CESifo WORKING PAPERS

7051
2018

May 2018

Somatic Distance, Cultural Affinities, Trust and Trade

Jacques Melitz, Farid Toubal

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7051*
Category 8: Trade Policy

Somatic Distance, Cultural Affinities, Trust and Trade

Abstract

Somatic distance, or differences in physical appearance, proves to be extremely important in the gravity model of bilateral trade in conformity with results in other areas of economics and outside of it in the social sciences. This is also true quite independently of survey evidence about bilateral trust. These findings are obtained in a sample of the 15 members of the European Economic Association in 1996. Robustness tests also show that somatic distance has a more reliable influence on bilateral trade than the other cultural variables. The article finally discusses the interpretation and the breadth of application of these results.

JEL-Codes: F100, F400, Z100.

Keywords: somatic distance, cultural interactions, trust, language, bilateral trade.

*Jacques Melitz**
ENSAE & CEPII
Paris / France
j.melitz@hw.ac.uk

Farid Toubal
ENS de Paris-Saclay,
CREST & CEPII
Paris / France
ftoubal@ens-cachan.fr

*corresponding author

April 2018
We would like to thank Anca Cristea, Anthony Edo, Thierry Mayer and José de Sousa for valuable comments.

In sociology, homophily is taken as an established fact: individuals tend to associate and bond with similar others, as in the proverb "birds of a feather flock together." Also, "Homophily in race and ethnicity creates the strongest divides in our personal environments (McPherson et al 2001, p. 415)." The authors of this oft-cited article go on: "Age, religion, education, occupation, and gender [follow] in roughly that order." Indeed, discrimination based on race and ethnicity is readily apparent in economics too, especially in labor studies. (See Bertrand and Mullainathan (2004) and Edo *et al* (2017) concerning job applications; Lang and Lehmann (2012) and Borowczyk-Martins *et al* (2017) concerning wage earnings and employment.) Yet in studies of bilateral trade between countries, homophily is mostly absent. It emerged only recently in a highly influential article by Guiso, Sapienza and Zingales (2009) (hereafter GSZ). Since they wrote, two works have picked up on the theme: Spring and Grossmann (2014) (in a critical spirit) and Yu *et al* (2014). GSZ introduced somatic distance or difference in physical appearance on the basis of an Italian source: a work by Biasutti in four volumes, dating 1954 (first edition), which summarizes and extends a huge literature on racial differences in physical anthropology. However, while making use of Biasutti, GSZ and the two aforementioned studies, subordinate the whole issue of somatic distance by treating it as affecting bilateral trade strictly via trust: that is, as an instrument for trust in a 2SLS interpretation. In this contribution, we shall instead introduce somatic distance as a direct influence on bilateral trade.

In his path-breaking work on *The Economics of Discrimination* dating 1957, Becker modeled discrimination in the labor market as founded on the distaste of employers for interacting with certain groups of workers. His view has undergone subsequent modification, with allowances for the significance of beliefs about other people and some rational elements in labor market discrimination (for example, Phelps (1972) and Arrow (1973)). But as the literature in the preceding paragraph is witness, his basic position stands: racial prejudice as such retains a firm position in explanations of racial discrimination. It is clear that such prejudice partly stems from and partly breeds distrust. But there is no precedent for viewing the prejudice and, broadly, somatic distance in general, as impinging on international trade strictly through the filter of trust. No one would argue that sex discrimination is only a matter of trust. There should be little question either that discrimination based on phys-

ical appearance is not strictly a matter of trust. In the specific case of international trade, it is standard, since Armington (1969), to allow that national preferences for different trade partners may intervene in explaining bilateral trade. Should somatic distance be one of the reasons for these national preferences, personal affinities could well be the source, independently of trust. The mechanism could take several forms. For example, an increase in the number of foreign markets to which national firms export raises their fixed costs and necessitates choices. In making these choices, "animal spirits" may operate and thus explain why somatic distance enters. As another example, exporters of consumption goods might find that their wares have more appeal to foreigners who resemble them, partly because of similar tastes, but partly also because of a preference for associating with them in commerce.

