

---

<sup>23</sup> In unreported regressions, we estimated the columns of Table 1 on the larger samples which are possible when not including all regressors. The effect of genetic distance remains very similar.

<sup>24</sup> The persistent negative effect of distance on bilateral trade flows has been referred to as the distance puzzle. It has spurred a large literature which tries to explain this fact, e.g., Lin and Sim (2012), Yotov (2012), and Larch et al. (2016). None of these papers investigates the impact of genetic distance over time on bilateral trade flows.

early years, such as former Soviet Union countries. Results are similar to our panel regressions: Genetic distance negatively affects RTA formation in all columns except for year 1970 in column (1), probably due to the relatively small number of RTAs in 1970, see Figure 1. For the remaining years the effect remains effectively constant. Geographic distance has a negative effect on RTA formation which increases for the years 2000, 2005, 2010, and 2012. The influence of our other measures of cultural differences are mostly not significant, in line with our panel results. Colonial status variables lose their significance over time, in line with the diminishing importance of colonial relationships for trade flows as documented by Head et al. (2010). This effect seems to spill over into RTA formation as well. Similar to the cultural differences proxies, the measures for the differences in political systems do not have consistently estimated effects over time. The difference in GDP per capita negatively influences RTA formation, particularly in later years. As in the panel results, geopolitical motives are key drivers behind RTA formation, as both military alliance and UN vote correlation have significant and positive effects for most years.

In addition to the years presented in Table 2, we estimate Equation (3) for all years in our sample beginning in 1970. We plot the estimated coefficients for each year for both (log) genetic and (log) geographic distance in Figure 2.

Figure 2 Inserts Here

Both geographic and genetic distance have a persistent and negative impact on RTA formation, with the mentioned exception of genetic distance in 1970. Interestingly, until

1990, we cannot reject the null hypothesis that the impact of genetic distance on RTA formation is as large as the impact of geographic distance. With the end of the Cold War, the negative impact of geographic distance becomes stronger, whereas the effect of genetic distance remains relatively constant. Overall, cultural differences and communication or negotiation costs proxied by genetic distance seem to act as a significant and economically important barrier to RTA formation. This effect appears to be stable over time.

#### **4 Robustness Checks**

In the following, we probe our results for robustness. We present results of these robustness checks for our panel regressions in Table 3. For convenience, column (1) reproduces column (6) of Table 1, our most stringent specification. In column (2), we present standardized beta coefficients (standard errors are those of the unstandardized coefficients). The coefficients confirm our previous interpretation that the effect of genetic distance is economically meaningful: If genetic distance increases by one standard deviation, the probability of RTA formation is reduced by 0.165 standard deviations, roughly half the size of the impact of a one standard deviation increase in geographic distance.

Our sample from 1970 to 2012 includes the end of the Cold War and the collapse of the Soviet Union (USSR). These events have significantly changed the geopolitical

environment in which trade agreement negotiations take place: Gowa and Mansfield (1993) argue that this shift from a bipolar to a multipolar world affects the formation (and dissolution) of trade agreements. This shift is also clearly visible in the number of country pairs with RTAs which has picked up after the dissolution of the USSR in 1991, see Figure 1. The countries which emerged from the former USSR increase our sample and results may be driven by these new countries. In column (3), we therefore rerun our regression from column (1) after excluding all former Soviet Union countries. Coefficient estimates hardly change, and genetic distance still has a significant dampening effect on RTA formation.

As is customary in the literature, our data set models the probability that a country pair signs an RTA, treating every country separately, including EU member countries. Trade policy and the conclusion of RTAs is an exclusive competence of the European Union, with the European Commission leading negotiations, not individual EU member countries. One could argue that this implies dropping country pairs involving individual EU member states and replacing them with country pairs between an aggregate EU and third countries. At the same time, EU trade policy making is complex: The European Parliament and the Council of the EU (or Council of Ministers before the Lisbon Treaty) have to adopt RTAs negotiated by the Commission. The Council consists of a representative of each member country and makes decisions concerning trade policy based on qualified majority voting, but some trade-related issues actually give a de facto veto power to individual member countries. This decision process has undergone several

reforms during the EU's history.<sup>25</sup> In practice, as evidenced by Brexit, even if de jure trade policy is an exclusive competence, EU decisions are shaped by participation constraints for the individual member states, as members who disagree with decisions can ultimately leave the EU. Still, we consider the EU as a single entity as a robustness check. For this, we have to create a weighted measure of genetic distance. We use the population share of an EU member state in a given year as the weight, i.e., we calculate genetic distance between the EU and a third country  $j$  in year  $t$  as:

$$Genetic\ Distance_{EUj,t} = \sum_{i=1}^{n_t} s_{i,t} \times Genetic\ Distance_{ij}$$

where  $n_t$  is the number of EU members in year  $t$  and  $s_{i,t}$  is the share of population of EU member country  $i$  in  $t$ :

$$s_{i,t} = \frac{Population_{i,t}}{\sum_{i=1}^{n_t} Population_{i,t}}$$

We also apply this weighting method to all other variables. We present results in column (4). We see that our main conclusions are not altered: The impact of genetic distance remains significant and of similar size. As expected, the impact of geographic distance is dampened, as the share of observations with RTAs decreases, and the individual EU countries all have  $RTA_{ijt} = 1$  and are all geographically close. So, our results are not driven by how we treat the decision-making process of EU member countries.

