I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11488

Long-Term Relatedness between Countries and International Migrant Selection

Tim Krieger
Laura Renner
Jens Ruhose

APRIL 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11488

Long-Term Relatedness between Countries and International Migrant Selection

Tim Krieger
University of Freiburg

Laura Renner
University of Freiburg

Jens Ruhose
Leibniz Universität Hannover and IZA

APRIL 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Long-Term Relatedness between Countries and International Migrant Selection*

This paper studies the effect of the long-term relatedness between countries, measured by their genetic distance, on educational migrant selection. Analyzing bilateral migrant stocks of the 15 main destination countries and 85 sending countries for the year 2000, we find that migrant selection and genetic distance follow a nonlinear J-shaped pattern: at low levels of genetic distance, increases in genetic distance reduce the positive selection of migration. However, at higher levels of genetic distance, this pattern is reversed and migration becomes more positively selected. We complement this finding by showing that the net benefits of genetic distance are strongly decreasing for low-skilled migrants with increasing genetic distance, while high-skilled migrants are less responsive to genetic distance in general. Results are robust to conditioning on bilateral control variables, including various destination- and sending-country-specific fixed effects and applying an instrumental-variables approach that exploits exogenous variation in genetic distances in the year 1500.

JEL Classification: F22, J61, Z1

Keywords: long-term relatedness, genetic distance, culture, international migration, selection

Corresponding author:
Jens Ruhose
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover
Germany
E-mail: ruhose@wipol.uni-hannover.de

* Forthcoming in the Journal of International Economics https://doi.org/10.1016/j.jinteco.2018.04.002. We are grateful to the editor Robert W. Staiger, two anonymous referees, David Dorn, Benjamin Elsner, Oliver Falck, Gabriel Felbermayr, Paola Giuliano, Volker Grossmann, Gordon Hanson, Ingrid Kubin, Volker Nitsch, Emily Oster, Jens Suedekum, and seminar and conference participants at the University of Freiburg, EGIT (Munich), ESPE (Braga), EEA (Toulouse), Verein für Socialpolitik (Hamburg), BevÖkA (Nuremberg), EALE (Ljubljana), the University of Jena and Dresden Technical University for their most helpful comments and discussions. We also thank Ingo Isphording and Sebastian Otten for sharing their language distance data with us. Earlier versions of the paper circulated under the titles "Culture, Selection, and International Migration" and "Genetic Distance and International Migrant Selection".

1 Introduction

In 2014, about 4.3 million people migrated permanently to OECD countries, which has increased the stock of the total foreign-born population in those countries to 120 million (OECD, 2016). International migration to these countries is dominated by migrants with higher skill levels as they are more likely to migrate than migrants with lower skills (Grogger and Hanson, 2011). High-skilled migrants are essential for economic development in destination countries that rely on innovation-driven economic growth (Coe and Helpman, 1995; Nelson and Phelps, 1966), while a disproportionate loss of high-skilled migrants could have detrimental development effects in the sending countries (Bhagwati and Hamada, 1974; Wilson, 2008).[1] Therefore, it is important to study and understand the determinants of international migrant selection.

Since Borjas (1987), a large literature has evolved that attempts to explain migrant selection by returns-to-skills differences between destination and sending countries.[2] However, the focus on differences in earning opportunities and a standard set of migration costs (e.g., geographical distance, existence of relatives abroad, household assets, and credit availability; cf. Kaestner and Malamud (2014); Patt et al. (2017)) as the main determinants of migration decisions and the resulting pattern of migrant selection appears too narrow.[3] It ignores that the costs and benefits of migration may also be affected by non-economic factors, such as cultural barriers to migration (Bauernschuster et al., 2014; Falck et al., 2012, 2017) or the benefits of living in a culturally different environment (Benson and O'Reilly, 2009a,b). Because policy makers are virtually unable to change these cultural factors, knowing to what extent and how they influence migrant selection is important for the design of labor-market and migration policies.

In this paper, we study how the *long-term relatedness* between countries, that is, their cultural distance or proximity, affects the selection of migrants. We measure migrant selection by using education-specific migrant stocks for the 15 main destination countries and 85 sending countries for the year 2000 (Docquier et al., 2007; Grogger and Hanson, 2011). The migrant skill mix that a destination country receives is positively (negatively) selected when the migrant stock is relatively more (less) skilled than the population in the sending country. To measure long-term relatedness, we use the genetic distance between two countries (Spolaore and Wacziarg, 2009, 2016b), which is based on the matrix of bilateral genetic distances between populations calculated by Cavalli-Sforza et al. (1994).

[1]However, recent research on 'beneficial brain drain' questions that this type of migration is necessarily harmful to the sending countries, see, e.g., Beine et al. (2008).

[2]For recent examples, see, e.g.: Belot and Hatton (2012); Chiquiar and Hanson (2005); Fernández-Huertas Moraga (2011); Grogger and Hanson (2011); Kaestner and Malamud (2014); Parey et al. (2017); Patt et al. (2017).

[3]For the year 2014, the OECD (2016) reports that only 14 percent of all migrants can be seen as labor migrants who are assumed to migrate for purely economic reasons. Most other migration happens for, e.g., family reasons, humanitarian reasons, and by accompanying families of workers.

Genetic distance measures how different the distribution of genes is between two populations. The calculation of genetic distance is only based on 'neutral' genes, which are considered as being 'neutral' because they change randomly. Those random changes in the distribution of genes take place regularly over time, which allows the interpretation of genetic distance as a molecular clock that measures the time span since two populations shared a common ancestor (Kimura, 1968). Studying the divergence in neutral genes between populations yields therefore a measure of the general relatedness of countries. Because cultural traits and habits are similarly transmitted across generations, genetic distance represents a summary statistic for a wide array of cultural traits transmitted intergenerationally (Spolaore and Wacziarg, 2009, 2016b). The view that genes and culture develop together over time is also present in the *dual inheritance theory* in social anthropology (Boyd and Richerson, 1985; Henrich and McElreath, 2003). In empirical work, several recent studies made use of genetic distance as a proxy for long-term relatedness or cultural distance between countries (see, e.g., Adserà and Pytliková, 2015; Desmet et al., 2011; Spolaore and Wacziarg, 2009, 2013, 2016b,c).

We contribute to the literature on the economic effects of long-term relatedness by documenting that the selection of migrants is related to genetic distance. In fact, our study documents a nonlinear J-shaped pattern between the bilateral migrant skill mix in the destination country and genetic distance: at low levels of genetic distance, increases in genetic distance slightly reduce the positive selection of migrants. At higher levels of genetic distance, however, this relationship is reversed and the migrant skill mix becomes much more positively selected when genetic distance increases. We complement this analysis by exploring the migration propensities for high- and low-skilled migrants separately. There we find that high-skilled migrants are generally less responsive to genetic distance than low-skilled migrants. At low levels of genetic distance, the results indicate that genetic distance attracts low-skilled migrants slightly more than high-skilled migrants. At high levels of genetic distance, the migration propensity is particularly low for low-skilled migrants. Thus, the findings suggest that the J-shaped pattern is mainly driven by strongly decreasing net benefits from cultural distance for low-skilled migrants (see Section 2 for a discussion of the benefits and costs of cultural distance).

While we find a strong nonlinearity regardless of the set of control variables, the J-shaped pattern between migrant selection and genetic distance appears only after the inclusion of measures for geographical distances and differences in aggregate income and returns to skills between the destination and the sending country. Because these covariates represent major alternative explanations for the pattern of migrant selection, this finding suggests that genetic distance is not an important determinant of the *aggregate* selection pattern when genetic distance is relatively low. However, the J-shaped pattern is robust to the inclusion of a large set of further covariates and an instrumental variables approach, which uses exogenous variation in genetic distance in the year 1500 to correct for a potential endogeneity bias that is induced by past migration waves and other omitted variables (Spolaore and Wacziarg, 2009). Because genetic distance remains a significant predictor of migrant selection throughout, we conclude that normally unobserved cultural traits, habits, and norms systematically affect migrant selection.

Importantly, even after controlling for a number of variables typically used to measure cultural differences (e.g., linguistic distance, common language, religion, and colonial history), we still find an independent and significant effect of genetic distance on migrant selection. We also verify whether the effect is mediated through skill-selective migration policies. For example, destination countries may welcome culturally distant populations only if they are sufficiently skilled. Controlling for standard and self-compiled measures of bilateral migration policies, we do not find that genetic distance is mediated through this channel. This does not rule out that the selection pattern is partly driven by informal and unobservable policies and local behavior against migrants from culturally distant populations. However, as long as these demand-side factors are the result of cultural distance, we still identify a total effect of genetic distance on migrant selection.

Contributing to the growing literature on the economic consequences of cultural traits and habits,[4] we provide the first study that links the literature on the determinants of migrant selection to the literature that examines the social and economic effects of the long-term relatedness between countries. While previous studies have found that differences in cultural traits are especially successful in explaining the size and the direction of economic exchange,[5] this study complements the literature by showing that long-term relatedness is a decisive factor for explaining migrant selection.

The remainder of paper is organized as follows. Section 2 lays out our theoretical considerations. Section 3 introduces the data, discusses genetic distance as a measure for the long-term relatedness between countries, and provides first descriptive evidence on the relationship between genetic distance and migrant selection. Section 4 explains the econometric setup and the identification strategy. In Section 5, we provide the results of our analysis. Section 6 conducts several extensions and robustness checks. Section 7 concludes.

2 Theoretical Considerations

The standard approach in migration economics postulates that utility-maximizing individuals decide to migrate to a different location if, and only if, the benefits of migration

[4]See, e.g., Ashraf and Galor (2013); Burchardi and Hassan (2013); Guiso et al. (2006); Ottaviano and Peri (2005); Spolaore and Wacziarg (2013); Tabellini (2010).

[5]For example, the literature contains studies that use long-term relatedness to explain income differences between countries (Spolaore and Wacziarg, 2009), migration flows (Belot and Ederveen, 2011; Dahl and Sorenson, 2010; Falck et al., 2012, 2017; Mayda, 2009), the diffusion of technology (Comin et al., 2012; Spolaore and Wacziarg, 2012), trade patterns (Guiso et al., 2009; Felbermayr and Toubal, 2010), and investment behavior (Guiso et al., 2009).

exceed migration costs (Borjas, 1987; Sjaastad, 1962). Benefits and migration costs differ between high- and low-skilled individuals and can be both monetary and non-monetary (Sjaastad, 1962; Schwartz, 1973). On the benefit side, the economics literature has studied extensively the role of economic factors, such as higher returns to skills, higher wage levels, and lower unemployment rates, as inputs in the migration decision (Borjas, 1987; Chiquiar and Hanson, 2005; Clark et al., 2007; Patt et al., 2017). Benefits of migration can also be non-monetary. For example, many cross-border moves have the purpose to reunite with family members. On the cost side, there is a wide range of monetary and non-monetary costs of migration (Falck et al., 2017); e.g., direct costs related to moving people and household goods abroad; indirect or opportunity costs through earnings forgone during the relocation process; and psychological costs resulting from the disutility associated with leaving behind one's family and social networks (Borjas, 1987; Schwartz, 1973; Sjaastad, 1962).

Migrant selection, that is, the ratio of high- to low-skilled migration, is a consequence of the skill dependence of (monetary and non-monetary) migration benefits and costs. Chiquiar and Hanson (2005), for example, show that if migration costs are decreasing with skills (or earnings), migrants may be positively or negatively selected, depending on the extent of migration costs and the shape of the skill distribution. We extend this standard selection framework by introducing (cultural) long-term relatedness between two countries as a determinant of migrant selection. In this context, long-term relatedness between countries can be understood as the cultural distance or proximity between the two populations living in these two countries (Spolaore and Wacziarg, 2016b,c).[6]

To assess the effect of long-term relatedness on migrant selection, it is important to discuss whether and how cultural distance can affect migration costs and benefits of high- and low-skilled migrants. We proceed by describing how cultural distance can enter the migrant decision for migrants at different skill levels. However, we have to keep in mind that what matters for migrant selection is the *relative* importance of cultural distance for high- and low-skilled migrants and not its *absolute* effect on either skill group. It is also important to mention that demand-side factors in the destination country can drive migration costs and benefits. For example, destination countries can affect migration benefits by adjusting net earnings in their countries through tax policies (Akcigit et al., 2016; Kleven et al., 2014). Destination countries can also affect migration costs by, e.g., imposing migration policies which usually favor specific skill groups (Ortega and Peri, 2013; Bertoli et al., 2011; Bertoli and Fernández-Huertas Moraga, 2015). In fact, if destination countries do not allow specific migrants to enter the country at all, one can argue that migration costs are prohibitively high. Because selective migration policies are very common in many countries, we put special emphasis on the role of these policies.

Cultural distance can increase migration costs because the less related two populations

[6] We discuss how to measure and interpret long-term relatedness between countries in the next section.

are, the more demanding it is to cross formal and informal barriers drawn by differences in cultural traits and habits between destination and sending countries (Belot and Ederveen, 2011; Falck et al., 2012).[7] That is, expected migration costs rise at the individual level if the destination-country population is perceived as very different from one's fellow citizens at home. It is also possible that destination countries design migration policies that favor migrants from countries that are culturally close, but impose higher barriers for migrants from countries that are not. These policies are potentially associated with substantial monetary and non-monetary costs for the migrant.

The effect of relatedness on migration costs may well differ by skill level (see, e.g., Bauernschuster et al., 2014, on internal migration). As low-skilled migrants have fewer capacities to cope with the differences between culturally unrelated countries, they ought to be less willing to move to culturally distant countries than their high-skilled counterparts. The latter may, for instance, be better at information gathering and processing, providing them with a larger set of possible destinations than low-skilled individuals. They may also have a greater ability to achieve bilingualism or find it easier to invest in destination-country-specific human capital. Also, migration policies potentially favor high-skilled individuals regardless where they come from and discriminate between migrants by cultural distance when they are low-skilled. While all these arguments clearly suggest that migration costs increase along with increasing cultural distance, they should increase at a much lower rate for high-skilled migrants than for low-skilled migrants. Holding migration benefits constant, we therefore expect the migrant skill mix to become more positively selected with increasing cultural distance.