Once we admit that somatic distance has a direct place in a gravity equation for bilateral trade, the variable emerges as highly significant. It remains so in the presence of other cultural factors, reflecting language, religion, law, and the history of wars, as well as sample evidence from questionnaires about trust. Indeed, it is more robust than the rest. Trust, based on questionnaire evidence, and the history of wars never matter. Since immigrants are particularly important in studying cultural interactions, we introduce them. This has a seriously damaging effect on two important cultural variables, same legal origin and common religion. Adding a population-weighted measure of physical distance in the presence of immigrants removes the significance of common native language too. Somatic distance is hardly affected throughout.

All these results occur in a European sample close to GSZ's. There are two strong reasons for sticking close to this sample. The more important is that we want to control for trust and we know no alternative to GSZ's measure. But secondly, even if we were to drop GSZ's trust variable, we could not extend the analysis very far, only to the rest of Europe outside the European Economic Area (EEA) in 1996, since Biasutti's data for somatic distance permits going no further. This would essentially add Eastern Europe, Iceland and Switzerland.

The next section offers the test evidence, the following one provides robustness tests concerning the significance of somatic distance, and the last one engages in general discussion and interpretation.

I. Tests and Evidence

The theoretical basis for the gravity model of international trade is now sufficiently well known to permit us to pass directly to the estimating equations. As mentioned, our sample size depends on GSZ's trust variable. They drew this variable from a number of Eurobarometer surveys of the trust of people in one country in natives of another in the then-current EEA. The exact question was: "I would like to ask you a question about how much trust you have in people from various countries. For each, please tell me whether you have a lot of trust, some trust, not very much trust, or no trust at all." The surveys took place in 8 separate years from 1970 to 1996 (there have been no further surveys since) and they cover 15 countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, and the United Kingdom). Of the 8 surveys, GSZ retained 7, and only 5 of the 15 countries appear in all 7, 4 of them only once. As a result, they could only draw 595 observations at most. We draw more, 690, because they include one variable limiting their sample that we shall neglect (Press coverage). GSZ are also careful to filter out country-specific fixed effects in the raw data so as to focus strictly on *bilateral* trust. They further eliminate variations in trust by calendar year affecting the entire sample. We shall imitate both steps.

As a start, the estimating equation is:

$$\text{Ln Exports}_{jit} = \alpha + \beta \text{ Ln Distance}_{ij} + \delta \text{ Border}_{ij} + \gamma \text{ Trust(Q)}_{ijt} + \eta_{it} + \lambda_{jt} + u_{ijt} \qquad (1)$$

Ln Exports_{jit} is the log of the exports of country *j* to country *i* in survey year *t*. Trust(Q)_{ijt} is the trust of country *i*, the importer, in country *j*, the exporter, in the year *t* based on questionnaire evidence (Q). η_{it} and λ_{jt} are importer-year and exporter-year fixed effects. u_{ijt} is the residual. For the bilateral exports data, we relied on UN COMTRADE. For distance, we followed GSZ in taking the distances between the two capitals. Common border is the usual 0-1 dummy variable. There are no zeros for the dependent variable. Eq. (1) is there to show the impact of Trust, based on the questionnaire evidence, on trade in the absence of any other cultural variables. It is the sole reflection of any cultural influences in the equation. All the relevant descriptive statistics are reported in the last table, Table 4.

As seen from the test of eq. (1) in column (1) of Table 1, Trust(Q) is totally insignificant. Distance

and common border are very significant, as generally true, but the coefficient of Distance is far below the usual value of one or over (in absolute terms) in the gravity model, and the coefficient of common border is unusually high relative to distance.

Next, we repeat the same test for Somatic Distance after substituting this distance for Trust(Q). As regards the measure of somatic distance (drawn for Biasutti 1954, vol.2), let us quote GSZ in full (GSZ 2008, p. 1107):

> As an alternative measure of distance [to the DNA measure] between two populations, we derive an index of somatic distance, based on the average frequency of specific traits in the indigenous population reported by Biasutti (1954). For height, hair color (pigmentation), and cephalic index (the ratio of the length and [to the] width of the skull). Biasutti (1954) draws a map of the prevailing traits in each country in Europe. For each trait, European Union countries fall into three different categories. For hair color we have "Blond prevails," "Mix of blond and dark," and "Dark prevails." We arbitrarily assign the score of 1 to the first, 2 to the second, and 3 to the third. When one country's somatic characteristics belong to more than one category, we take the country's most prevalent category. We then compute the somatic distance between two countries as the sum of the absolute value of the difference between each of these traits.