---

<sup>25</sup> For an overview of EU decision making and its history concerning trade policy issues, see chapter 12 in Baldwin and Wyplosz (2015).



Giuliano et al. (2014) argue that geographic features affecting transportation costs in the distant past have also led to separations of populations and thus affected genetic distance. Indeed, geographic distance highly correlates with genetic distance.<sup>26</sup> In their analysis of 22 European countries, they find that genetic distance does not exert a significant effect on trade flows once one controls for geographic distance.<sup>27</sup> Genetic distance is lower within Europe than in our worldwide sample of 133 countries, see Table A2 in the Appendix. In column (5), we use 20 European countries which are part of the sample of Giuliano et al. (2014) and for which we observe our regressors.<sup>28</sup> Our results partly agree with theirs: We do find a significant and negative effect of genetic distance on RTA formation between European countries but the effect is less than one tenth of the size of the effect in the worldwide sample. Hence, genetic distance affects RTA formation to a much lesser extent within Europe. Our results would indicate that the negative effect of genetic distance on RTA formation is nonlinear and increases with higher genetic distance.

---

<sup>26</sup> In our sample, the correlation between genetic distance and geographic distance in levels across all years is 0.410, and 0.514 in logarithms.

<sup>27</sup> Using a different specification, Melitz and Toubal (2019) do find that genetic distance matters even for trade flows between European countries.

<sup>28</sup> The countries included are Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Russia, Spain, Sweden, Switzerland, and United Kingdom.

## 5 Discussion

We now summarize the main take-aways from our analysis. We confirm previous findings in the literature that geographic distance has a dampening effect on RTA formation, whereas sharing a common border increases the probability of having an RTA. The effects are also large, so we confirm that geography is alive and well not only for trade flows, as established by the gravity literature, but also for trade policy. We also find that RTAs are somewhat more likely between countries with similar levels of economic development as measured by GDP per capita. Our results also confirm that geopolitical motivations are major drivers of trade integration, as military alliances and similar foreign policies as reflected by UN voting behavior all increase the probability of RTA formation. Also, countries with similar levels of democracy tend to sign RTAs with each other. This effect is not only driven by European integration but is observed worldwide. We also document the effects of cultural drivers of RTA formation. Importantly, genetic distance has a consistent negative impact on RTA formation, even after controlling for many other, typically used proxies for cultural differences. Its effect is stable over time and quite large: about half the size of the effect of geographic distance. Its effect is also consistent across many specifications, and it is particularly pronounced for RTA formation between countries with a large difference in genetic distance, larger than those differences within Europe.

Genetic distance seems to reflect some aspect of cultural differences and communication costs that is not easily captured by other proxies of cultural differences typically used in the literature. It therefore seems sensible to include it as a proxy variable in studies of RTA formation.

## **6 Conclusion**

Negotiations of trade agreements are often time-consuming and do not always lead to agreement. Understanding the determinants of successfully concluded regional trade agreements can help to identify drivers and potential pitfalls for future trade agreements. This paper examines the role of genetic distance between the populations of countries on RTA formation. Genetic distance measures how genetically related two populations are in terms of their last common ancestor. It is a readily available proxy for communication costs arising from differences in culture and preferences. Trade negotiations are particularly affected by these costs as they involve infrequent, high stakes interactions between often changing high-level politicians or bureaucrats from different cultural backgrounds where establishing trust and a common understanding may be difficult. We find that country pairs with larger genetic distances between their populations have a lower probability of signing an RTA. This effect is stable over time, and is distinct from the impact of geographic distance on RTA formation. It is robust to controlling for other determinants of RTA formation typically used in the literature

and holds across different subsamples. These results are consistent with a larger literature which documents the impact of cultural differences proxied by genetic distance on economic outcomes. They cannot and should not be interpreted as evidence for genetic determinism of trade policy. They also do not imply that countries should use genetic distance to make trade policy decisions. This would be a grave misreading of our results. Our analysis highlights the potential usefulness of genetic distance as a readily available proxy for difficult to measure bilateral communication and negotiation costs due to cultural differences across countries. It also documents the importance of cultural considerations in analyzing successful negotiations of trade agreements.