Cultural distance may also increase the benefits of migration because some people have a desire to live in a different cultural environment, i.e., because of their pronounced intercultural interest or 'love of adventure' (Krieger and Lange, 2010). In that sense, cultural distance could yield some positive utility for the migrant and can also be characterized as 'lifestyle migration' (Benson and O'Reilly, 2009a, 2016) or 'amenity migration' (Gosnell and Abrams, 2011). The literature essentially argues that individuals value destinations also for their natural and cultural environments because these factors may improve quality of life. This view is also compatible with Rosen (1974) and Roback (1982) who argue that non-economic amenities, such as attractive scenery, a pleasant climate, clean air, and good schools, also enter the migrant decision.[8] These theoretical considerations indicate that a higher cultural distance acts as a (non-economic) benefit (or amenity) in the migration decision. In principle, destination countries can actively affect these benefits, for example, by advertising their own way of life in potential sending countries. In contrast to relatedness-induced migration costs, it seems likely that relatedness-induced migration

[7]Note that *negative migration costs* that are induced by cultural distance are categorized as migration benefits and that *negative migration benefits* from cultural distance are categorized as migration costs.

[8]These models are mainly used to explain internal migration and persistent earnings differentials in developed countries.

benefits do not strictly increase in line with cultural distance because migrants must be able to connect to the new cultural environment to receive a benefit. If the other culture is very different, this is unlikely and the cultural environment would become a burden.

The literature suggests that lifestyle or amenity migration is mainly undertaken by *relatively* affluent and privileged individuals who have enough assets and resources to afford to migrate for non-economic reasons (Benson and O'Reilly, 2009b, 2016; Hoey, 2005; Torkington, 2010). Better education is also associated with higher civic, political, and cultural participation and engagement (Oreopoulos and Salvanes, 2011), which suggests that the optimal cultural match receives a higher weight in the utility function of high-skilled individuals compared to low-skilled individuals. That does not imply, however, that low-skilled individuals do not migrate for non-economic reasons (Benson and O'Reilly, 2016). For example, because low-skilled individuals are disproportionally disadvantaged and vulnerable, they may seek other cultural environments where they feel culturally and socially more connected. However, we do not expect the corresponding benefits to extend beyond a certain threshold of cultural distance because research in the field of attitudes towards immigrants clearly suggest that low-educated individuals are themselves biased against (especially low-skilled) foreigners (Hainmueller and Hiscox, 2007; Mayda, 2006) and that this opposition is stronger against migrant populations that are ethnically more distinct (Dustmann and Preston, 2007). All these arguments imply that it is unlikely that low-skilled migrants receive (or are aware of) the benefit of moving to cultural environments that are very distant from to their cultural origin. By contrast, we expect high-skilled individuals who have a larger set of possible destinations to receive positive benefits also from cultural environments that are more distant. Holding migration costs constant, we therefore expect that the migrant skill mix should become more positively selected with increasing cultural distance.

The main message from this section is that there are theoretical reasons to predict that cultural distance has an effect on migrant selection. The discussion indicates that high-skilled migrants should dominate the migration flow in the presence of high cultural distances because it is likely that they experience lower relatedness-induced migration costs and relatively greater migration benefits than low-skilled migrants. However, the prediction is less clear-cut in the presence of lower levels of cultural distance. While relatedness-induced migration benefits and costs should also be low at those levels, the *relative* benefits and costs for high- and low-skilled migrants are hard to predict. Ultimately, the effect of cultural distance on migrant selection depends on whether high- or low-skilled migrants react more strongly to cultural distance. Therefore, we always analyze the migration behavior of both skill groups separately.

3 Measurement Concepts and Data

3.1 Genetic Distance and Long-Term Relatedness

To measure long-term relatedness between countries, we use the genetic distance between the countries' populations.[9] Following Spolaore and Wacziarg (2009, 2016b), genetic distance represents a summary statistic for a wide array of cultural traits transmitted intergenerationally. An important theoretical basis for using genetic distance as a proxy variable for differences in cultural traits comes from the *dual inheritance theory* in social anthropology. This theory points specifically to the parallels between genes and culture. Boyd and Richerson (1985) and Henrich and McElreath (2003) argue that culture is a system of inheritance, following evolutionary developments as genes do. In addition, geological and ecological barriers strengthen the differentiation between groups and therefore can affect genes and culture in the same way. Finally, cultural differences and genetic differences are mutually reinforcing. One example of this is that marriage mostly takes place within the same ethnic or religious group (Falck et al., 2012).

Genetic differences and cultural differences are similar in the sense that they are both transmitted from generation to generation and both change rather slowly. The longer two populations develop separately, the more time there is for them to develop in different directions and the greater the distance in genes and culture (Cavalli-Sforza et al., 1994). This does not assume that genes determine culture or that culture determines genes, but it does indicate important parallels in the development of genes and culture. Like genes, deeply rooted beliefs and behaviors (e.g., family structures), which are already imitated and learned from an early age, probably also change very slowly.[10]

In this paper, we use the *F_{ST} genetic distance* from Spolaore and Wacziarg (2009), who, in turn, refer to the seminal work by Cavalli-Sforza et al. (1994).[11] Cavalli-Sforza et al. (1994) assemble a matrix of bilateral genetic distances between populations, which they use to analyze the timing of the emergence of different populations across the world. Thus, intuitively, their measure is proportional to the time span since two populations have shared a common ancestor and therefore delivers information about the relatedness of populations (Spolaore and Wacziarg, 2009, p. 481). For the purpose of a cross-country analysis, Spolaore and Wacziarg (2009) assign the matrix of genetic distances between populations to countries within today's boundaries, while weighting populations according

[9] Online Appendix Table B-1 provides a detailed list of definitions and data sources for all variables. Krieger et al. (2018) provide data and replication files for this study.

[10] Several studies on the persistence of culture point to the existence of deeply-rooted beliefs, which change only very slowly (e.g., Alesina et al., 2013; Voigtländer and Voth, 2012).

[11] Details on the construction of this data can be found in Spolaore and Wacziarg (2009, 2016b) and in Krieger et al. (2015). We do not have information on genetic distance for the Czech Republic and therefore drop this country as a sending country from the analysis. We check the robustness of our results by using alternative genetic distance data from Pemberton et al. (2013) that rely on variation in human microsatellites (Spolaore and Wacziarg, 2016a). Results are shown in Online Appendix Table B-2 and are similar to the results obtained by using the data from Cavalli-Sforza et al. (1994).

to the ethnic composition in a country in the 1990s (data come from Alesina et al. (2003)). The F_{ST} genetic distance can take values between 0 (no genetic distance) and 1 (maximum genetic distance) and is multiplied by 10,000.

Panel A of Table 1 shows summary statistics for the genetic distance data. One standard deviation in genetic distance is represented by 572 points, with a mean of 716. Based on the genetic distance between the USA and Germany (352), one standard deviation indicates a shift in genetic distance similar to that between the USA and Mexico (904), the USA and Thailand (920), or the USA and Turkey (927).

Spolaore and Wacziarg (2009) also provide a F_{ST} genetic distance based on populations in 1500. Because populations in 1500 were close to the world populations used by Cavalli-Sforza et al. (1994), this limits measurement error in the assignment of genetic distances to populations because populations at that time were unaffected by later mass migration flows. In their analysis, Spolaore and Wacziarg (2009) propose the genetic distance based on populations in 1500 as an instrumental variable for genetic distance in the 1990s. In fact, both genetic distances are highly correlated ($r = 0.76$).

3.2 Selection of International Migrants

To examine the relationship between the selection of migrants and genetic distance, we use the bilateral migration data from Docquier et al. (2007) for the year 2000. Their data provides information on the number of migrants and residents by destination and sending countries and by education level (primary, secondary, and tertiary). As in Grogger and Hanson (2011), we restrict our analysis to the 15 main immigrant destination countries: Australia, Austria, Canada, Denmark, Finland, France, Germany, Ireland, the Netherlands, New Zealand, Norway, Spain, Sweden, the UK and the US. Due to data availability, the sample of sending countries is equal to 85.[12]

Starting from utility maximization and assuming that the error structure follows an i.i.d. extreme value distribution, it can be shown that the log odds of migrating to destination country d versus staying in sending country s is equal to the log of the share of the population of skill level $j \in \{H(igh), L(ow)\}$ from s that has migrated to d, that is E^j_{sd}, over the population with skill level j in s that remains in s, that is E^j_s (McFadden, 1974). Hence, $\ln\left(E^j_{sd}/E^j_s\right)$ gives the migration propensity from country s to country d (by skill level). In the following, we also refer to migration propensity as the *scale* of migration (Grogger and Hanson, 2011). Migrant selection is measured by migrant skill mix, that is, $\ln\left(E^H_{sd}/E^L_{sd}\right) - \ln\left(E^H_s/E^L_s\right)$. Migrants are positively selected when the share of migrants from country s is disproportionally high-skilled, that is, when the scale of high-skilled migrants is larger than the scale of low-skilled migrants, $\ln\left(E^H_{sd}/E^L_{sd}\right) - \ln\left(E^H_s/E^L_s\right) > 0$.

[12]See Appendix Table A-1 for the list of sending countries. We do not observe the migrant stock for 158 out of potentially 1,260 country pairs. Online Appendix Section C provides further results that use imputations of the migrant skill mix to provide evidence that these missings do not affect our conclusions.

Migrants are negatively selected if the reverse is true, i.e., $\ln\left(E_{sd}^{H}/E_{sd}^{L}\right) - \ln\left(E_{s}^{H}/E_{s}^{L}\right) < 0$.

Panel B of Table 1 provides summary statistics for emigration shares by skill level and for the migrant skill mix. The first row indicates that it is, on average, 81 percent more likely to see high-skilled emigration versus low-skilled emigration. In fact, in only 10.3 percent of the country pairs we observe that negative selection characterizes the migrant stock in the destination country. The average emigration share of the primary-educated population is equal to 0.003. That means that, on average, 0.3 percent of the sending country's low-skilled population lives abroad. That share is equal to 2.9 percent for the high-skilled population. Together, this indicates again the strong positive selection in international migration.

Figure 1 gives a first impression of the unconditional correlation between the scale of high- and low-skilled migration and genetic distance. To ease interpretation, we use non-parametric binned scatter plots. The relationship between genetic distance and the log odds of primary-educated emigration is clearly negative, indicating that, on average, a larger genetic distance is associated with lower migration of low-skilled individuals. If we find this relationship also after conditioning on confounding variables, this suggests that low-skilled migrants perceive genetic distance primarily as a migration cost instead of a migration benefit. The unconditional relationship between genetic distance and the log odds of tertiary-educated migrants does not indicate a significant relationship.

We can combine the measures for the scale of migration by skill level and use the difference between the two as a measure of the migrant skill mix that destination country d receives from sending country s. Figure 2 plots the relationship between the skill mix of migrants and genetic distance. As could be expected from Figure 1, the relationship is positive on average. A larger genetic distance is associated with a higher share of high-skilled migrants (relative to the high-skilled population in the sending country) compared to the share of low-skilled migrants (relative to the low-skilled population in the sending country). The positive relationship between genetic distance and migrant selection seems much stronger for country pairs with larger genetic distances compared to country pairs with smaller genetic distances. We focus further on this relationship in the multivariate analysis because the described unconditional correlation may not reveal a causal relationship if there are important confounding factors.

3.3 Other Variables Influencing Migrant Selection

We consider several covariates that could be associated with both migrant selection and genetic distance. Panel C of Table 1 documents summary statistics for all variables. First of all, geographical barriers between two countries ought to influence the flow of migrants as they increase transportation and adaptation costs. Geographical barriers could also be a reason for the observed genetic distance as populations developed along those barriers (for trade and genetic distance, see Giuliano et al., 2014).[13] For example, Schwartz (1973) argues that geographical distance is related to psychological migration costs and finds that increasing education diminishes the negative effect of geographical distance on migration. Therefore, our regressions account for the (log) geographical distance (in km) and whether the destination and the sending country share a common border (contiguity). The data comes from Head et al. (2010). To capture nonlinearities in geographical distance, we include country differences in absolute longitude, absolute latitude and average temperature and precipitation (data from Ashraf and Galor, 2013).[14]

We use wage data from Grogger and Hanson (2011) to capture the influence of skill premia, which is a key factor in explaining migrant selection (Borjas, 1987; Chiswick, 1999; Grogger and Hanson, 2011; Parey et al., 2017; Patt et al., 2017). Grogger and Hanson (2011) provide comparable wage measures for the 80th and 20th income percentile for each sending and destination country in our sample. We use the difference between the destination and the sending country in the 80th/20th wage ratio to proxy for monetary incentives of selective migration. Spolaore and Wacziarg (2009) show that income differences across countries converge in genetic proximity. Hence, another confounding factor may be the difference in GDP per capita and not the difference in the returns to skills. Therefore, we also control for differences in GDP per capita. The data come from Head et al. (2010).

The next variable concerns language barriers. Adserà and Pytliková (2015) show that the language distance between the sending and the destination country is a major obstacle for migration flows. Learning a new language may be easier for high-skilled people, thereby affecting migrant selectivity. However, language differences can also partly be determined by cultural differences. Hence, language differences could mediate the effect of genetic distance on migrant selection. Using the Levensthein language distance (data from Isphording and Otten, 2013), which is conceptually closely related to genetic distance, we test whether the effect of genetic distance is purely due to language differences. The measure compares the pronunciation of a set of words with the same meaning across languages and can be understood as the number of cognates, that is common roots, between two languages. The final Levensthein language distance is achieved by averaging over the set of words and gives a percentage measure of dissimilarity.[15] The closer the languages of sending and destination country, the smaller the Levensthein distance. The smallest language distance in our sample is between Finland and Estonia, while Denmark and Jordan have the maximum value.