Column 2 shows the result of substituting Somatic Distance for Trust(Q). As we see, Somatic Distance is extremely important with the right sign and the estimates for physical Distance and common border are moderately lower than before (in absolute terms), but remain highly significant.

Following, we use a mix of Trust(Q) and Somatic distance, while adding a range of controls for other cultural influences besides the obvious one of common language. As regards these other controls, GSZ make a whole series of interesting suggestions. They introduce five variables, all of them possibly for the first time in the gravity literature: namely, first, same legal origins; second, the history of wars between countries going back to the year 1000; third, common religion; fourth, common linguistic roots based on the *Ethnologue* classification of language trees; and last, a second measure of genetic distance besides somatic distance that depends on DNA sequences (Cavalli-Sforza et al 1996). For Common language, they resort to Common official language. Their measure of Same legal origin comes from La Porta et al (1998), who distinguish between French, German, Scandinavian, and English origins. These first two are dummy variables. The history of wars will not detain us since GSZ dropped the variable early on because it proved insignificant (as has not always been true since) and we do too. Common religion comes from the *World Value Surveys* of the World Bank, which distinguish between Catholic, Protestant, Jewish, Hindu, Buddhist, Orthodox, no reli-

gion and other affiliation. Common religion is also a 0-1 indicator variable. Linguistic common roots rests on the Fearon-Laitin (2003) index based on language trees. We tried alternative measures for all these variables except DNA sequences before deciding to stick to GSZ's with one outstanding exception, concerning language, which we will justify.

Column 3 shows the results of adding the aforementioned cultural influences except for the two that GSZ drop early on: the history of wars, and genetic distances stemming from differences in DNA sequences. The reason why they drop DNA sequences, as well as the history of wars, is that the variable is completely dominated by Somatic distance when the two measures of genetic distance serve together (as we find as well). For expository reasons, we momentarily retain GSZ's measure of common language in column 3. As seen, the coefficient of Trust(Q) is still insignificant, as in column 1, but with the wrong negative sign. Common official language and Same legal origin both enter positively and very significantly at the 99 percent confidence level. Common religion does the same but only at the 90 percent confidence level. Linguistic common roots is totally insignificant. Finally, Somatic distance remains negative and significant at the 99 percent confidence level just as before in column 2.[1]

Column (4) focuses on language. At the time GSZ wrote, the only measure of common language in wide use rested on official status. Rose (2000) had recently pioneered this measure in applying the gravity model to worldwide evidence. Widely serviceable measures of common language based on native language and spoken language only came shortly after. Since they came (or concurrently), Melitz and Toubal (2014) have shown the superiority of both of these measures to official language in measuring a common language. The point bears special note at present. In the current sample of the EEA membership of 1996, for example, English is an official language strictly in the United

[1] GSZ obtain moderately better results for Trust(Q) in their OLS estimates than ours, but the differences are easy to explain. They stem from two sources: first, GSZ's addition of Press coverage, costing many observations (and perhaps secondarily their addition of Transportation costs); and second, their exclusion of somatic distance and common religion from their OLS estimates of bilateral trade in order to reserve both for use as instruments for trust(Q) in a subsequent 2SLS specification. If we replicate their OLS equations in the first three columns of their Table IV (GSZ 2009, pp. 1116-7), we get a somewhat higher coefficient (0.42 instead of 0.36) and higher standard error (0.21 instead of 0.17) for Trust(Q) than they do in their column 1, a slightly lower coefficient (0.27 instead of 0.29) and higher standard error (0.19 instead of 0.17) in their column 2, and a slightly lower coefficient (0.22 instead of 0.25) and the identical standard error (0.19) in their column 3. As thus clear, our estimates and theirs for Trust(Q) are quite close on their specification. In addition, if we introduce Press coverage (thereby losing nearly 100 observations) and transportation costs in our own specification, which contains both common religion and somatic distance, none of our conclusions are affected.

Kingdom and Ireland and therefore, based on official status, English is a means of communication strictly between these two. As another example, German is official strictly in Germany, Austria, and Switzerland; yet we know of the importance of German in the Netherlands, Denmark and the Scandinavian countries.