## References

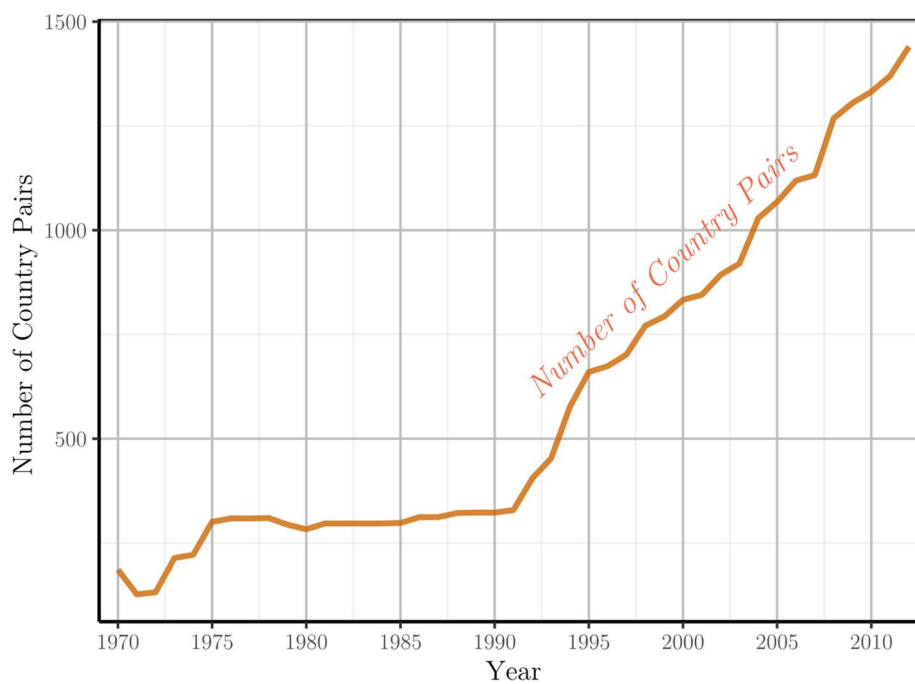
- Ahlerup, P., & Olsson, O. (2012). The roots of ethnic diversity. *Journal of Economic Growth*, 17(2), 71–102.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2), 155–194.
- Ashraf, Q., & Galor, O. (2013). Genetic diversity and the origins of cultural fragmentation. *American Economic Review*, 103(3), 528–533.
- Baier, S. L., & Bergstrand, J. H. (2004). Economic Determinants of Free Trade Agreements. *Journal of International Economics*, 64(1), 29–63.
- Baier, S. L., & Bergstrand, J. H. (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.
- Baier, S. L., Bergstrand, J. H., & Mariutto, R. (2014). Economic determinants of free trade agreements revisited: Distinguishing sources of interdependence. *Review of International Economics*, 22(1), 31–58.
- Baldwin, R., & Wyplosz, C. (2015). *The Economics of European Integration*. 5th edition. McGraw-Hill Education.

- Bergkamp, L., & Kogan, L. (2013). Trade, the Precautionary Principle, and Post-Modern Regulatory Process. *European Journal of Risk Regulation*, 4(November), 493–507.
- Bergstrand, J. H., Egger, P., & Larch, M. (2016). Economic determinants of the timing of preferential trade agreement formations and enlargements. *Economic Inquiry*, 54(1), 315–341.
- Bove, V., & Gokmen, G. (2018). Genetic distance, trade, and the diffusion of development. *Journal of Applied Econometrics*, (November 2017), 3–9.
- Brander, J. A. (1986). Rationales for Strategic Trade and Industrial Policy. In P. R. Krugman (Ed.), *Strategic Trade Policy and the New International Economics* (pp. 23–46). MIT Press.
- Buchan, N. R., Croson, R. T. A., & Dawes, R. M. (2002). Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange. *American Journal of Sociology*, 108(1), 168–206.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics*, 29(2), 238–249.
- Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press.
- Chaudhry, A., & Ikram, R. (2015). Does genetic proximity to high growth countries affect a country’s own growth? *Economic Modelling*, 51, 444–453.
- Chen, M. X., & Joshi, S. (2010). Third-country effects on the formation of free trade agreements. *Journal of International Economics*, 82(2), 238–248.
- Davies, R. B., & Guillin, A. (2014). How far away is an intangible? Services FDI and distance. *World Economy*, 37(12), 1731–1750.
- Desmet, K., Le Breton, M., Ortuño-Ortín, I., & Weber, S. (2011). The stability and breakup of nations: A quantitative analysis. *Journal of Economic Growth*, 16(3), 183–213.
- Disdier, A.-C., & Head, K. (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. *The Review of Economics and Statistics*, 90(1), 37–48.
- Egger, P., & Larch, M. (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics*, 76(2), 384–399.
- Egger, P., Larch, M., Staub, K. E., & Winkelmann, R. (2011). The Trade Effects of Endogenous Preferential Trade Agreements. *American Economic Journal: Economic Policy*, 3(3), 113–143.
- Egger, P., & Tarlea, F. (2015). Multi-way clustering estimation of standard errors in gravity models. *Economics Letters*, 134(C), 144–147.
- Felbermayr, G. J., & Toubal, F. (2010). Cultural proximity and trade. *European Economic Review*, 54(2), 279–293.
- Gächter, S., Herrmann, B., & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B*, 365(1553), 2651–2661.