[13]The correlation matrix in Online Appendix Table B-3 shows that the correlation between genetic distance and log geographical distance is $r = 0.43$.

[14]For robustness, we also run regressions by including geographical distance linearly and introducing geographical distance squared and cubic (see Online Appendix Table B-4). The results are very similar.

[15]When languages do not even have random similarities the value can be above 100 percent, e.g., Vietnamese to English (104,06).

In a robustness check, we also use an alternative language distance by Spolaore and Wacziarg (2016b) that follows the cladistic approach. Every language is assigned to several language families, which are further broken down into smaller sub-categories. By this means, one can construct language family trees and count the number of common nodes along the language tree for every language pair. Two countries that share no common nodes have the maximum cladistic language distance (in our sample, Armenia-Ireland with 0.87). We use the cladistic language distance that weights languages according to their share in the respective country. For example, Italian and French share four nodes ("Indo-European—Italic—Romance—Italo-Western") and have a weighted cladistic language distance of 0.21 (Spolaore and Wacziarg, 2016b). The Levensthein language distance and the cladistic language distance are highly correlated with $r = 0.78$.

Furthermore, because English is widely taught in many schools in different countries, we introduce a dummy for anglophone destination countries. Finally, we control for a shared official language, which is the case when at least 9 percent of the population speak the same language (data from Head et al., 2010).[16]

Another important factor for migrant selection is the presence of a diaspora or migrant network in the destination country. Existing networks increase information access and can reduce migration costs. This reduction in migration costs results in increased migration flows with relatively low average education levels from sending countries with larger migrant networks in destination countries compared to sending countries with smaller networks (Beine et al., 2011). The calculation of migrant networks follows Belot and Hatton (2012) who calculate the share of all migrants (coming from all education levels) from a sending country in the destination country relative to all residents in the sending country, $\sum_j E^j_{sd} / \sum_j E^j_s$. To avoid a correlation by construction with the dependent variable, we use data on migrants and residents in the year 1990 for the calculations. However, arguably, like bilateral language differences, migrant networks are potentially a product of cultural distances. Thus, introducing migrant networks to the model is potentially endogenous and explains (some of) the effect of genetic distance on migrant selection.

As outlined in Section 2, political and legal barriers as well as the general (non-)openness of a destination country could contribute to migration costs, which may be easier to bear for high-skilled migrants. In particular, visa restrictions and citizenship regulations are an important way to control immigration for countries such as the United States, Canada, and Australia. To test whether these demand-side regulations mediate the effect of genetic distance on migrant selection, we compile measures of privileged access to work visas and privileged access to citizenship at the country-pair level (see Section 6.1 for the construction of the variables). However, migration policies could also be a major confounding factor. Therefore, throughout the analysis, we control for travel

[16]Online Appendix Table B-3 shows that the Levensthein language distance and the existence of a common language are positively correlated with genetic distance.

visa restrictions by using a dummy, which is one if the destination country has imposed a travel visa restriction on the sending country and zero otherwise (data from Neumayer, 2006). We also use dummies for country pairs that are signatories to the Schengen agreement and for country pairs that were in a colonial relationship (data from Head et al., 2010). To measure the general openness of a country toward immigration, we include the log of the aggregate inflow of foreigners and the log number of asylum-seekers into the country, both retrieved from the International Migration Dataset of the OECD.

Measures of the general country skill level complete our baseline model because countries with a more similar skill mix could be more likely to interact with each other. Thus, we use the difference between the destination country and the sending country in terms of years of schooling and the share of people who have completed tertiary education, both taken from Barro and Lee (2013).[17]

4 Econometric Setup

4.1 Estimation

The aim of this study is to explain the migrant skill mix in destination country d from sending country s through variations in genetic distance. The theoretical considerations in Section 2 suggest that the effect of genetic distance on the migrant skill mix could be nonlinear. We allow for this possibility by including genetic distance linearly (GD_{sd}) and genetic distance squared (GD^2_{sd}) in Equation (1).[18]

$$\ln \frac{E^H_{sd}}{E^L_{sd}} - \ln \frac{E^H_s}{E^L_s} = \beta_0 + \beta_1 GD_{sd} + \beta_2 GD^2_{sd} + \mathbf{X}'_{sd}\phi + \epsilon_{sd} \tag{1}$$

We stepwise include a vector of control variables, $\mathbf{X}$, explained in Section 3.3. The error term ϵ_{sd} in Equation (1) is clustered at the destination country level to allow for arbitrary correlation within destination countries.[19]

The coefficients of interest in Equation (1) are the coefficients on genetic distance, β_1 and β_2, respectively. From our theoretical considerations, we clearly predict that the migrant skill mix should become more selected when genetic distance is high. Therefore, the prediction is that β_2 is positive. However, because both benefits and costs are likely to be small at low levels of genetic distance for all skill groups, the prediction of the sign of β_1 is not clear-cut. The coefficient is positive if the net benefits increase more (or decrease less) for high-educated migrants than for low-educated migrants with increasing

[17] Including these measures mean that we have to drop 13 sending countries from the analysis. However, because the omitted countries are not important sending countries, the results are unchanged when including them in models without the education variables.

[18] Grogger and Hanson (2011) derive the basic equation without the genetic distance terms formally based on individual utility maximization.

[19] Clustering at the destination-by-sending country level or using two-way clustering (Cameron et al., 2011) at the destination and the sending country level does not affect the results.

genetic distance. This would mean that migration becomes more and more positively selected with increasing genetic distance. In contrast, the coefficient is negative if the net benefits increase more (or decrease less) for low-educated migrants than for high-educated migrants with increasing genetic distance. This would mean that migration becomes less positively selected at lower levels of genetic distance, but more positively selected at higher levels of genetic distance.

The coefficients reveal the causal effect of genetic distance on the migrant skill mix if and only if genetic distance is not correlated with the error term. In the next section, we discuss when this identifying assumption fails to hold and lay out the identification strategy.

4.2 Identification

In this section, we explain the major threats to our identification strategy and how we approach them. First of all, reverse causality may be a problem if the migrant flow leads to the current genetic distance. We can exclude simple reverse causality because we use genetic distance in 1990 to explain migrant selection in 2000. In addition, as sampled populations tend to be indigenous, reverse causality is unlikely because very large migration waves are necessary to change the genetic distance between countries (Cavalli-Sforza et al., 1994).

Second, the measurement of genetic distance can be more or less precise for different countries. For example, genetic distance should be measured more accurately when the genetic variation within both countries is lower. The measurement error that is introduced through the imprecise measurement of genetic distance causes a bias in the coefficients on genetic distance toward zero (Spolaore and Wacziarg, 2016c).[20]

Third, and most problematically, there may be omitted variables that are correlated with both genetic distance and migrant skill mix and therefore bias the coefficients on genetic distance. Because genetic distance is a rather fundamental concept, it is important to distinguish between factors that are mediating channels—and whose omission does not bias the causal (reduced-form) effect of genetic distance on migrant selection—and factors that are causing an omitted variable bias. Bilateral migration policies, for example, could be a mediating factor. If genetic distance causes migration policies and migration policies cause migrant selection, then we can say that the effect of genetic distance is mediated by migration policies. Including both migration policy variables and genetic distance into the model, the model would ascribe some of the variation of the migrant skill mix to the variation of the migration policy variables. By contrast, omitted variables that are only related to genetic distance and the migrant skill mix may bias the observed relationship.

It is not always clear whether a variable represents a mediating factor or (if not

[20]Of course, this reasoning only applies if the measurement error is classical, i.e., if it is not correlated with genetic distance and the migrant skill mix.

included) an omitted variable, respectively. For example, on the one hand, language differences may be an important mediating channel as genetic distance could have shaped language differences as well. On the other hand, however, language differences, if they cause genetic distances and are not included, could also be an omitted variable. Bilateral migration policies are another example. Many countries discriminate against migrants by skill level while they are more willing to accept low-skilled migrants from, e.g., neighboring countries and from countries that share similar values.[21] As long as these policies are a product of lower genetic distance, we regard this channel as a mediating factor. However, it could be that migration policies correlate for other reasons with the genetic distance and with the migrant skill mix. In this case, migration policies would be an omitted variable.

While we cannot exactly distinguish between confounding and mediating factors, the objective of the empirical approach is to test whether genetic distance provides useful information about usually unobserved cultural traits and habits over and above what can be measured with other covariates that are related to cultural factors. Therefore, we investigate how the inclusion of potentially (confounding or mediating) variables change the relationship between genetic distance and migrant selection. If they do not matter a lot, we conclude that these variables are neither an omitted variable nor an important mediating factor. Importantly, this approach verifies whether there is one single variable (or a small set of variables) that can explain the effect of genetic distance on migrant selection. However, as genetic distance should serve as a proxy for complex differences in (normally unobserved) cultural traits and habits, we do not necessarily expect to find such a variable or set of variables, respectively.

To resolve the omitted variable problem and the measurement error issue, we use exogenous variation in genetic distance based on populations as they were in the year 1500 to explain genetic distance in 1990 (Spolaore and Wacziarg, 2009, 2016c). This approach matters especially for destination countries such as the United States and Australia, where native populations in 1500 were not at all influenced by later mass migration or colonization. To the extent that these countries are also more likely to have implemented selective immigration policies, we may find that ordinary-least-squares estimates are biased upward. Given the measurement concerns from above, however, the overall bias is ambiguous.

To be a valid instrument, the genetic distance in 1500 has to fulfill two conditions: First, it must be sufficiently correlated with the genetic distance in 1990, which we can test. Second, the exclusion restriction requires that the genetic distance in 1500 has an effect on the migrant skill mix in 2000 only through the genetic distance in 1990 (see Spolaore and Wacziarg (2009) for a detailed discussion of the validity of the instrumental

[21] An example for a policy that also accepts low-skilled migrants is the free movement of labor in the European Union.

variables approach). It is unlikely that the migrant skill mix in 2000 has a causal impact on genetic distance in 1500. It is also unlikely that genetic distance in 1500 affects the current migrant skill mix through other channels than current genetic distance. Thus, we are confident that the exclusion restriction holds in our setting.

Empirically, we estimate the model in two steps. In the first step, we predict the genetic distance in 1990 by using the variation in genetic distance in 1500, controlling for the full set of control variables. Because we want to identify a nonlinear relationship, Equations (2) and (3) give two first stage regressions of the two stage-least-squares procedure. The first model explains genetic distance and the second model explains the squared genetic distance both by linear and squared genetic distance in 1500 and conditional on the full set of control variables. An important prerequisite for this setup is that the squared genetic distance term in 1500 has enough predictive power (independently of the linear term) to identify the squared genetic distance. If this is not the case, the model is not identified.

$$GD_{sd} = \lambda_0 + \lambda_1 GD_{1500,sd} + \lambda_2 GD^2_{1500,sd} + \mathbf{X}'_{sd}\omega + \nu_{sd} \tag{2}$$

$$GD^2_{sd} = \psi_0 + \psi_1 GD_{1500,sd} + \psi_2 GD^2_{1500,sd} + \mathbf{X}'_{sd}\varpi + \varphi_{sd} \tag{3}$$

Once we have predicted genetic distance and squared genetic distance from the first stages, we include the fitted values into the second stage regression (Equation (4)). In this step, we use only the variation in genetic distance triggered by the variation in 1500.

$$\ln \frac{E^H_{sd}}{E^L_{sd}} - \ln \frac{E^H_s}{E^L_s} = \gamma_0 + \gamma_1 \widehat{GD_{sd}} + \gamma_2 \widehat{GD^2_{sd}} + \mathbf{X}'_{sd}\zeta + \mu_{sd} \tag{4}$$

Note that we estimate the first and the second stage within the same routine to account for the predicted values in the second stage, which is important to receive correct standard errors.

5 Results

5.1 Explaining Migrant Selection

Table 2 shows the results of estimating Equation (1) by ordinary least squares. This exercise gives a first impression of which variables are important for explaining migrant selection and develops our baseline model.

Column (1) of Table 2 shows the unconditional correlation between genetic distance and migrant selection. We observe that the coefficient on genetic distance is positive and highly significant (cf. also Figure 2). This indicates that a higher genetic distance of a country pair is associated on average with a more positively selected migrant skill mix. We have standardized both genetic distance and the migrant skill mix such that each of them has a standard deviation of one.[22] The interpretation of the coefficients is now in terms of standard deviations.

To model potential nonlinearity in genetic distance, next we introduce a squared term in Column (2).[23] We observe that the squared term is positive and highly significant, indicating that the migrant stock becomes more positively selected as genetic distance increases. At the same time, the coefficient on the linear term has decreased by half. In the next two columns, we add variables that are regarded as important control variables because they represent competing theories about migrant selection (see Section 3.3). The first set of variables is concerned with geographical differences and the second set of variables is concerned with differences in economic conditions.

Column (3) shows that geographical controls are important for explaining migrant selection as the R-squared increases from 0.275 to 0.456. Geographical distance is significantly associated with more positive selection, while country contiguity is negatively related to the migrant skill mix. The reason why the difference in absolute latitude matters more than the difference in absolute longitude is that most of our destination countries are in the Northern hemisphere. Thus, there is less variation along the longitudinal dimension than the latitudinal dimension.[24] The differences in the climate variables are not important for migrant selection. Including geographical controls, the coefficient on the linear genetic distance term drops substantially from 0.252 to –0.054, whereas the coefficient on the nonlinear term increases from 0.080 to 0.108. As expected, geographical conditions are a confounding factor because geographically distant populations have tended to split apart from each other earlier over time and therefore have drifted apart genetically to a greater degree.