In light of these concerns, we simply add Common native language in column (4) to the rest of the variables in column (3) of Table 1. In order do so, we constructed a special measure of the variable so as to adapt it to the study period 1970-1996. Since most of the observations in this period come in the latter part, the new measure rests on the information about native languages in the 15 relevant countries in the 1988, 1996 and 2000 editions of *Ethnologue*. First, we averaged the percentages of native speakers of the relevant languages over those three years. Next, we calculated the sum of the *products* of the population shares country pair by country pair. (We needed to sum the products because a country pair could contain native speakers of both German and French in both countries, for example.)[2] The resulting measure refers to the probability that two people at random from two different countries in the sample would have the same native language.

As seen from column (4), Common native language enters significantly at the 95 percent confidence level. In its presence the significance of Common official language vanishes completely. Otherwise, there is little change except that Common religion becomes significant at the 99 percent level. In general, with regard to column (4), Common native language is the largest of the four significant cultural influences on trade. Based on standardized beta coefficients, its impact is 7 percent, those of Same legal origin and Somatic distance are 6 percent, and that of Common religion is 5 percent.

II. Robustness tests

For the rest, our econometric analysis centers on the robustness of Somatic distance, its robustness alone and as compared with the other cultural variables. Table 2 begins with a repeat of the last column of Table 1 without common official language. In column (2), we take advantage of supplementary information about somatic distance on GSZ's website (GSZ 2017). There, GSZ provide an interesting alternative measure. Whereas the current measure rests on height, hair color, and cephalic

[2] Admittedly, this procedure can lead to a problem of double-counting because of bilinguals and trilinguals, etc., but that is only an important concern for common spoken language not for common native language (see Melitz and Toubal 2014, p. 354).

index (HHC), the website offers the possibility of testing based on height and hair color (HH) alone. This obviously can shed light on the separate importance of the cephalic index. As we see in column 2, measuring somatic distance based on height and hair color alone makes no mentionable difference. We shall continue in the remainder of our tests to show results for Somatic distance resting on both HHC and HH alone.

Next, we introduce a particularly strong robustness test, to our minds: we admit immigrants. Not only does this variable regularly enter highly significantly in previous estimates of bilateral trade, but its presence tends to lower, blur or even eliminates the influence of other cultural variables. Our measure of Immigrants is the stock of people in country *j*, the exporter, who were born in country *i*, the importer (Özden *et al.* 2011). This variable (Immigrants) is obviously subject to simultaneity bias since exports from country *j* into country *i* may encourage emigration from *i* to *j*. The reverse influence of trade on Immigrants is thus also expected to be positive. Consequently, failure to correct for it (and we do not know how to do so), in principle, should lead to an exaggerated positive coefficient of the variable. All the other coefficient estimates in the equation could be affected.

With these caveats in mind, column (3) shows what happens when the variable enters. As ever, Immigrants is extremely important. In its presence, Common religion ceases to matter at all. The coefficient of Same legal origin drops too though the variable remains important somewhat below the 95 percent confidence level. The coefficient of Common native language drops as well, but the variable remains important at the 95 percent confidence level. Of all the cultural variables, only Somatic distance (HHC) is relatively unaffected. Its coefficient drops but it remains significant at the 99 percent confidence level. The last column of Table 2 substitutes Somatic distance (HH) for Somatic distance (HHC). Now HH behaves moderately more poorly than HHC. However, with HH present, Same legal origin becomes insignificant and Common religion remains so.

Table 3 takes up an important suggestion in Spring and Grossmann (2016, pp. 107-108). They observe that Head and Mayer (2002) had argued for the use of population-weighted distances in limited regional samples like ours on the ground that the other measures of distance tend to exaggerate the effects of Common border. We find, as Spring and Grossmann did, that use of the population-weighted measure gives more plausible results in the GSZ sample.