- Gelfand, M. J., Severance, L., Fulmer, C. A., & Al-Dabbagh, M. (2012). Explaining and Predicting Cultural Differences in Negotiation. In G. Bolton & R. Croson (Eds.), *The Oxford Handbook of Economic and Conflict Resolution* (pp. 332–356). Oxford University Press.
- Gelfand, M. J., Severance, L., Lee, T., Bruss, C. B., Lun, J., Abdel-Latif, A.-H., et al. (2015). Culture and getting to yes: The linguistic signature of creative agreements in the United States and Egypt. *Journal of Organizational Behavior*, *36*, 967–989.
- Giuliano, P., Spilimbergo, A., & Tonon, G. (2014). Genetic distance, transportation costs, and trade. *Journal of Economic Geography*, *14*(1), 224–225.
- Gowa, J., & Mansfield, E. D. (1993). Power Politics and International Trade. *American Political Science Review*, *87*(2), 408–420.
- Guiso, L., Sapienza, P., & Zingales, L. (2009). Cultural Biases in Economic Exchange? *Quarterly Journal of Economics*, *124*(3), 1095–1131.
- Head, K., Mayer, T., & Ries, J. (2010). The Erosion of Colonial Trade Linkages after Independence. *Journal of International Economics*, *81*(1), 1–14.
- Henrich, J. (2000). Does Culture Matter in Economic Behaviour? Ultimatum Game Bargaining Among the Machiguenga of the Peruvian Amazon. *American Economic Review*, *90*(4), 973–979.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review Papers and Proceedings*, *91*(2), 73–78.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, *319*(5868), 1362–1367.
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889.
- Jones, D. (2003). Kinship and Deep History: Exploring Connections between Culture Areas, Genes, and Languages. *American Anthropologist*, *105*(3), 501–514.
- Knack, S., & Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, *112*(4), 1251–1288.
- Kohl, T., & Brouwer, A. E. (2014). The development of trade blocs in an era of globalization. *Environment and Planning A*, *46*(7), 1535–1553.
- Kreutz, J. (2010). How and when armed conflicts end: Introducing the UCDP conflict termination dataset. *Journal of Peace Research*, *47*(2), 243–250.
- Larch, M., Norbäck, P. J., Sirries, S., & Urban, D. M. (2016). Heterogeneous Firms, Globalisation and the Distance Puzzle. *World Economy*, *39*(9), 1307–1338.
- Leblang, D. (2010). Familiarity breeds investment: Diaspora networks and international investment. *American Political Science Review*, *104*(3), 584–600.
- Lin, F., & Sim, N. C. S. (2012). Death of distance and the distance puzzle. *Economics Letters*, *116*(2), 225–228.
- Magee, C. S. (2003). Endogenous Preferential Trade Agreements: An Empirical Analysis. *Contributions to Economic Analysis & Policy*, *2*(1), article 15.