In Column (4) of Table 2, we control for an important competing channel by introducing the difference in the 80/20 income ratio and the difference in GDP per capita. Estimates broadly follow the literature as both coefficients predict higher migrant selection—even though they are not strongly significant, which is mainly because the differences in the 80/20 income ratio and GDP per capita are highly correlated with each other ($r = 0.86$). Introducing both variables increases the coefficient on the nonlinear genetic distance term from 0.108 to 0.141 and decreases the linear genetic distance term from –0.054 to –0.213. In this specification, the linear term is also significantly different from zero.[25]

[22]For ease of interpretation, we do the same with geographical distance (taking logs afterwards) and language distance. This has the advantage that the coefficients between these important variables are directly comparable.

[23]The squared term is based on the standardized genetic distance variable.

[24]Table 1 shows that the standard deviation in the difference in absolute latitude is more than three times the standard deviation in the difference in absolute longitude.

[25]Because both measures are highly correlated, this effect does not depend on including either the difference in 80/20 income ratio or the difference in GDP per capita, respectively. Including only the difference in the 80/20 income ratios results in a coefficient on the linear term of –0.184 (0.101), significant at ten percent, and a coefficient on the squared term of 0.134 (0.017), significant at one percent. Including

Differences in geographical and economic conditions seem to be major confounding factors and therefore important control variables in our analysis. However, they mainly have an impact on the linear genetic distance term, indicating that they matter especially for migrant selection at lower levels of genetic distance. Given that benefits and costs associated with genetic distance should be relatively modest at those levels, it is plausible that geographical barriers—which are more easily crossed by high-skilled migrants—and economic incentives—which usually attract high-skilled migrants more than low-skilled migrants—dominate the overall pattern of migrant selection. Because genetic distance also correlates with geographical barriers and economic differences, it is important to control for these main confounding factors to isolate the effect of genetic distance on migrant selection. We now introduce the remaining variables that complement our baseline model.

By adding a dummy for an anglophone destination, a dummy for whether the two countries share a common language, and the Levensthein language distance to the model, Column (5) of Table 2 shows whether genetic distance is only a proxy for language differences. Interestingly, language distance—conditional on the other two language variables—does not show up significantly.[26] Checking the robustness of this result by using cladistic language distance, we can confirm that language distance is not significantly related to the migrant skill mix once we account for genetic differences (Column (6)).

Next, we introduce the measure for migrant networks into the model. As expected, the availability of migrant networks shows up as a highly negative significant predictor of migrant selection (Column (7)). Not surprisingly, introducing this variable also affects the coefficients on genetic distance. While the difference in the squared genetic distance terms between the models in Columns (5) and (7) is not statistically significant, the difference in the linear genetic distance terms is marginally significant at the ten percent level. Nevertheless, both genetic distance variables remain highly significantly correlated with migrant selection. This implies that the availability of migrant networks seems to reduce skill-specific migration costs relatively independently of genetic distance.

As discussed in Section 3.3, legal restrictions on immigration may be a major confounding factor. A dummy for bilateral travel visa restrictions enters with a positive and highly significant coefficient. We also add a dummy for a Schengen country pair and a dummy for a former colony. The introduction of the variables in Column (8) has no effect on the coefficients on genetic distance. In Column (9) of Table 2, we introduce

only the difference in GDP per capita results in a coefficient on the linear term of –0.216 (0.093), significant at five percent, and a coefficient on the squared term of 0.142 (0.016), significant at one percent.

[26] The coefficients on genetic distance are very similar when we include a squared term for the Levensthein language distance. Dropping the genetic distance variables from the regression in Column (5) of Table 2 leads to a marginally significant positive coefficient on language distance. Introducing the squared language distance term to this model leads to non-significant coefficients on both the linear and the squared language distance. This reassures us that genetic distance provides different information than language distance.

two measures for general openness, that is inflow of foreigners and of asylum-seekers in the destination country. Neither do these changes significantly affect the coefficients on genetic distance much. In further analysis, we use self-compiled measures of immigration policies that are based on destination-country-specific legislation (see Section 6.1).

The last column, Column (10), shows our fully specified model, which we use in all further applications in the paper. Here, we see that the coefficient on the difference in years of schooling is significantly positive, but the coefficients on genetic distance are again virtually unaffected.

To sum up, the results of the OLS model indicate a highly significant nonlinearity. As genetic distance increases, the migrant stock becomes less positively selected at lower levels of genetic distance and more positively selected at higher levels of genetic distance. After conditioning on geographical and economic differences, the coefficients on genetic distance are very stable over the specifications. Because most of the covariates we have introduced are arguably more mediating factors than omitted variables, we conclude that genetic distance captures an independent effect of cultural differences that is usually unobserved. The next section examines in detail the robustness of the OLS results with regard to potential endogeneity bias.

5.2 Dealing with Potential Endogeneity

The OLS results in Column (10) of Table 2 describe only a causal effect of genetic distance on migrant selection when genetic distance is uncorrelated with the error term in Equation (1). Following the discussion in Section 4.2, the major concern is omitted variable bias in the relationship between genetic distance (measured in 1990) and the migrant skill mix (measured in 2000). To address a potential bias, we use the instrumental variables (IV) approach suggested by Spolaore and Wacziarg (2009), exploiting the exogenous variation in genetic distance in 1500.

We report the corresponding IV results in Table 3. The first column replicates the OLS results for comparison. Columns (2) and (3) show the first-stage results. As discussed above, it is important that both the linear and the squared genetic distance are identified separately by the first stages. In Column (2), we find that the linear term of the genetic distance in 1500 is strongly associated with the linear term in genetic distance in 1990, whereas the squared genetic distance in 1500 has no predictive power. The first stage for the squared genetic distance in Column (3) shows that the squared term of the genetic distance in 1500 is a strong predictor for the squared genetic distance in 1990, while the linear term of the genetic distance in 1500 predicts the squared genetic distance too. To test whether the instruments are strong, we report the joint Kleibergen-Paap F statistic of the excluded instruments. This statistic shows a value of 42.0, which is well above conventional levels. Hence, we conclude that both the linear and the squared genetic distance are well identified.

The reduced form in Column (4) shows that there is also a direct relationship between the instruments and the outcome. This is reassurance of a causal effect that runs from the instrument through the endogenous variable. Finally, Column (5) shows the result from the IV estimation. While the effect is slightly stronger, we observe that the IV result confirms the nonlinear effect from the OLS regression in Column (1).

Plotting the conditional expectation of the migrant skill mix, i.e., $\mathbb{E}\left[\ln\left(E_{sd}^{H}/E_{sd}^{L}\right) - \ln\left(E_{s}^{H}/E_{s}^{L}\right)|\mathbf{X}\right]$, helps to interpret the model results. Figure 3 (left axis) documents a J-shaped pattern between genetic distance and the prediction of migrant selection conditional on all control variables. For expositional purpose, we add sample averages of the migrant skill mix, the scale of tertiary-educated migration, and the scale of primary-educated migration to the conditional expectations. Hence, the figure provides conditional expectations of the three outcomes at every point in the genetic distance distribution relative to their sample averages (see Online Appendix Figure B-1 for plots centered around zero). We see that the model predicts positive selection regardless of the level of genetic distance. But at low levels of genetic distance, we find that the migrant skill mix becomes slightly less positively selected as genetic distance increases. However, the association is rather modest at low levels of genetic distance compared to the increasingly positive selection at higher levels of genetic distance.

In the bottom of Table 3, we report marginal effects at the 10th, 50th, and 90th percentile of the genetic distance distribution. For a genetic distance at the 90th percentile, increasing genetic distance by one standard deviation increases positive selection by 45.6 percent of a standard deviation if estimated with OLS (Column (1)) and by 54.4 percent of a standard deviation if estimated with IV (Column (5)). Destination-sending-country pairs with those high genetic distance levels are, e.g., Germany—Vietnam, United Kingdom—Jamaica, and Australia—Malaysia. In contrast, for genetic distance at the 10th percentile, increasing genetic distance by one standard deviation decreases positive selection by 21.7 percent of a standard deviation if estimated with OLS (Column (1)) and by 33.2 percent of a standard deviation if estimated with IV (Column (5)). Destination-sending-country pairs with those low genetic distance levels are, e.g., Germany—France, United Kingdom—Sweden, and Australia—Netherlands.

In Figure 4, we evaluate the baseline model (Column (5) of Table 3) for each (standardized) genetic distance from 0 to 4.71 in 0.1 steps to show marginal effects along the entire genetic distance distribution. Following the J-shape pattern in Figure 3, the marginal effect is negative for low genetic distance levels and increases with genetic distance. Based on the parameters from the nonlinear model, we calculate that point estimates of marginal effects are positive for (standardized) genetic distances above 1.04 (69 percent of the sample). However, marginal effects are statistically indistinguishable from zero for genetic distances between 0.54 and 1.31. Because the median of the standardized genetic distance is 1.21, marginal effects are indistinguishable from zero or slightly negative for 26 percent or 35 percent, respectively, of the country pairs in the sample. As already revealed by the results above, genetic distances are only a barrier to migration when they are sufficiently high.

Using the marginal effects at the bottom of Table 3 (Column (5)), we can perform some effect size calculations. At the 90th percentile of the genetic distance distribution, positive migrant selection increases by 42.6 percent ($= 0.544 \cdot 1.567 \cdot 1/2$) if genetic distance increases by half of a standard deviation.[27] Evaluated at the 90th percentile ($E^H_{sd}/E^L_{sd} = 4.629$) and assuming that E^H_s/E^L_s stays constant,[28] this leads to an increase in the ratio of high- to low-skilled migrants by 1.97 ($= 0.426 \cdot 4.629$) high-skilled migrants for each low-skilled migrant. At the 50th percentile of the genetic distance distribution, migrant selection increases only by 4.9 percent ($= 0.063 \cdot 1.567 \cdot 1/2$) if genetic distance increases by half of a standard deviation. Evaluated at the 50th percentile ($E^H_{sd}/E^L_{sd} = 1.278$), this leads to an increase in the ratio of high- to low-skilled migrants by 0.06 ($= 0.049 \cdot 1.278$) high-skilled migrants for each low-skilled migrant. Finally, at the 10th percentile of the genetic distance distribution, positive migrant selection decreases by 26.0 percent ($-26.0 = -0.332 \cdot 1.567 \cdot 1/2$) if genetic distance increases by half of a standard deviation. Evaluated at the 10th percentile ($E^H_{sd}/E^L_{sd} = 0.970$), this leads to a decrease in the ratio of high- to low-skilled migrants by 0.25 ($-0.25 = -0.260 \cdot 0.970$) high-skilled migrants for each low-skilled migrant.

Investigating the scale of primary-educated and tertiary-educated migrants, we try to elicit who of them is responsible for the nonlinear association between genetic distance and the migrant skill mix. Figure 3 illustrate the conditional expectations of the emigration propensities for high- and low-skilled migrants (estimates are from the last two Columns of Table 3). As expected, there are substantial differences in the size of the emigration propensities, with larger propensities for high-skilled emigration over the entire genetic distance distribution. At higher levels of genetic distance, low-skilled migrants increasingly avoid migration, whereas high-skilled migrants are much less affected. At lower levels of genetic distance, the emigration propensities of high- and low-skilled migrants seem to react relatively similarly to higher genetic distances.[29]

We confirm these findings by plotting the corresponding marginal effects for migration propensities of both education groups along the genetic distance distribution in Figure 5 and by splitting the sample by country pairs above and below the median genetic distance in Table 4. Both indicate that migration propensity increases with genetic distance for both education groups at low levels of genetic distance. Thus, holding geographical and

[27] We choose an increase in genetic distance by half of a standard deviation to make no out-of-sample prediction.

[28] The percentile position refers to the migrant skill mix, i.e., to $\ln\left(E^H_{sd}/E^L_{sd}\right) - \ln\left(E^H_s/E^L_s\right)$. Then, E^H_{sd}/E^L_{sd} is the average high- to low-skilled migrant share at this percentile position.

[29] Online Appendix Figure B-1, plotting centered conditional expectations, shows that the migration propensity for low-skilled migrants reacts slightly stronger to genetic distance than the migration propensity of high-skilled migrants.

economic conditions, language differences, migration networks, and other control variables constant, it seems that larger genetic distances are a net benefit on average and attract migrants in general. The results indicate that the net benefits of genetic distance is slightly larger for low-educated migrants than for high-educated migrants. At higher levels of genetic distance, however, increasing genetic distance prevents low-skilled migrants from migrating but leaves high-skilled migrants largely unaffected.

6 Extensions and Robustness Checks

6.1 Selective Migration Policies

As already sketched in the theoretical considerations, immigration policies may not distinguish between high- and low-skilled migration when the sending country is culturally related, but favor high-skilled migration over low-skilled migration when the sending country is culturally unrelated. To capture those endogenous migration policies, we gather information on working visa and citizenship regulations for the destination countries in our sample. Specifically, we verify for both categories whether a destination country treats migrants from specific sending countries more favorably with respect to the skill level and occupational status (see Online Appendix Section D for details).

From our self-compiled data, we construct two dummy variables: one dummy for country pairs where the destination country offers favorable treatment for migrants from a sending country for granting a work visa, *privileged access to work visas*, and one dummy for receiving citizenship, *privileged access to citizenship*.[30] Regressing both dummies on the selection term without any further controls yields negative significant coefficients (see footnote), indicating that privileged access to work visas and privileged access to citizenship are associated with lower migrant selection on average.[31] However, using the two additional dummies conditional on all other control variables in our baseline regression, we observe that both variables have only limited predictive power and have even positive coefficients (Column (2) of Table 5). In other specifications, we find that the coefficient on privileged access to work visas becomes highly statistically significant (Column (5) to (7) of Online Appendix Table B-5). However, neither in these specifications do we find that the coefficients on genetic distance are substantially different. We conclude that the two migration policy variables do not confound the relationship between genetic distance and migrant selection.