The first two columns of Table 3 begin by removing Immigrants in order to distinguish between the effect of the new measure of distance and that of Immigrants. Let us focus first on Distance and Common border, where the basic change lies. The coefficient of Common border in columns (1) and (2) (where the difference between the two figures is low and disregarded) goes down from its comparable level in Table 2, columns (1) and (2), as expected, but only moderately so. On the other hand, the coefficient of Distance rises substantially in absolute terms, from −0.42 (or −0.43) to −0.65 (or 0.66). This approach to −1 strikes us as good reason to favor population-weighted distance since −1 or higher (in absolute terms) is more typical in gravity tests. Viewing the cultural variables next, we find that the switch to population-weighted distance affects Common religion adversely. The variable now remains important at the 90 percent level whereas it was so before at the 95 percent level in the comparable columns (1) and (2) of Table 2. Common native language experiences a similar fate. Its coefficient drops, though it remains significant at the 95 percent confidence level. Only Same legal origin and Somatic distance are relatively unaffected. The bottom line is that all four cultural variables remain important with the substitution of population-weighted distance for distance between capitals.

The picture changes when population-weighted distance and Immigrants are mixed together, as we find in the last two columns of Table 3. Now Common native language becomes insignificant for the very first time. Common religion also returns to insignificance where it was before with Immigrants alone (columns (3) and (4) of Table 2). Same legal origin holds up relatively better. Its coefficient and significance drop as compared with columns (1) and (2) of Table 3 and the variable remains significant either at the 95 percent (column (3)) or the 90 percent confidence level (column (4)). Only Somatic distance remains basically unaffected. Its coefficient is essentially untouched and it is still significant at the 99 percent level in one case (HHC) and the 95 percent level in the other (HH). All in all, we are prone to attach importance to the moderately better performance of Somatic distance HHC than HH in Tables 2 and 3 and thus the presence of the cephalic index in measuring Somatic distance, partly because its absence notably raises the coefficients of Common border and Common native language (suggesting that these last two variables partly reflect the cephalic index in its absence).

As we look back on the entire set of results in Tables 1, 2, 3, Somatic distance behaves distinctly better than the other cultural variables. Apart from the outlying value of −0.13 when the variable served as the sole reflection of any cultural influences (Table 1, column 1), its coefficient varies from −0.10 to −0.07 from start to finish, and its significance falls below the 99 percent confidence level to the 95 percent one only when it is measured strictly on the basis of height and hair color. By contrast, all of the other cultural variables truly wilt at some point. Common language does so in the last two columns of Table 3. Same legal origin does so in the last column of Table 2 and it only remains significant at the 90 percent confidence level in column (3) of this table and in column (4) of Table 3. Common religion moves from good performance at first in Table 2 to poor performance in the last two columns of the table, and from acceptable performance in Table 3 to poor performance in the last two columns of the table.

III. Discussion

General discussion may begin with the implications of the insignificance of Trust(Q). On a strictly formal level, there are two possible interpretations. One is that trust has no bearing on bilateral trade. The other is that Trust(Q) is a poor estimate of trust in bilateral trade and therefore that trust remains embedded in the other cultural influences (as a group) in the estimates. There is little doubt that the second alternative is to be preferred. It would be difficult to deny the importance of trust in trade simply because of poor results in a sample of answers to questions about trust (subject to sample variance and demanding respondents to put up no stake whatever) in the teeth of all the previous historical evidence, going back many centuries, to the contrary. Nor can we easily overlook the theoretical and common sense grounds for this historical evidence to the contrary. Significantly, though, the answer has a drawback. It forces us to recognize that Somatic distance, however independent it may be of common language, common religion and common law in our results, most likely reflects an element of trust as well as cultural affinities. Yet removing this ambiguity would be a formidable task. For the time, we cannot do so but must simply point back to the earlier evidence we summoned of homophily and discrimination in labor markets.[3] This evidence greatly undermines any conceivable effort to pin the influence of Somatic distance in foreign trade entirely on rational calculations of

[3]As regards homophily, Fehr (2014) also provides related evidence of differences in altruism and cooperation depending on in-group status.

expected returns. In addition, even under laboratory conditions of contrived trust games, psychologists have great difficulty distinguishing effects of trust from ones of risk aversion and philanthropy (or a mix of personal ethic and self-esteem) (see, for example, Fehr 2014). Isolating the impact of trust in the uncontrolled environment of bilateral trade in economics would seem even more daunting. On the basis of the general state of knowledge, it makes good sense to ascribe the impact of physical differences between peoples on their choices of trade partners in our results largely to social preferences, prejudices and the like, quite apart from trust.