- Mansournia, M. A., Geroldinger, A., Greenland, S., & Heinze, G. (2018). Separation in Logistic Regression: Causes, Consequences, and Control. *American Journal of Epidemiology*, 187(4), 864–870.
- Márquez-Ramos, L., Martínez-Zarzoso, I., & Suárez-Burguet, C. (2011). Determinants of Deep Integration: Examining Socio-political Factors. *Open Economies Review*, 22(3), 479–500.
- Marshall, M. G., Gurr, T. R., & Jagers, K. (2016). POLITY™ IV PROJECT. Political Regime Characteristics and Transitions, 1800-2016. Dataset Users' Manual. Version 2016 <p4v2016 and p4v2016d>. *Center for Systemic Peace Paper*.
- Martin, P., Mayer, T., & Thoenig, M. (2012). The geography of conflicts and regional trade agreements. *American Economic Journal: Macroeconomics*, 4(4), 1–35.
- Mayer, T., & Zignago, S. (2011). Notes on CEPII's Distances Measures: The GeoDist Database. *CEPII Working Paper 2011 - 25*.
- Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, 93(2), 351–363.
- Melitz, J., & Toubal, F. (2019). Somatic distance, trust and trade. *Review of International Economics*, 27(3), 786–802.
- Pemberton, T. J., DeGiorgio, M., & Rosenberg, N. A. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation. *G3: Genes, Genomes, Genetics*, 3(5), 891–907.
- Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review*, 81(5), 1068–1095.
- Spolaore, E., & Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2), 469–592.
- Spolaore, E., & Wacziarg, R. (2016a). Ancestry, Language and Culture. In *The Palgrave Handbook of Economics and Language* (pp. 174–211).
- Spolaore, E., & Wacziarg, R. (2016b). War and Relatedness. *Review of Economics and Statistics*, 98(5), 925–939.
- Spolaore, E., & Wacziarg, R. (2018). Ancestry and development: New evidence. *Journal of Applied Econometrics*, 33(5), 748–762.
- Stone, L., & Lurquin, P. F. (2007). *Genes, Culture, and Human Evolution. A Synthesis*. Blackwell Publishing.
- Wooldridge, J. M. (2002). *Introductory Econometrics. A Modern Approach. Second Edition*. South-Western.
- Yotov, Y. V. (2012). A simple solution to the distance puzzle in international trade. *Economics Letters*, 117(3), 794–798.
- Zou, X., Tam, K. P., Morris, M. W., Lee, S. L., Lau, I. Y. M., & Chiu, C. Y. (2009). Culture as Common Sense: Perceived Consensus Versus Personal Beliefs as Mechanisms of Cultural Influence. *Journal of Personality and Social Psychology*,

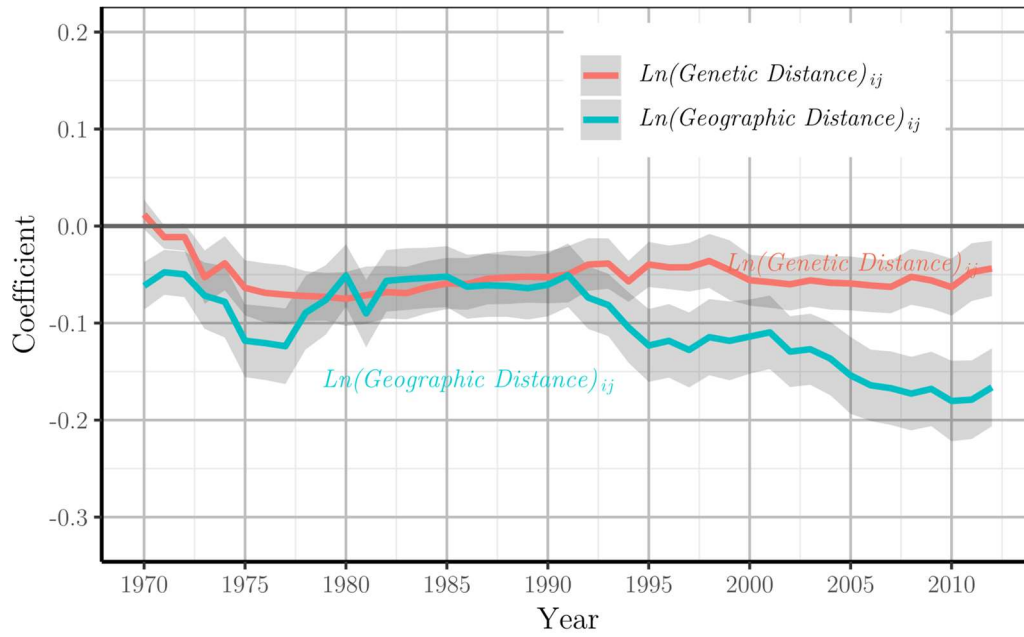
Figure 1. Total Number of Country Pairs with RTAs, 1970-2012



Note: Graph depicts the total number of distinct country pairs which are covered by an RTA (free trade agreement and/or customs union). Number of countries is  $N = 133$ , hence the total number of distinct country pairs is  $N \times (N - 1) / 2 = 8778$ .



Figure 2. Coefficients of  $\ln(\text{Genetic Distance})_{ij}$  and  $\ln(\text{Geographic Distance})_{ij}$  (1970-2012)



Note: Coefficients from cross-sectional OLS regressions based on Equation (3) for each year. Red and blue lines are the coefficients of genetic and geographic distance, respectively. The grey areas are the 95% confidence intervals for the estimated coefficients (1.96 times the standard error of the estimated coefficient). The samples for each year are the respective observations from the sample used in Table 1.