What can explain this (non-)effect? In general, almost all destination countries have

[30] We also construct those two dummies without Commonwealth linkages because UK immigration regulations have become increasingly rigid also for these countries. Using these dummies leads to very similar results (results not shown).

[31] The coefficient on privileged access to work visas is equal to -0.778 (0.127), significant at the one percent level, and the coefficient on privileged access to citizenship is equal to -0.274 (0.145), significant at the ten percent level.

selective migration policies that favor high-skilled migrants and specific occupation groups, regardless where these migrants come from. Most sending-country-specific migration policies focus on country pairs (groups) that are either regionally close (e.g., the Nordic Passport Union or the free movement of labor within EU member states) or have shared historical origins (e.g., UK and the Commonwealth countries). Only a very small fraction of destination countries have additional bilateral entry regulations. In our dataset, only 14 percent of the country pairs have privileged access to work visas and only three percent privileged access to citizenship. The low number of specific regulations and treatments is in line with the fact that it is against the law in many destination countries to discriminate against migrants on cultural, religious, and ethnic grounds. Therefore, our refined measures of migration policies have no additional explanatory power over and above the control variables we have already used. However, this does not rule out that migration policies and local behavior toward migrants can be implicit and therefore are not codified in official documents. However, as long as this implicit behavior is driven by genetic distances, we still identify the total effect of genetic distance on migrant selection. In Section 6.3 below, we partly account for these unobserved policies by showing that our selection pattern holds when comparing country pairs with migrants from the same world region, which rules out that our results are driven by unobserved variation between world regions.

6.2 Robustness Checks

In Columns (3) and (4) of Table 5, we look at two kinds of constraints in the sending country that could also increase positive selection. On the one hand, poverty constraints could hinder low-skilled migration. Thus, we include the average predicted poverty rate in the sending country. The construction follows Belot and Hatton (2012) and uses data from the World Bank Development Indicators (Column (3)). On the other hand, political freedom in the sending country could be a major factor for high-skilled individuals because the literature shows that higher education also leads to stronger political interest and participation (Brade and Piopiunik, 2016; Dee, 2004; Milligan et al., 2004). Therefore, limited political freedom may push especially high-skilled individuals out of the country. We include the Freedom House Index for sending countries in Column (4).[32] The average poverty rate enters significantly with the correct sign, whereas the political freedom index is not significant. However, nonlinearity in genetic distance is not very much affected by including these control variables.

In further analysis, we also consider differences in religious orientation because we suspect that individuals are more likely to migrate to countries with a similar religious

[32] Meierrieks and Renner (2017) show that higher economic freedom in the sending country leads to less high-skilled migration than low-skilled migration. We do not find that controlling for economic freedom in the sending country significantly affects the coefficients on genetic distance.

background. We also control for differences in the economic structure of destination and sending countries because a similar industrial structure may also stem from a lower cultural distance and could induce larger migrant flows. Finally, we also consider the migratory difference in the distance to Addis Ababa between the sending and the destination country as a control to capture the difference in general within-country genetic diversity (Ashraf and Galor, 2013). None of these robustness checks significantly change the nonlinear J-shaped relationship between genetic distance and migrant selection (results are shown in Online Appendix Table B-5).

An important assumption for the validity of our estimates is the *assumption of irrelevant alternatives*. The assumption states that the presence of irrelevant alternative destination countries that are not included should neither affect the estimates. We can verify the stability of the parameters by piecewise omitting one destination country. Comparing the coefficients and significance levels in Appendix Table A-2 across the different samples shows that they are all in the same ballpark.

6.3 Region-Specific Effect Heterogeneity

Table 6 shows region-specific effect heterogeneity. Column (2) of Table 6 excludes four countries (Australia, Canada, USA, and the UK) known to have particularly selective immigration policies relative to the remainder of the destinations in our sample. The coefficients drop substantially, but nonlinearity remains highly significant and also within the sampling error of the baseline model (Column (1)). The effect is also reduced when we limit the sample to EU member states only (additionally excluding the UK) in Column (3).[33] Column (4) shows the results with destination fixed effects. This specification captures time-persistent differences between destination countries (such as strictness of immigration policies). In this specification, the linear term on genetic distance is no longer significant, which is to some degree due to a lower coefficient, but also due to a higher standard error. Nevertheless, including destination fixed effects also confirms the nonlinearity in the baseline model.[34]

In Columns (5) to (7) of Table 6, we include different sending-country-specific fixed effects. This analysis eliminates time-persistent differences between the sending regions. For example, it is possible that low-skilled migrants from specific regions or countries are not welcome in any of the destination countries for various reasons (including the

[33]The variation in genetic distance between European countries is very low, which prevents an meaningful within-EU analysis. The median genetic distance between European countries is smaller than the 10th percentile of the world sample.

[34]Appendix Table A-3 provides further fixed effects regressions for comparison. Using OLS and IV, we estimate models with destination-country fixed effects, sending-country fixed effects, and destination-by-sending-country fixed effects. Even though coefficients on genetic distance are substantially reduced (in absolute terms), we find in all specifications that the linear genetic distance term is negative (not significant) and that the squared genetic distance term is always positive (significant at least at the ten percent level).

implicit migration policies outlined above). In Column (5), we begin by including sending-continent fixed effects (see Appendix Table A-1 for the assignment of sending countries to continents and regions). This has only small effects on the coefficients on genetic distance. In the next column, we account for differences between sending world regions that are more detailed than continents. While these fixed effects reduce the size of the coefficients, we can still confirm the nonlinear relationship from the baseline model.[35] Ultimately, we estimate the model by including sending-country fixed effects (Column (7)). This leads to an estimation approach that compares within sending countries the extent of migrant selection to the 15 different destination countries. In this specification, the regression refers more to *sorting* into different destinations than to *selection* in general (see Grogger and Hanson, 2011, for a detailed discussion about sorting versus selection). Nevertheless, we can confirm the selection pattern even in this demanding specification.[36] We conclude from this analysis that there is consistent evidence of a substantial nonlinear relationship between genetic distance and migrant selection.

7 Conclusion

This paper provides the first evidence on the impact of long-term relatedness between countries, measured by genetic distance, on the selectivity of international migration. It thereby connects the literature on the economic outcomes of long-term relatedness with the literature that examines the determinants of migrant selection. Using bilateral education-specific migrant stocks of the main 15 destination countries for the year 2000, we find that genetic distance affects migrant selection nonlinearly. Conditional on geographical and economic differences, our results imply that the association follows a J-shape pattern: at low levels of genetic distance, increases in genetic distance reduce the positive migrant selection. This effect is reversed at higher levels of genetic distance, and the migrant stock becomes increasingly more positively selected with increasing genetic distance. Further analysis by skill group shows that the relationship between genetic distance and migrant selection is mainly driven by strongly decreasing *net* benefits from genetic distance for low-skilled migrants. When genetic distance is low, we find that low-skilled migrants are slightly more attracted by increases in genetic distance than high-skilled migrants. However, at higher levels of genetic distance, low-skilled migrants increasingly avoid migrating to culturally distant destinations, indicating rising migration costs induced by increasing cultural distance. By contrast, high-skilled migrants are

[35] Another possibility to check the robustness of the results regarding migrants from different sending regions is to omit one sending region at a time. Appendix Table A-4 shows that the results do not depend on one single sending region.

[36] Because sending-country fixed effects remove the main variation in genetic distance—that comes from variation between sending countries and not from variation between destination countries—the first stages, however, suffer from a potential weak instrumental variable problem (indicated by a low F statistic).

generally not strongly affected by cultural distance.

The result is robust to different bilateral covariates. Importantly, we document an independent effect over and above many variables that were previously used by other researchers as proxies for cultural differences (such as nonlinear geographical distances, languages differences, religious differences, and the availability of networks). By using traditional and self-compiled measures of skill-selective migration policies, arguably the most important instrument for controlling migration in destination countries, we document that they also cannot explain the pattern between genetic distance and migrant selection. This does not preclude the possibility that migrant selection is affected by informal and unobservable destination-specific policies or other demand-side factors that are correlated with cultural distance. However, various destination- and sending-country-specific fixed-effect specifications confirm our baseline results, ruling out that the effect is entirely due to destination-specific or origin-specific unobserved behavior or policies toward migrants. Finally, we test whether remaining omitted variables bias our results. We confirm our baseline results by employing an instrumental variables approach that uses exogenous variation in genetic distances in 1500 to predict current genetic distance. This strategy specifically avoids using variation that is driven by later mass migration and colonization.

We conclude that genetic distance, as a proxy for usually unobserved differences in cultural factors, for instance, differences in (more complex) informal networks, cultural norms, traits, and habits, enter the individual migration decision and thereby cause specific selection patterns. Because these patterns are deeply rooted in the populations' norms and belief systems, which develop over long periods, they tend to shape migration flows in a 'natural' way. Policy makers who are interested in designing migration policies to attract certain types of migrants have to take such deeply rooted factors into account. For example, destination countries can already expect to receive a more positively selected migrant intake from sending countries that are not closely related. However, it is possible that relatedness-induced migration costs are so high that only very few highly-skilled migrants find it attractive to migrate. To attract a larger number of high-skilled migrants, the destination country needs to find ways to lower the cultural distance or to provide other compensation that increases migration benefits or decreases migration costs, respectively. Destination countries may also be interested in low-skilled migrants (e.g., for seasonal work and vacant jobs in low-skilled occupations). Because these migrants react more strongly to relatedness-induced migration benefits and costs than high-skilled migrants, it is even more important to factor in cultural differences in migration policies targeted at low-skilled migrants. Either way, not acknowledging cultural differences between countries could lead to migration policies that fail to achieve the destination country's desired skill mix of the migrant stock.

References

Adserà, A. and Pytliková, M. (2015). The Role of Language in Shaping International Migration. *Economic Journal*, 125(586):F49–F81.

Akcigit, U., Baslandze, S., and Stantcheva, S. (2016). Taxation and the International Mobility of Inventors. *American Economic Review*, 106(10):2930–2981.

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

Alesina, A., Guiliano, P., and Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *Quarterly Journal of Economics*, 128(2):469–530.

Ashraf, Q. and Galor, O. (2013). The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *American Economic Review*, 103(1):1–46.

Barro, R. and Lee, J.-W. (2013). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104:184–198.

Bauernschuster, S., Falck, O., Heblich, S., Suedekum, J., and Lameli, A. (2014). Why Are Educated and Risk-Loving Persons More Mobile Across Regions? *Journal of Economic Behavior & Organization*, 98:56–69.

Beine, M., Docquier, F., and Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, 95(1):30–41.

Beine, M., Docquier, F., and Rapoport, H. (2008). Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers. *Economic Journal*, 118(528):631–652.

Belot, M. and Ederveen, S. (2011). Cultural Barriers in Migration between OECD Countries. *Journal of Population Economics*, 25(3):1077–1105.

Belot, M. and Hatton, T. J. (2012). Immigrant Selection in the OECD. *Scandinavian Journal of Economics*, 114(4):1105–1128.

Benson, M. and O'Reilly, K., editors (2009a). *Lifestyle Migration. Expectations, Aspirations and Experiences.* Ashgate.

Benson, M. and O'Reilly, K. (2009b). Migration and the Search for a Better Way of Life: A Critical Exploration of Lifestyle Migration. *Sociological Review*, 57(4):608–625.

Benson, M. and O'Reilly, K. (2016). From Lifestyle Migration to Lifestyle in Migration: Categories, Concepts, and Ways of Thinking. *Migration Studies*, 4(1):20–37.

Bertoli, S. and Fernández-Huertas Moraga, J. (2015). The Size of the Cliff at the Border. *Regional Science and Urban Economics*, 51:1–6.

Bertoli, S., Fernández-Huertas Moraga, J., and Ortega, F. (2011). Immigration Policies and the Ecuadorian Exodus. *World Bank Economic Review*, 25(1):57–76.

Bhagwati, J. and Hamada, K. (1974). The Brain Drain, International Integration of Markets for Professionals and Unemployment: A Theoretical Analysis. *Journal of Development Economics*, 1(1):19–42.

Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *American Economic Review*, 77(4):531–553.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process.* University of Chicago Press.

Brade, R. and Piopiunik, M. (2016). Education and Political Participation. *CESifo DICE*

Report, 14(1):70–73.

Burchardi, K. B. and Hassan, T. A. (2013). The Economic Impact of Social Ties: Evidence from German Reunification. *Quarterly Journal of Economics*, 128(3):1219–1271.

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

Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press.

Chiquiar, D. and Hanson, G. H. (2005). International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113(2):239–281.

Chiswick, B. R. (1999). Are Immigrants Favorably Self-Selected? *American Economic Review*, 89(2):181–185.

Clark, X., Hatton, T. J., and Williamson, J. G. (2007). Explaining U.S. Immigration, 1971–1998. *Review of Economics and Statistics*, 89(2):359–373.

Coe, D. T. and Helpman, E. (1995). International R&D Spillovers. *European Economic Review*, 39(5):859–887.

Comin, D. A., Dmitriev, M., and Rossi-Hansberg, E. (2012). The Spatial Diffusion of Technology. NBER Working Paper 18534.

Dahl, M. S. and Sorenson, O. (2010). The Migration of Technical Workers. *Journal of Urban Economics*, 67(1):33–45.

Dee, T. S. (2004). Are there Civic Returns to Education? *Journal of Public Economics*, 88(9-10):1697–1720.