Next, what shall we make of our better results for somatic distance than all the other cultural variables? How much credence can we give to this outcome? Rather little, we think. The major reason is the limitation of our sample to the 15 European members of the EEA in 1996. The importance of this limitation shows up in a number of ways. One is the impact of substituting population-weighted distance for distance between national capitals. Much previous work shows that once a sufficiently wide world sample of countries serves, it makes no difference in gravity equations whether distances are measured based on capitals, central geographical locations, most populated cities, or otherwise. If population-weighted distances rarely serve in tests of *world* samples, as they seem to do, it is only because researchers know that the added sophistication would be pointless. We have checked this point in our 193-country sample in Melitz and Toubal (2014) by substituting population-weighted distance for distance between most-populated cities, our earlier choice. The differences in the coefficients and standard errors for distance and common borders are hardly worthy of discussion.

As a second indication of the importance of our limited sampling, past work with worldwide samples also tends to yield better results for common language than those here. In a study focusing on the proper measurement of common language, we found, in sharp contrast to the present, that common official languages, common native languages and differences in linguistic roots all emerge as simultaneously important (*ibid*.). True, somatic distance was absent in this earlier work, but all the other cultural variables (all of them besides trust) here present are there too.

Common religion and Same legal origin deserve a separate word. Right now, the only possible common religions are Catholic and Protestant. In a world sample, there would be room for other important shared religions such as Muslim, Buddhist and Orthodox. The results could be better or

worse. Similarly, GSZ's measure of Same legal origin, which distinguishes between French, German, Scandinavian, and English origins, would be difficult, if not impossible, to apply on a global scale. Any global measure of legal systems would need to rest on a different classification, for example, JuriGlobe's between civil law, common law, Muslim law, and mixed systems. Once again, the variable might behave differently. For all these reasons, we must beware of concluding that Somatic distance is the most reliable of the cultural influences.

Still, what about the importance of somatic distance as such? In our view, the variable's behavior in our tests provides powerful evidence in its favor. One might have thought that in a limited sampling of 15 European countries, like ours, the significance of the variable could not appear because of insufficient variance. But the results fly in the face of this prior. In fact, height and hair color alone do almost as well as the variable with a cephalic index too. Furthermore, with or without a cephalic index, Somatic distance outperforms the other cultural variables. Upon separate examination of the issue of variance as such, the three relevant European maps in Biasutti (vol. 2 of the 3d edition, 1959, insets between pages 40-41, 42-43, and 48-49) permit sorting each of the three elements of Somatic distance into three separate groupings, and based on GSZ's method of scoring, the coefficient of variance of Somatic distance is about 0.5 with HHC and .67 with HH alone. Those figures are high enough, it seems, to answer any puzzlement about our ability to discern the impact of the variable. Of course, with wider geographical sampling, there would be still more groupings of each element (that would even be so for all of Europe alone), which would mean more scope for Somatic distance to bear its influence and greater ease of detecting its impact. Once again, based on pure statistics, this could work in any direction. But empirically speaking, it would be surprising if effects of differences in physical appearance on trade that appear clearly in European evidence alone were to blur or disappear with added evidence from Africa, the Middle East and Asia. On the contrary, we would expect the importance of the variable to show up better with broader sampling.

There remains the vexing issue of the absence of skin color in the measure of somatic distance. Many people, we included, would expect this element of somatic distance to be, in fact, the most important. Yet there is no available index for it. A repair of this problem would be very welcome. Perhaps internationally comparable measures of skin color could rest on differences in melanin.

REFERENCES

Armington, Paul, 1969. "A theory of demand for products distinguished by place of production," *International Monetary Fund Staff Papers*, 16: 159-176.

Arrow, Kenneth (1973). "The theory of discrimination," in Orley Ashenfelter and Albert Rees, eds. *Discrimination in labor markets*, Princeton: Princeton University Press: 3-33.

Becker, Gary (1957). *The economics of discrimination*, Chicago: University of Chicago Press, first edition.

Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American Economic Review*, 94(4): 991-1013.

Biasutti, Renato (1954). *Le Razze e i Popoli Della Terra*, vol. 2, Turin : UTET; 3d. ed. 1959.