Table 1. OLS Panel Regression Coefficient Estimates (1970-2012)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Ln(Genetic Distance)<sub>ij</sub></i>	-0.160*** (0.010)		-0.090*** (0.009)	-0.088*** (0.009)	-0.073*** (0.010)	-0.055*** (0.010)
<i>Ln(Geographic Distance)<sub>ij</sub></i>		-0.186*** (0.015)	-0.140*** (0.013)	-0.136*** (0.015)	-0.136*** (0.015)	-0.112*** (0.015)
<i>Common Border<sub>ij</sub></i>				0.077*** (0.026)	0.070*** (0.025)	0.086*** (0.024)
<i>Common Native Language<sub>ij</sub></i>				0.040 (0.055)	0.053 (0.059)	0.032 (0.061)
<i>Common Official Language<sub>ij</sub></i>				0.004 (0.010)	-0.006 (0.014)	-0.003 (0.014)
<i>Ln(Linguistic Distance)<sub>ij</sub></i>				0.027* (0.016)	0.031* (0.017)	0.030* (0.016)
<i>Ln(Religious Distance)<sub>ij</sub></i>				0.018 (0.024)	0.039* (0.021)	0.059*** (0.019)
<i>Colonial Relationship<sub>ij</sub></i>					-0.015 (0.030)	-0.005 (0.032)
<i>Common Coloniser<sub>ij</sub></i>					0.035* (0.020)	0.026 (0.020)
<i>DIF GDP Per Capita<sub>ijt</sub></i>					-0.021*** (0.005)	-0.008** (0.003)
<i>Common Legal Origin<sub>ij</sub></i>					0.017** (0.008)	0.005 (0.008)
<i>DIF Democracy<sub>ijt</sub></i>						-0.017** (0.007)
<i>DIF Auto<sub>ijt</sub></i>						0.001 (0.007)
<i>DIF Polity<sub>ijt</sub></i>						0.006 (0.007)
<i>DIF Parreg<sub>ijt</sub></i>						0.001 (0.004)
<i>DIF Parcomp<sub>ijt</sub></i>						-0.008 (0.008)
<i>DIF Polcomp<sub>ijt</sub></i>						0.006** (0.003)
<i>UN Vote Correlation<sub>ijt</sub></i>						0.211*** (0.046)
<i>War Duration<sub>ijt</sub></i>						-0.000 (0.000)
<i>Millitary Alliance Relationship<sub>ijt</sub></i>						0.121*** (0.045)
<i>Country – Year Fixed Effects</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	511886	511886	511886	511886	511886	511886
<i>adj. R<sup>2</sup></i>	0.331	0.375	0.408	0.410	0.419	0.449

Note: Two-way cluster-robust standard errors clustered at the origin and destination country in parentheses. \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Regressions are based on Equation (2).