Desmet, K., Breton, M., Ortuño-Ortín, I., and Weber, S. (2011). The Stability and Breakup of Nations: A Quantitative Analysis. *Journal of Economic Growth*, 16(3):183–213.

Docquier, F., Lowell, B. L., and Marfouk, A. (2007). A Gendered Assessment of the Brain Drain. Mimeo.

Dustmann, C. and Preston, I. P. (2007). Racial and Economic Factors in Attitudes to Immigration. *B.E. Journal of Economic Analysis & Policy*, 7(1):1–39.

Falck, O., Heblich, S., Lameli, A., and Südekum, J. (2012). Dialects, Cultural Identity, and Economic Exchange. *Journal of Urban Economics*, 72(2-3):225–239.

Falck, O., Lameli, A., and Ruhose, J. (2017). Cultural Biases in Migration: Estimating Non-Monetary Migration Costs. *Papers in Regional Science*, forthcoming.

Fearon, J. D. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, 8(2):195–222.

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

Fernández-Huertas Moraga, J. (2011). New Evidence on Emigrant Selection. *Review of Economics and Statistics*, 93(1):72–96.

Giuliano, P., Spilimbergo, A., and Tonon, G. (2014). Genetic Distance, Transportation Costs, and Trade. *Journal of Economic Geography*, 14(1):179–198.

Gosnell, H. and Abrams, J. (2011). Amenity Migration: Diverse Conceptualizations of Drivers, Socioeconomic Dimensions, and Emerging Challenges. *GeoJournal*, 76(4):303–322.

Grogger, J. and Hanson, G. H. (2011). Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics*, 95(1):42–57.

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives*, 20(2):23–48.

Guiso, L., Sapienza, P., and Zingales, L. (2009). Cultural Biases in Economic Exchange. *Quarterly Journal of Economics*, 124(3):1095–1131.

Hainmueller, J. and Hiscox, M. J. (2007). Educated Preferences: Explaining Attitudes Toward Immigration in Europe. *International Organization*, 61(2):399–442.

Head, K., Mayer, T., and Ries, J. (2010). The Erosion of Colonial Trade Linkages after Independence. *Journal of International Economics*, 81(1):1–14.

Henrich, J. and McElreath, R. (2003). The Evolution of Cultural Evolution. *Evolutionary Anthropology*, 12(3):123–135.

Hoey, B. A. (2005). From Pi to Pie: Moral Narratives of Noneconomic Migration and Starting Over in the Postindustrial Midwest. *Journal of Contemporary Ethnography*, 34(5):586–624.

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

Kaestner, R. and Malamud, O. (2014). Self-Selection and International Migration: New Evidence from Mexico. *Review of Economics and Statistics*, 96(1):78–91.

Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, 217(5129):624–626.

Kleven, H. J., Landais, C., Saez, E., and Schultz, E. (2014). Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark. *Quarterly Journal of Economics*, 129(1):333–378.

Krieger, T. and Lange, T. (2010). Education Policy and Tax Competition with Imperfect Student and Labor Mobility. *International Tax and Public Finance*, 17(6):587–606.

Krieger, T., Renner, L., and Ruhose, J. (2015). Genetic Distance and International Migrant Selection. CESifo Working Paper No. 5453.

Krieger, T., Renner, L., and Ruhose, J. (2018). Long-Term Relatedness between Countries and International Migrant Selection: Replication Files, Mendeley Data, v1. http://dx.doi.org/10.17632/mjkg85jyc7.1.

Mayda, A. M. (2006). Who Is against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants. *Review of Economics and Statistics*, 88(3):510–530.

Mayda, A. M. (2009). International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows. *Journal of Population Economics*, 23(4):1249–1274.

McFadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics*, 3(4):303–328.

Meierrieks, D. and Renner, L. (2017). Stymied Ambition: Does a Lack of Economic Freedom Lead to Migration? *Journal of Population Economics*, 30(3):977–1005.

Milligan, K., Moretti, E., and Oreopoulos, P. (2004). Does Education Improve Citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics*, 88(9-10):1667–1695.

Nelson, R. R. and Phelps, E. S. (1966). Investment in Humans, Technological Diffusion, and Economic Growth. *American Economic Review*, 56(1/2):69–75.

Neumayer, E. (2006). Unequal Access to Foreign Spaces: How States Use Visa Restrictions to Regulate Mobility in a Globalized World. *Transactions of the British Institute of Geographers*, 31(3):72–84.

OECD (2016). International Migration Outlook 2016. OECD Publishing, `http://dx.doi.org/10.1787/migr_outlook-2016-en`.

Oreopoulos, P. and Salvanes, K. G. (2011). Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives*, 25(1):159–184.

Ortega, F. and Peri, G. (2013). The Effect of Income and Immigration Policies on International Migration. *Migration Studies*, 1(1):47–74.

Ottaviano, G. I. and Peri, G. (2005). The Economic Value of Cultural Diversity: Evidence from US Cities. *Journal of Economic Geography*, 6(1):9–44.

Parey, M., Ruhose, J., Waldinger, F., and Netz, N. (2017). The Selection of High-Skilled Emigrants. *Review of Economics and Statistics*, 99(5):776–792.

Patt, A., Ruhose, J., Wiederhold, S., and Flores, M. (2017). International Emigrant Selection on Occupational Skills. IZA Discussion Paper No. 10837.

Pemberton, T. J., DeGiorgio, M., and Rosenberg, N. A. (2013). Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation. *G3-Genes/Genomes/Genetics*, 3(5):891–907.

Roback, J. (1982). Wages, Rents, and the Quality of Life. *Journal of Political Economy*, 90(6):1257–1278.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1):34–55.

Schwartz, A. (1973). Interpreting the Effect of Distance on Migration. *Journal of Political Economy*, 81(5):1153–1169.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70(5):80–93.

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

Spolaore, E. and Wacziarg, R. (2012). Long-Term Barriers to the International Diffusion of Innovations. In Frankel, J. and Pissarides, C., editors, *NBER International Seminar on Macroeconomics*, chapter 1, pages 11–46. University of Chicago Press.

Spolaore, E. and Wacziarg, R. (2013). How Deep Are the Roots of Economic Development? *Journal of Economic Literature*, 51(2):325–369.

Spolaore, E. and Wacziarg, R. (2016a). Ancestry and Development: New Evidence. Working Paper.

Spolaore, E. and Wacziarg, R. (2016b). Ancestry, Language, and Culture. In Ginsburgh, V. and Weber, S., editors, *The Palgrave Handbook of Economics and Language*, chapter 6. London: Palgrave Macmillan.

Spolaore, E. and Wacziarg, R. (2016c). War and Relatedness. *Review of Economics and Statistics*, 98(5):925–939.

Tabellini, G. (2010). Culture and Institutions: Economic Development in the Regions of Europe. *Journal of the European Economic Association*, 8(4):677–716.

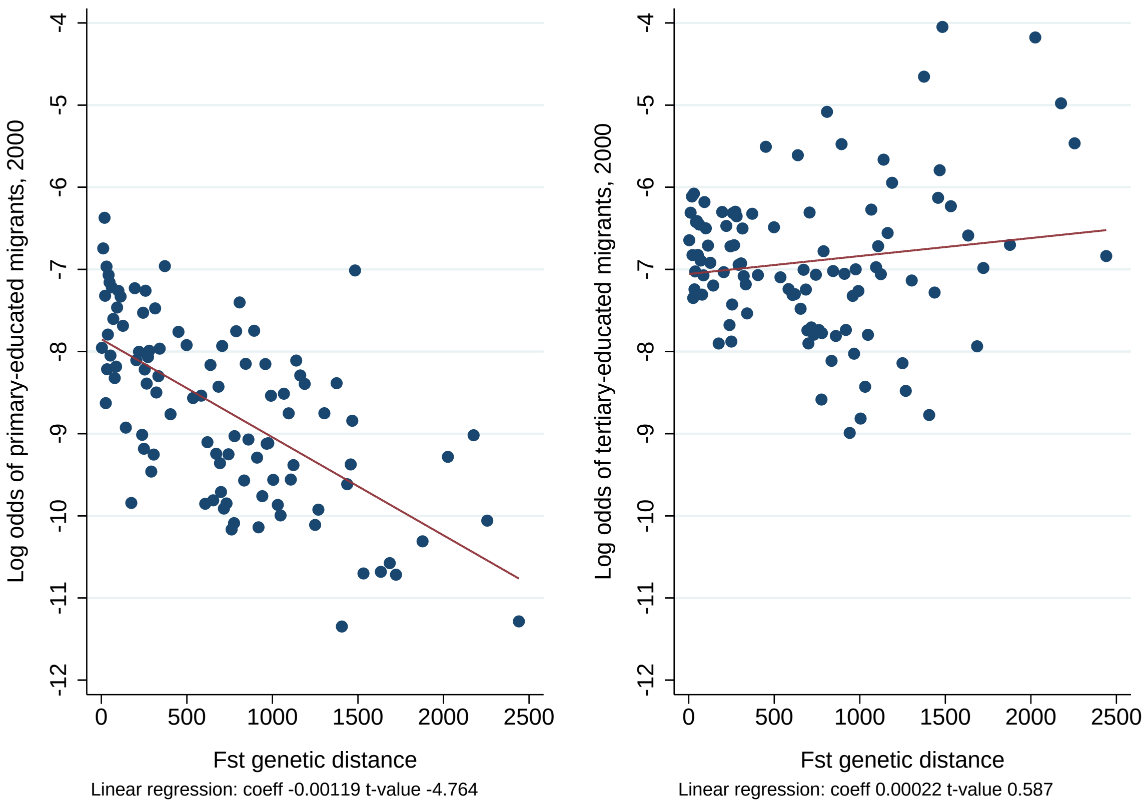
Torkington, K. (2010). Defining Lifestyle Migration. *Dos Algarves*, 19:99–111.

Voigtländer, N. and Voth, H.-J. (2012). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany. *Quarterly Journal of Economics*, 127:1229–1392.

Wilson, J. D. (2008). A Voluntary Brain-Drain Tax. *Journal of Public Economics*, 92(12):2385–2391.