Borowczyk-Martins, Daniel, Jake Bradley, and Linas Tarasonis (2017). "Racial discrimination in the U.S. labor market: Employment and wage differentials by skill," *Labour Economics*, 49: 106-127.

Cavalli-Forza, Luca, Paolo Menozzi, and Alberto Piazza (1996). *The History and Geography of Human Genes*, Princeton: Princeton University Press.

Edo, Anthony, Nicolas Jacquemet, and Constantine Yannelis (2017). "Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications," *Review of Economics of the Household*: 1-28.

Ethnologue. https://www.ethnologue.com/

Fehr, Ernst (2009). "On the economics and biology of trust," *Journal of the European Economic Association* 7(2-3): 235-266.

Fearon, James and David Laitin (2003). "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 97(1): 75-90.

Guiso, Luigi, Paola Sapienza and Luigi Zingales (2009). "Cultural biases in economic exchange?" *Quarterly Journal of Economics* 74(3): 1095-1132.

_____ (2017). "Online appendix for cultural biases in economic exchange?" http://www.kellogg.northwestern.edu/faculty/sapienza/htm/onlineappendix.pdf

Head, Keith and Thierry Mayer (2002). "Illusory border effects: Distance mismeasurement inflates estimates of home bias in Trade," CEPII *Working Papers* no. 2002-01.

Lang, Kevin and Jee-Yeon Lehmann (2012). "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Literature*, 50(4): 959-1006.

La Porta, Rafael, Florencio López-de-Silanes, Andrei Schleifer and Robert Vishny (1998). "Law and finance," *Journal of Political Economy*, 106(6): 1113-1155.

McPherson, Miller, Lynn Smith-Lovin, and James Cook (2001). "Birds of a feather: Homophily in social networks,"*Annual Review of Sociology*, 27: 415-444

Melitz, Jacques and Farid Toubal (2014). "Native language, spoken language, translation and trade," *Journal of International Economics* 93(2): 351-363.

Özden, Çağlar, Christopher R. Parsons, Maurice Schiff and Terrie L. Walmsley (2011). "Where on Earth is Everybody? *World Bank Economic Review* 25(1): 12-56.

Phelps, Edmund (1972). "The statistical theory of racism and sexism," *American Economic Review*, 62(4): 659-61.

Rose, Andrew (2000). "One market one money: The effect of common currencies on international trade," *Economic Policy* 15: 7-33.

Spring, Eva, and Volker Grossman (2016). "Does bilateral trust across countries really affect international trade and factor mobility?" *Empirical Economics*, 50(1): 103-136.

Yu, Shu, Sjoerd Beugelsdijk, and Jakob de Haan (2015). "Trade, trust and the rule of law," *European Journal of Political Economy* 37: 102-115.

TABLES

Table 1. Baseline tests

	(1)	(2)	(3)	(4)
Distance (log)	-0.50***	-0.40***	-0.40***	-0.41***
	(0.09)	(0.09)	(0.08)	(0.08)
Common border	0.72***	0.63***	0.30***	0.28***
	(0.15)	(0.13)	(0.11)	(0.10)
Trust(Q)	0.30		-0.26	-0.21
	(0.20)		(0.18)	(0.18)
Common official language			0.48***	0.04
			(0.15)	(0.17)
Common native language				1.04**
				(0.45)
Same legal origin			0.31***	0.24**
			(0.09)	(0.10)
Common religion			0.26*	0.35**
			(0.14)	(0.14)
Linguistic common roots			0.33	0.28
			(0.27)	(0.27)
Somatic distance (HHC)		-0.13***	-0.10***	-0.09***
		(0.03)	(0.03)	(0.03)
	Yes	Yes	Yes	Yes
	Yes	Yes	Yes	Yes
Observations	690	690	690	690
Adj. R^2	0.947	0.962	0.963	0.963
Number of country pairs	207	207	207	207

Notes : All regressions contain exporter/year and importer/year fixed effects. Robust standard errors in parentheses have been adjusted for clustering by country pair. ***,** ,* different from 0 at the respective 1%, 5% and 10% significance levels.