Table 2. Cross-section OLS Coefficient Estimates for Specific Years

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1970	1980	1990	2000	2005	2010	2012
<i>Ln(Genetic Distance)<sub>ij</sub></i>	0.012 (0.008)	-0.075*** (0.014)	-0.053*** (0.014)	-0.056*** (0.013)	-0.059*** (0.014)	-0.063*** (0.015)	-0.044*** (0.015)
<i>Ln(Geographic Distance)<sub>ij</sub></i>	-0.062*** (0.013)	-0.051*** (0.016)	-0.061*** (0.017)	-0.114*** (0.020)	-0.154*** (0.020)	-0.180*** (0.021)	-0.166*** (0.020)
<i>Common Border<sub>ij</sub></i>	0.089** (0.041)	-0.003 (0.038)	0.055 (0.038)	0.154*** (0.031)	0.085** (0.036)	0.081** (0.039)	0.106** (0.042)
<i>Common Native Language<sub>ij</sub></i>	0.039 (0.069)	-0.060 (0.070)	-0.086 (0.068)	0.128 (0.110)	0.156 (0.099)	0.213** (0.093)	0.149* (0.085)
<i>Common Official Language<sub>ij</sub></i>	0.092*** (0.033)	-0.023 (0.016)	-0.010 (0.017)	0.037* (0.020)	0.021 (0.019)	-0.002 (0.018)	-0.003 (0.015)
<i>Ln(Linguistic Distance)<sub>ij</sub></i>	0.028*** (0.008)	0.032*** (0.011)	-0.020 (0.020)	0.045 (0.037)	0.045 (0.032)	0.066** (0.026)	0.040** (0.016)
<i>Ln(Religious Distance)<sub>ij</sub></i>	0.080** (0.031)	0.015 (0.029)	0.025 (0.024)	0.071*** (0.024)	0.087*** (0.030)	0.071** (0.035)	0.046 (0.035)
<i>Colonial Relationship<sub>ij</sub></i>	0.103 (0.076)	-0.054** (0.026)	-0.043* (0.025)	-0.021 (0.051)	0.005 (0.046)	0.008 (0.041)	0.023 (0.051)
<i>Common Coloniser<sub>ij</sub></i>	-0.050* (0.029)	0.016 (0.019)	0.002 (0.017)	0.057* (0.034)	0.045 (0.032)	0.018 (0.029)	0.004 (0.028)
<i>DIF GDP Per Capita<sub>ij</sub></i>	-0.002 (0.003)	-0.002 (0.003)	-0.005 (0.004)	-0.009* (0.005)	-0.017*** (0.006)	-0.015** (0.006)	-0.017*** (0.006)
<i>Common Legal System<sub>ij</sub></i>	-0.017 (0.012)	-0.012 (0.016)	0.006 (0.011)	0.006 (0.010)	0.007 (0.012)	0.020 (0.013)	-0.008 (0.012)
<i>DIF Democracy<sub>ij</sub></i>	0.022** (0.009)	-0.000 (0.003)	-0.024*** (0.005)	-0.039*** (0.013)	-0.028** (0.013)	-0.027** (0.012)	-0.015 (0.012)
<i>DIF Auto<sub>ij</sub></i>	0.025*** (0.009)	0.013*** (0.003)	-0.008 (0.006)	-0.033*** (0.011)	-0.014 (0.013)	-0.009 (0.012)	0.006 (0.013)
<i>DIF Polity<sub>ij</sub></i>	-0.018** (0.008)	-0.005** (0.002)	0.014*** (0.004)	0.030*** (0.011)	0.022* (0.012)	0.015 (0.010)	0.004 (0.011)
<i>DIF Parreg<sub>ij</sub></i>	-0.010* (0.005)	0.006 (0.006)	-0.004 (0.004)	0.004 (0.006)	0.013* (0.007)	0.009 (0.007)	0.009 (0.007)
<i>DIF Parcomp<sub>ij</sub></i>	0.047*** (0.014)	-0.014 (0.012)	-0.025** (0.011)	-0.035** (0.015)	0.001 (0.012)	0.002 (0.014)	0.003 (0.016)
<i>DIF Polcomp<sub>ij</sub></i>	-0.022*** (0.006)	0.000 (0.005)	0.019*** (0.004)	0.021*** (0.005)	-0.004 (0.006)	0.000 (0.007)	-0.006 (0.007)
<i>UN Vote Correlation<sub>ij</sub></i>	0.079** (0.033)	0.128** (0.063)	0.233*** (0.066)	0.271*** (0.069)	0.236*** (0.071)	0.097 (0.065)	0.133* (0.076)
<i>War Duration<sub>ij</sub></i>	-0.000 (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
<i>Millitary Alliance Relationship<sub>ij</sub></i>	0.060** (0.030)	0.247*** (0.068)	0.188*** (0.066)	0.072 (0.048)	0.087* (0.050)	0.089* (0.053)	0.098* (0.051)
<i>Country Fixed Effects</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	6428	8328	10038	14910	16426	16944	15664
<i>adj. R<sup>2</sup></i>	0.262	0.466	0.449	0.466	0.489	0.457	0.451

Note: Two-way cluster-robust standard errors clustered at the origin and destination country in parentheses. \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Regressions are based on Equation (3). The samples for each year are the respective observations from the sample used in Table 1.

Table 3. OLS Panel Regression Coefficient Estimates (1970-2012): Robustness Checks