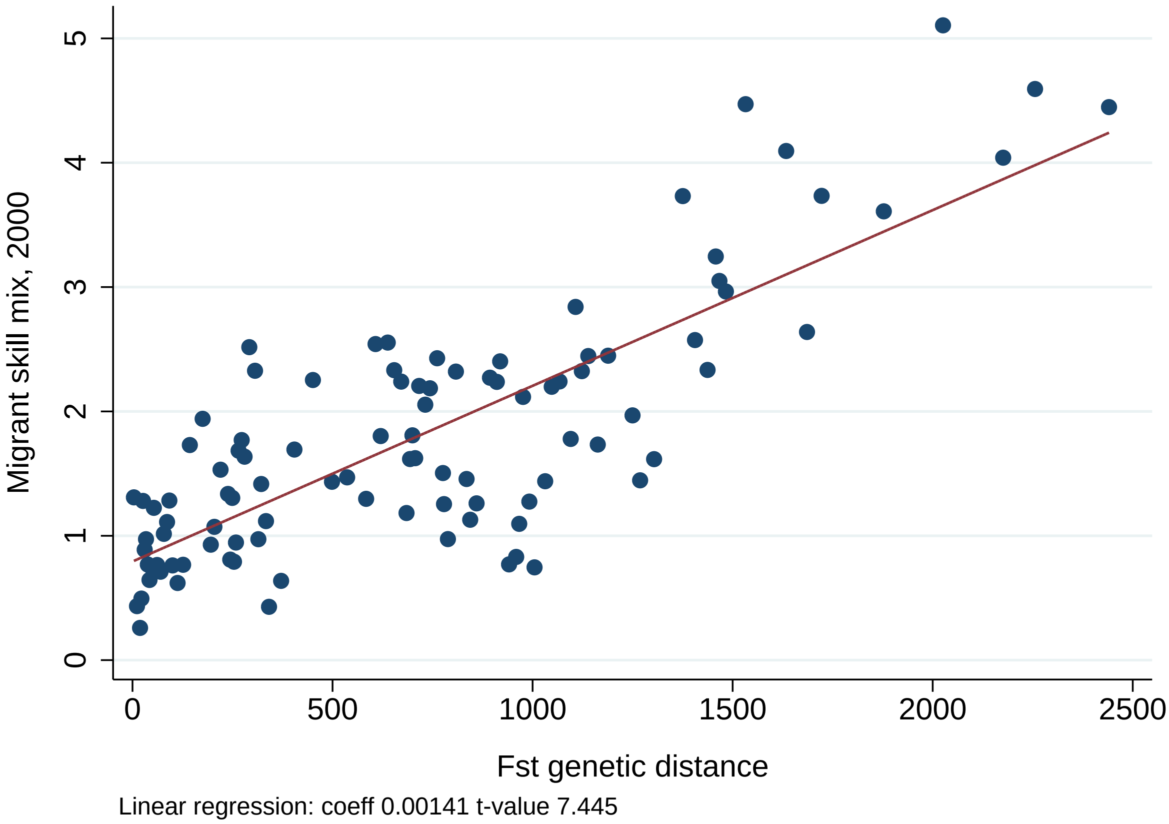
Figures and Tables

Figure 1: ***Genetic Distance and Emigration Odds by Education Level, 2000***



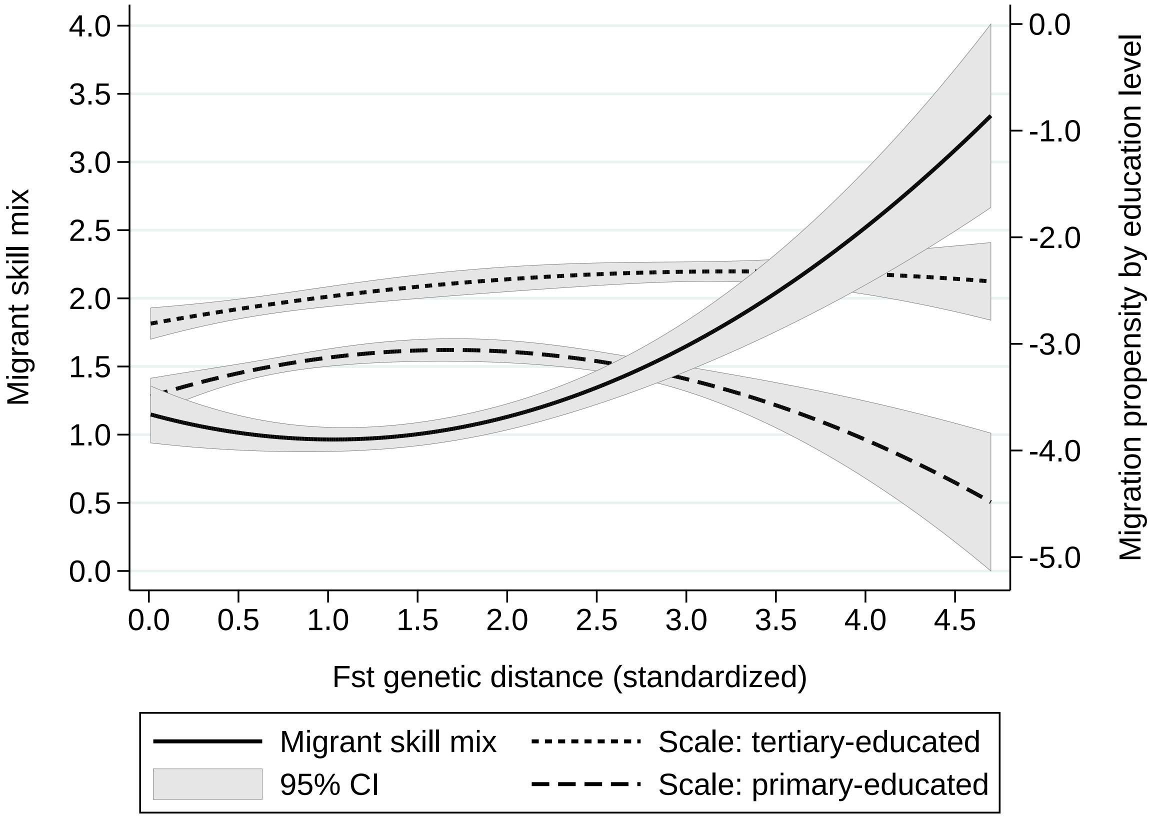
Notes: The figure on the left (right) shows the relationship between the log odds of emigration for primary-educated (tertiary-educated) migrants and genetic distance. Genetic distance is binned into 100 equal sized bins. The plots show the mean of genetic distance and log emigration odds within each bin. Coefficients and t-statistics are from OLS regressions on the microdata. Data on migrant stocks by skill level are from Docquier et al. (2007) and the genetic distance data are from Spolaore and Wacziarg (2009).

Figure 2: ***Genetic Distance and Emigrant Selection, 2000***



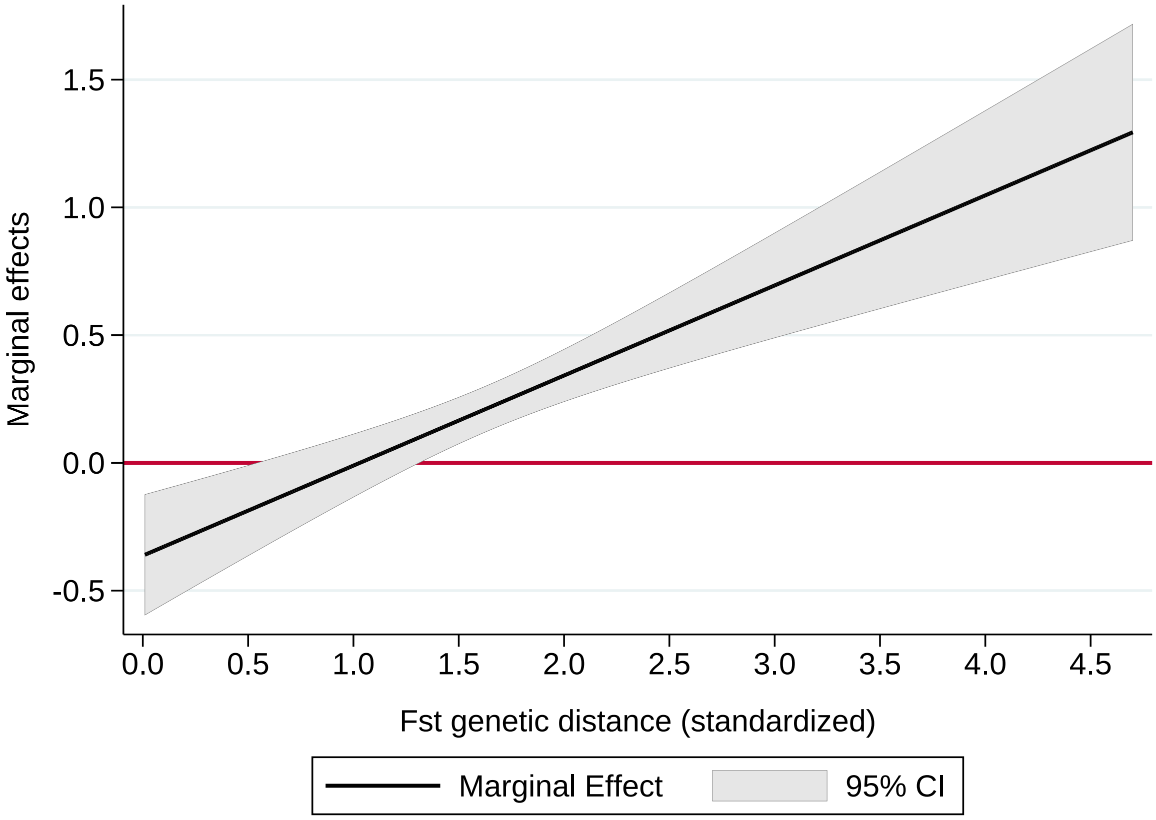
Notes: The figure shows the relationship between migrant skill mix and genetic distance. Higher values in the migrant skill mix indicate a more positively selected migrant population relative to the population in the sending country. Genetic distance is binned into 100 equal sized bins. The plots show the mean of genetic distance and log emigration odds within each bin. Coefficients and t-statistics are from OLS regressions on the microdata. Data on migrant stocks by skill level are from Docquier et al. (2007) and the genetic distance data are from Spolaore and Wacziarg (2009).

Figure 3: ***Conditional Expectation of Emigrant Selection***



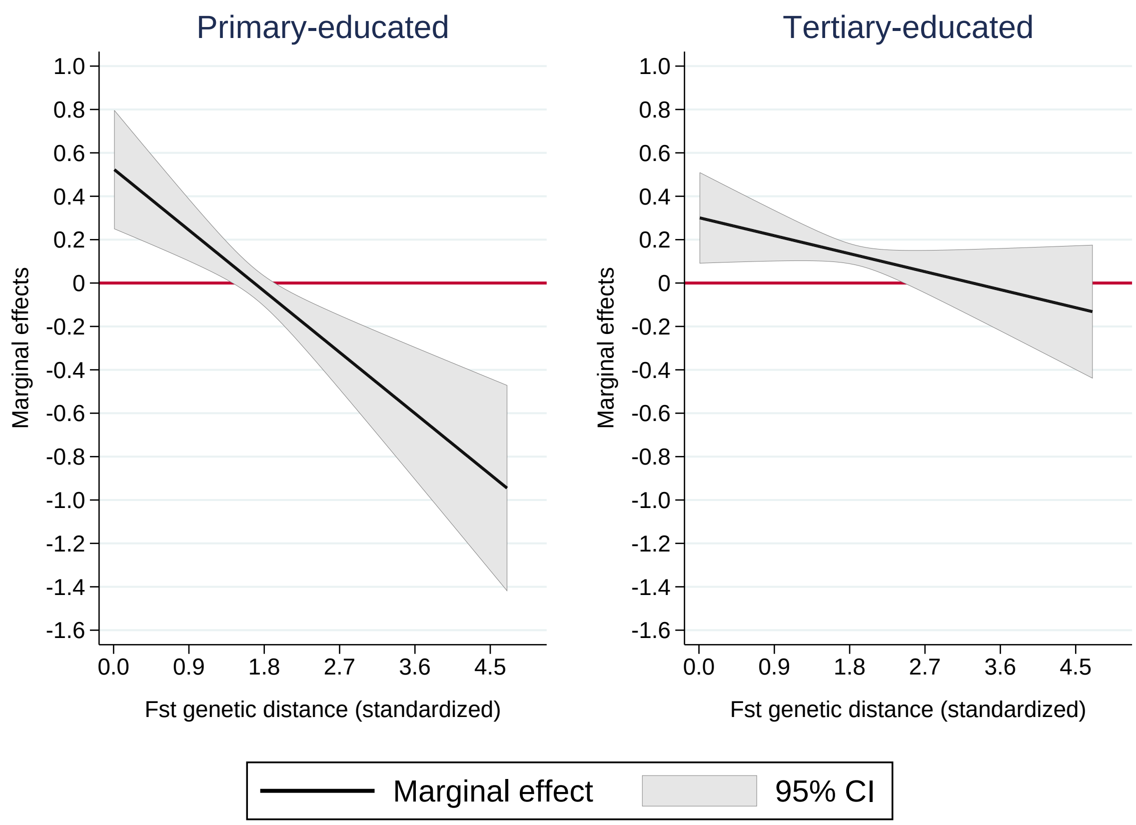
Notes: The figure shows the conditional expectation of the migrant skill mix, i.e., $\mathbb{E}\left[\ln\left(E_{sd}^H/E_{sd}^L\right) - \ln\left(E_s^H/E_s^L\right)|\mathbf{X}\right]$ (left axis) and the conditional expectation for the scale of emigration by education level, i.e., $\mathbb{E}\left[E_{sd}^j/E_s^j|\mathbf{X}\right]$ with $j \in$ {primary-educated, tertiary-educated} (right axis). To construct the figure, we use the estimates that condition on the full set of control variables from Columns (5), (6), and (7) of Table 3. Conditional expectations of the three outcomes are computed relative to their sample averages.

Figure 4: ***Marginal Effects on Migrant Skill Mix***



Notes: The figure shows marginal effects of genetic distance on the migrant skill mix by the level of genetic distance. Migrant skill mix and genetic distance are standardized by their standard deviations. The figure is obtained by evaluating the non-linear model in Column (5) of Table 3 for each level of genetic distance. Marginal effects are computed by $\gamma_1 + \gamma_2 \cdot 2 \cdot GD$.

Figure 5: ***Marginal Effects on Migration Propensity by Education Level***



Notes: The figure shows marginal effects of genetic distance on the scale of migrants (log odds of emigration) by the level of genetic distance for primary-educated migrants (left panel) and tertiary-educated migrants (right panel). Education-specific log odds of emigration and genetic distance are standardized by their standard deviations. The figure is obtained by evaluating the non-linear model in Columns (6) and (7) of Table 3 for each level of genetic distance. Marginal effects are computed by $\gamma_1 + \gamma_2 \cdot 2 \cdot GD$.

Table 1: ***Summary Statistics***

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
Panel A: Genetic Distance Data					
F_{ST} genetic distance	716	572	0	2,695	1,102
F_{ST} genetic distance, 1500	989	650	0	3,557	1,102
Standardized					
F_{ST} genetic distance, standardized	1.252	1	0	4.710	1,102
F_{ST} genetic distance, standardized, 1500	1.522	1	0	5,473	1,102
Panel B: Migrant Selection Measures					
Migrant skill mix, $\ln(E^H_{sd}/E^L_{sd}) - \ln(E^H_s/E^L_s)$	1.805	1.567	-3.041	8.077	1,102
Primary-educated emigration share, E^L_{sd}/E^L_s	0.003	0.016	0	0.251	1,102
Log primary-educated emigration share	-8.704	2.488	-15.067	-1.380	1,102
Tertiary-educated emigration share, E^H_{sd}/E^H_s	0.029	0.228	0	4.798	1,102
Log tertiary-educated emigration share	-6.899	2.453	-13.200	1.568	1,102
Standardized					
Migrant skill mix	1.152	1	-1.940	5.154	1,102
Log primary-educated emigration share	-3.499	1	-6.056	-0.555	1,102
Log tertiary-educated emigration share	-2.812	1	-5.381	0.639	1,102
Panel C: Control Variables					
Log geographic distance*	8.541	0.973	5.081	9.880	1,102
Contiguous	0.030	–	0	1	1,102
Δ absolute latitude	-9.274	68.585	-172	172	1,102
Δ absolute longitude	18.738	20.728	-39	62.633	1,102
Δ temperature	-8.546	10.787	-36.513	29.264	1,102
Δ precipitation	-27.232	64.335	-202.626	135.989	1,102
Δ 80/20 wage ratio	22.632	10.276	-7.934	47.999	1,102
Δ GDP per capita	15,856	12,799	-32,624	37,006	1,102
Anglophone destination	0.406	–	0	1	1,102
Common language	0.166	–	0	1	1,102
Levensthein language distance*	86.732	23.890	0	105.270	1,102
Cladistic language distance*	0.948	0.109	0.110	1	991
Migrant networks	0.004	0.018	0	0.263	1,102
Travel visa restriction	0.477	–	0	1	1,102
Schengen pair	0.159	–	0	1	1,102
Colony	0.064	–	0	1	1,102
Log inflow foreigners	11.286	1.236	8.979	13.421	1,102
Log inflow asylum-seekers	9.670	1.117	7.332	11.463	1,102
Δ years of schooling	2.637	2.975	-4.571	11.984	1,102
Δ share tertiary	6.116	8.842	-22.836	30.437	1,102
Privileged access to work visas	0.142	–	0	1	1,102
Privileged access to citizenship	0.027	–	0	1	1,102
Political freedom	2.798	1.651	1	7	1,082
Average poverty rate, predicted	25.507	18.2450	4.760	75.395	1,029

Notes: The F_{ST} genetic distance, which can take values between 0 and 1 in the data matrix provided by Cavalli-Sforza et al. (1994), is multiplied by 10,000. Δ represents the simple difference between destination and sending country, that is, $\Delta X = X_d - X_s$. *In the regression models, F_{ST} genetic distance, F_{ST} genetic distance, 1500, language distance, and geographical distance are standardized such that they have a standard deviation of 1 over all country pairs in the sample. Standard deviations are not reported for dummy variables. See Appendix Table B-1 for variable definitions and data sources.