Table 2. Robustness tests

	(1)	(2)	(3)	(4)
Distance (log)	-0.41***	-0.42***	-0.44***	-0.44***
	(0.08)	(0.08)	(0.07)	(0.07)
Common border	0.28***	0.34***	0.24***	0.29***
	(0.10)	(0.10)	(0.09)	(0.09)
Trust(Q)	-0.20	-0.21	-0.08	-0.08
	(0.17)	(0.19)	(0.13)	(0.14)
Common native language	1.11***	1.29***	0.71**	0.87***
	(0.31)	(0.34)	(0.30)	(0.33)
Same legal origin	0.24**	0.22**	0.16*	0.14
	(0.09)	(0.10)	(0.09)	(0.09)
Common religion	0.35**	0.35**	0.22	0.23
	(0.14)	(0.14)	(0.14)	(0.14)
Linguistic common roots	0.29	0.11	0.10	-0.05
	(0.27)	(0.29)	(0.25)	(0.27)
Somatic distance (HHC)	-0.09***		-0.079***	
	(0.03)		(0.026)	
Somatic distance (HH)		-0.10***		-0.087**
		(0.04)		(0.034)
Immigrants (stock in logs)			0.14***	0.14***
			(0.03)	(0.03)
Exporter × Year FE	Yes	Yes	Yes	Yes
Importer × Year FE	Yes	Yes	Yes	Yes
Observations	690	690	690	690
Adj. R^2	0.963	0.963	0.968	0.968
Number of country pairs	207	207	207	207

Notes : All regressions contain exporter/year and importer/year fixed effects. Robust standard errors in parentheses have been adjusted for clustering by country pair. ***,** ,* different from 0 at the respective 1%, 5% and 10% significance levels.

Table 3. Population-weighted distance

	(1)	(2)	(3)	(4)
Pop-Weighted Distance (log)	-0.65***	-0.66***	-0.62***	-0.64***
	(0.10)	(0.11)	(0.10)	(0.10)
Common border	0.24**	0.29***	0.22**	0.26***
	(0.10)	(0.09)	(0.09)	(0.09)
Trust(Q)	-0.26	-0.26	-0.15	-0.15
	(0.16)	(0.17)	(0.13)	(0.14)
Common native language	0.80**	0.93**	0.46	0.57
	(0.32)	(0.36)	(0.32)	(0.36)
Same legal origin	0.25***	0.23**	0.17**	0.16*
	(0.09)	(0.09)	(0.09)	(0.09)
Common religion	0.28*	0.28*	0.17	0.17
	(0.14)	(0.14)	(0.14)	(0.14)
Linguistic common roots	0.15	0.02	0.01	-0.10
	(0.26)	(0.28)	(0.25)	(0.26)
Somatic distance (HHC)	-0.07***		-0.07***	
	(0.03)		(0.02)	
Somatic distance (HH)		-0.08**		-0.07**
		(0.04)		(0.03)
Immigrants (stock in logs)			0.12***	0.12***
			(0.03)	(0.03)
Exporter × Year FE	Yes	Yes	Yes	Yes
Importer × Year FE	Yes	Yes	Yes	Yes
Observations	690	690	690	690
Adj. R^2	0.965	0.953	0.969	0.968
Number of country pairs	207	207	207	207

Notes : All regressions contain exporter/year and importer/year fixed effects. Robust standard errors in parentheses have been adjusted for clustering by country pair. ***,** ,* different from 0 at the respective 1%, 5% and 10% significance levels.

Table 4. Descriptive Statistics

	Dimension	Mean	Std. Dev.	Min	Max
Bilateral Trade	Logarithm	21.25	1.73	15.45	24.79
Trust(Q)	Real-non-neg.	2.73	0.29	2.01	3.65
Common official language	Binary	0.06	0.23	0.00	1.00
Common native language	Fractional	0.03	0.12	0.00	0.85
Distance	Logarithm	6.92	0.71	5.16	8.12
Common border	Binary	0.18	0.39	0.00	1.00
Same legal origin	Binary	0.35	0.48	0.00	1.00
Common religion	Fractional	0.30	0.26	0.00	0.87
Linguistic common roots	Binary	0.37	0.18	0.00	1.00
Somatic distance (HHC)	Real-non-neg.	2.46	1.22	0.00	5.00
Somatic distance (HH)	Real-non-neg.	1.64	1.11	0.00	4.00
Immigrants (stock)	Logarithm	9.14	2.07	3.74	13.38