	(1)	(2)	(3)	(4)	(5)
	Full Sample	Standardized	No USSR	EU as one	EU20
$Ln(\text{Genetic Distance})_{ij}$	-0.055*** (0.010)	-0.165*** (0.010)	-0.069*** (0.010)	-0.064*** (0.010)	-0.004** (0.002)
$Ln(\text{Geographic Distance})_{ij}$	-0.112*** (0.015)	-0.306*** (0.015)	-0.116*** (0.016)	-0.068*** (0.011)	0.002 (0.009)
$\text{Common Border}_{ij}$	0.086*** (0.024)	0.046*** (0.024)	0.085*** (0.026)	0.148*** (0.024)	0.018 (0.019)
$\text{Common Native Language}_{ij}$	0.032 (0.062)	0.016 (0.062)	0.021 (0.060)	0.159** (0.076)	0.010 (0.048)
$\text{Common Official Language}_{ij}$	-0.003 (0.014)	-0.004 (0.014)	0.005 (0.013)	-0.022* (0.013)	0.007 (0.015)
$Ln(\text{Linguistic Distance})_{ij}$	0.030* (0.016)	0.038* (0.016)	0.026* (0.015)	0.040** (0.019)	0.007 (0.055)
$Ln(\text{Religious Distance})_{ij}$	0.059*** (0.019)	0.054*** (0.019)	0.063*** (0.019)	0.064*** (0.015)	-0.043** (0.018)
$\text{Colonial Relationship}_{ij}$	-0.005 (0.032)	-0.002 (0.032)	-0.055*** (0.019)	0.110 (0.103)	0.029** (0.012)
$\text{Common Coloniser}_{ij}$	0.026 (0.020)	0.024 (0.020)	-0.006 (0.014)	0.048** (0.019)	0.000 (0.000)
$\text{DIF GDP Per Capita}_{ijt}$	-0.008** (0.003)	-0.052** (0.003)	-0.003 (0.003)	-0.005** (0.002)	-0.002 (0.004)
$\text{Common Legal System}_{ij}$	0.005 (0.008)	0.008 (0.008)	0.004 (0.010)	0.016** (0.006)	-0.047* (0.024)
$\text{DIF Democracy}_{ijt}$	-0.017** (0.007)	-0.214** (0.007)	-0.011* (0.007)	-0.005 (0.004)	-0.030 (0.019)
$\text{DIF Auto}_{ijt}$	0.001 (0.007)	0.012 (0.007)	-0.000 (0.007)	0.002 (0.003)	0.048*** (0.011)
$\text{DIF Polity}_{ijt}$	0.006 (0.007)	0.144 (0.007)	0.004 (0.006)	0.001 (0.003)	0.000 (0.000)
$\text{DIF Parreg}_{ijt}$	0.001 (0.004)	0.005 (0.004)	0.005 (0.004)	0.002 (0.002)	0.001 (0.012)
$\text{DIF Parcomp}_{ijt}$	-0.008 (0.008)	-0.035 (0.008)	-0.000 (0.007)	0.001 (0.004)	0.011 (0.034)
$\text{DIF Polcomp}_{ijt}$	0.006** (0.003)	0.069** (0.003)	0.004 (0.003)	0.003* (0.002)	0.017 (0.017)
$\text{UN Vote Correlation}_{ijt}$	0.211*** (0.046)	0.218*** (0.046)	0.247*** (0.049)	0.050*** (0.012)	-0.132 (0.086)
$\text{War Duration}_{ijt}$	-0.000 (0.000)	-0.010 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
$\text{Military Alliance Relationship}_{ijt}$	0.121*** (0.045)	0.107*** (0.045)	0.112** (0.047)	0.171*** (0.050)	0.005 (0.019)
$\text{Country} - \text{Year Fixed Effects}$	Yes	Yes	Yes	Yes	Yes
$N$	511886	511886	441920	390508	11276
$\text{adj. } R^2$	0.449	0.449	0.481	0.346	0.823

Note: Two-way cluster-robust standard errors clustered at the origin and destination country in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions are based on Equation (3). Column (2) repeats the estimation of column (1) but shows standardized beta coefficients. Column (3) drops former USSR countries. Column (4) treats the EU as a single entity. Column (5) restricts the sample to 20 European countries.

## Appendix

Table A1. List of Countries

AGO	ALB	ARE	ARG	ARM	AUS	AUT	AZE	BDI	BEL	BEN	BFA	BGD	BGR
BHR	BLR	BOL	BRA	BTN	CAF	CAN	CHE	CHL	CHN	CMR	COD	COL	CRI
CUB	CYP	CZE	DEU	DJI	DNK	DOM	DZA	ECU	EGY	ERI	ESP	EST	FIN
FJI	FRA	GAB	GBR	GEO	GHA	GIN	GMB	GNB	GRC	GTM	GUY	HND	HRV
HTI	HUN	IDN	IND	IRL	IRN	IRQ	ISR	ITA	JAM	JOR	JPN	KAZ	KEN
KGZ	KHM	KOR	LAO	LBN	LBR	LBY	LKA	LTU	LVA	MAR	MDA	MDG	MEX
MLI	MOZ	MRT	MUS	MWI	MYS	NER	NGA	NIC	NLD	NOR	NPL	NZL	OMN
PAK	PAN	PER	PHL	POL	PRT	PRY	RUS	RWA	SAU	SDN	SEN	SGP	SLE
SLV	SVK	SVN	SWE	TCD	THA	TJK	TKM	TTO	TUN	TUR	UGA	UKR	URY
USA	UZB	VEN	VNM	ZAF	ZMB	ZWE							

Note: Table shows the ISO codes of the 133 countries included in the sample used in Table 1.

Table A2. Descriptive Statistics of the Different Samples

Variable	Mean	Std. Dev.	Min.	Max.
$Genetic\ Distance_{ij}$ (EU20)	0.00491	0.00418	0.00002	0.01150
$Ln(Genetic\ Distance)_{ij}$ (EU20)	-6.13605	1.61321	-10.71991	-4.46583
$Genetic\ Distance_{ij}$ (Total 133 countries)	0.03458	0.01720	0.00002	0.08723
$Ln(Genetic\ Distance)_{ij}$ (Total 133 countries)	-3.58075	0.84302	-10.71991	-2.43916

Note: (EU20) refers to the sample used for Column (5) in Table 3. (Total 133 countries) refers to the sample used in Table 1.