Table 2: ***Nonlinear Effect of Genetic Distance on Migrant Selection: OLS Results***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
F_{ST} genetic distance$_{sd}$	0.516***	0.252*	–0.054	–0.213**	–0.240***	–0.232***	–0.198**	–0.218***	–0.230***	–0.240***
	(0.069)	(0.119)	(0.106)	(0.094)	(0.071)	(0.077)	(0.069)	(0.066)	(0.070)	(0.078)
F_{ST} genetic distance$_{sd}$, squared		0.080***	0.108***	0.141***	0.147***	0.145***	0.137***	0.138***	0.142***	0.135***
		(0.021)	(0.016)	(0.017)	(0.019)	(0.020)	(0.016)	(0.015)	(0.024)	(0.025)
Log geographic distance$_{sd}$			0.157**	0.169***	0.063	0.047	0.038	0.061	0.032	–0.008
			(0.068)	(0.052)	(0.050)	(0.047)	(0.045)	(0.049)	(0.060)	(0.054)
Contiguous$_{sd}$			–0.474***	–0.307*	–0.360**	–0.387***	–0.320**	–0.355**	–0.414***	–0.521***
			(0.130)	(0.152)	(0.139)	(0.119)	(0.144)	(0.143)	(0.128)	(0.109)
Δ absolute latitude$_{sd}$			0.005***	0.005***	0.003***	0.003***	0.003***	0.003***	0.003*	0.000
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Δ absolute longitude$_{sd}$			0.007	0.004	0.006	0.004	0.005	0.005	0.010**	0.004
			(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.002)
Δ temperature$_{sd}$ $\times 10^{-1}$			–0.171	–0.119	–0.067	–0.072	–0.085	–0.083	–0.036	–0.003
			(0.106)	(0.090)	(0.076)	(0.073)	(0.077)	(0.079)	(0.081)	(0.053)
Δ precipitation$_{sd}$ $\times 10^{-1}$			–0.004	–0.009	–0.012	–0.012	–0.014**	–0.013*	–0.008	–0.003
			(0.011)	(0.010)	(0.007)	(0.007)	(0.006)	(0.006)	(0.008)	(0.005)
Δ 80/20 wage ratio$_{sd}$				0.008	–0.010	–0.013	–0.007	–0.006	–0.007	–0.015
				(0.011)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)
Δ GDP per capita$_{sd}$ $\times 10^{-4}$				0.153*	0.298***	0.330***	0.288***	0.258***	0.260***	0.226***
				(0.079)	(0.077)	(0.082)	(0.078)	(0.072)	(0.066)	(0.058)
Anglophone destination$_d$					0.561***	0.599***	0.581***	0.618***	0.546**	0.656***
					(0.173)	(0.172)	(0.170)	(0.171)	(0.211)	(0.117)
Common language$_{sd}$					0.207*	0.218*	0.255**	0.248**	0.193*	0.175*
					(0.114)	(0.105)	(0.104)	(0.104)	(0.092)	(0.098)
Levensthein language distance$_{sd}$					0.022		–0.005	–0.017	–0.021	–0.051
					(0.025)		(0.031)	(0.029)	(0.031)	(0.037)
Cladistic language distance$_{sd}$						0.018				
						(0.024)				
Migrant networks$_{sd}$							–6.288***	–6.268***	–7.060***	–6.245***
							(1.403)	(1.411)	(1.836)	(1.803)
Travel visa restriction$_{sd}$								0.157**	0.117**	–0.033
								(0.054)	(0.047)	(0.053)
Schengen pair$_{sd}$								0.185*	0.196*	0.251**
								(0.091)	(0.096)	(0.091)
Colony$_{sd}$								–0.054	–0.073	–0.062
								(0.095)	(0.087)	(0.075)
Log inflow foreigners$_d$									0.192	0.088
									(0.111)	(0.122)
Log inflow asylum-seekers$_d$									–0.115	–0.024
									(0.118)	(0.142)
Δ years of schooling$_{sd}$										0.147***
										(0.021)
Δ share tertiary$_{sd}$										0.005
										(0.007)
R-squared	0.265	0.275	0.456	0.511	0.558	0.553	0.568	0.571	0.584	0.671
Observations	1,102	1,102	1,102	1,102	1,102	991	1,102	1,102	1,102	1,102
Cluster	15	15	15	15	15	14	15	15	15	15

Notes: The dependent variable is the (standardized) migrant skill mix in 2000, i.e., $\ln\left(E^H_{sd}/E^L_{sd}\right) - \ln\left(E^H_s/E^L_s\right)$ divided by the standard deviation. F_{ST} genetic distance, geographic distance, and language distance(s) are standardized such that they have a standard deviation of 1 over all country pairs in the sample. Δ represents the simple difference between destination and sending country, that is, $\Delta X = X_d - X_s$. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 3: ***Nonlinear Effect of Genetic Distance on Migrant Selection: Instrumental Variable Results***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Selection	Genetic distance		Selection		Scale	
		Linear	Squared			Primary	Tertiary
	OLS	FS	FS	RF	IV	IV	IV
F_{ST} genetic distance$_{sd}$	–0.240***				–0.363***	0.526***	0.301***
	(0.078)				(0.121)	(0.140)	(0.107)
F_{ST} genetic distance$_{sd}$, squared	0.135***				0.176***	–0.156***	–0.046*
	(0.025)				(0.034)	(0.040)	(0.028)
F_{ST} genetic distance$_{sd}$, 1500		0.748***	0.816**	–0.128*			
		(0.057)	(0.284)	(0.064)			
F_{ST} genetic distance$_{sd}$, 1500, squared		0.012	0.565***	0.095***			
		(0.009)	(0.123)	(0.017)			
Control variables	yes	yes	yes	yes	yes	yes	yes
R-squared	0.671	0.763	0.774	0.666	0.669	0.673	0.746
Observations	1,102	1,102	1,102	1,102	1,102	1,102	1,102
Cluster	15	15	15	15	15	15	15
Kleibergen-Paap F statistic					42.0	42.0	42.0
Marginal effects for F_{ST} genetic distance at							
10th percentile	–0.217**	0.748***	0.853***	–0.122*	–0.332***	0.498***	0.293***
	(0.074)	(0.057)	(0.277)	(0.063)	(0.115)	(0.133)	(0.102)
50th percentile	0.087**	0.787***	2.668***	0.184***	0.063	0.148***	0.190***
	(0.036)	(0.049)	(0.212)	(0.038)	(0.054)	(0.052)	(0.044)
90th percentile	0.456***	0.808***	3.636***	0.348***	0.544***	–0.280***	0.064
	(0.068)	(0.052)	(0.394)	(0.049)	(0.079)	(0.078)	(0.044)

Notes: The dependent variable in Columns (1), (4), and (5) is the (standardized) migrant skill mix in 2000, i.e., $\ln\left(E^H_{sd}/E^L_{sd}\right) - \ln\left(E^H_s/E^L_s\right)$ divided by the standard deviation. F_{ST} genetic distance is standardized such that it has a standard deviation of 1 over all country pairs in the sample. In Column (2), the dependent variable is the linear standardized genetic distance. In Column (3), the dependent variable is the squared standardized genetic distance. The dependent variables in Columns (6) and (7) are the (standardized) emigration propensities for primary-educated and tertiary-educated migrants, i.e., $\ln\left(E^j_{sd}/E^j_s\right)$ for education level j and divided by the standard deviation. *Control variables:* Log geographic distance, contiguous, Δ absolute latitude, Δ absolute longitude, Δ temperature, Δ precipitation, Δ 80/20 wage ratio, Δ GDP per capita, Levensthein language distance, anglophone destination, common language, migrant networks, travel visa restriction, Schengen pair, colony, log inflow foreigners, log inflow asylum-seekers, Δ years of schooling, and Δ share tertiary. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: ***Nonlinear Effect of Genetic Distance on Migrant Selection: Evidence from Sample Splits by Median Genetic Distance***

	(1)	(2)	(3)	(4)	(5)	(6)
	Selection		Scale primary		Scale tertiary	
	Above	Below	Above	Below	Above	Below
F_{ST} genetic distance$_{sd}$	0.591***	–0.293	–0.372***	0.602***	0.0004	0.424**
	(0.097)	(0.212)	(0.069)	(0.233)	(0.059)	(0.165)
Control variables	yes	yes	yes	yes	yes	yes
R-squared	0.686	0.629	0.681	0.659	0.792	0.713
Observations	551	551	551	551	551	551
Cluster	15	15	15	15	15	15
Kleibergen-Paap F statistic	117.9	66.2	117.9	66.2	117.9	66.2

Notes: The dependent variable in Columns (1) and (2) is the (standardized) migrant skill mix in 2000. The dependent variables in Columns (3) to (6) are the (standardized) emigration propensities for primary-educated and tertiary-educated migrants. *Above* and *Below* indicate sample splits by above and below the median genetic distance. See Table 3 for variable definitions and the list of control variables. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: ***Migration Policy and Robustness Checks***

	(1)	(2)	(3)	(4)
F_{ST} genetic distance$_{sd}$	–0.363***	–0.311***	–0.319***	–0.325***
	(0.121)	(0.108)	(0.095)	(0.092)
F_{ST} genetic distance$_{sd}$, squared	0.176***	0.165***	0.159***	0.161***
	(0.034)	(0.029)	(0.028)	(0.027)
Privileged access to work visas$_{sd}$		0.143*	0.190**	0.202**
		(0.082)	(0.079)	(0.081)
Privileged access to citizenship$_{sd}$		0.087	0.074	0.068
		(0.095)	(0.104)	(0.109)
Average poverty rate$_{s}$, predicted			0.006***	0.005***
			(0.002)	(0.002)
Political freedom$_{s}$				0.032*
				(0.017)
Control variables	yes	yes	yes	yes
R-squared	0.669	0.670	0.639	0.639
Observations	1,102	1,102	1,029	1,009
Cluster	15	15	15	15
Kleibergen-Paap F statistic	42.0	35.0	49.5	51.3
Marginal effects for F_{ST} genetic distance at				
10th percentile	–0.332***	–0.282***	–0.291***	–0.297***
	(0.115)	(0.104)	(0.091)	(0.088)
50th percentile	0.063	0.087	0.066	0.064
	(0.054)	(0.056)	(0.059)	(0.057)
90th percentile	0.544***	–0.537***	0.500***	0.503***
	(0.079)	(0.075)	(0.089)	(0.090)

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: ***Region-Specific Effect Heterogeneities***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F_{ST} genetic distance$_{sd}$	–0.363***	–0.241**	–0.145**	–0.207	–0.377***	–0.217	–0.290
	(0.121)	(0.107)	(0.063)	(0.140)	(0.136)	(0.133)	(0.302)
F_{ST} genetic distance$_{sd}$, squared	0.176***	0.118***	0.077***	0.124***	0.149***	0.109***	0.187***
	(0.034)	(0.030)	(0.018)	(0.036)	(0.041)	(0.036)	(0.070)
Control variables	yes	yes	yes	yes	yes	yes	yes
Excluding AUS, CAN, USA, UK		yes					
EU destinations only			yes				
Destination-country fixed effects (14)				yes			
Sending-continent fixed effects (5)					yes		
Sending-region fixed effects (9)						yes	
Sending-country fixed effects (84)							yes
R-squared	0.669	0.615	0.587	0.716	0.679	0.708	0.804
Observations	1,102	771	618	1,102	1,102	1,102	1,102
Cluster	15	11	9	15	15	15	15
Kleibergen-Paap F statistic	42.0	840.5	623.1	35.8	48.8	22.1	2.6
Marginal effects for F_{ST} genetic distance at							
10th percentile	–0.332***	–0.221**	–0.131**	–0.185	–0.351***	–0.198	–0.257
	(0.115)	(0.101)	(0.060)	(0.134)	(0.129)	(0.128)	(0.292)
50th percentile	0.063	0.043	0.041	0.092	–0.018	0.048	0.161
	(0.054)	(0.040)	(0.030)	(0.064)	(0.057)	(0.087)	(0.197)
90th percentile	0.544***	0.364***	0.251***	0.429***	0.388***	0.347***	0.670***
	(0.079)	(0.056)	(0.044)	(0.067)	(0.096)	(0.119)	(0.222)

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables and Appendix Table A-1 for continents and regions. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

A Appendix

Table A-1: ***Sending Countries by Continent***

(1)		(2)		(3)		(4)		(5)		(6)	
North America ($N = 12$)		South America ($N = 9$)		Asia ($N = 20$)		Europe ($N = 27$)		Africa ($N = 15$)		Oceania ($N = 2$)	
Country	Region	Country	Region	Country	Region	Country	Region	Country	Region	Country	Region
Canada	NAM	Bolivia	LAC	Armenia	CAS	Austria	WEU	Botswana	SSA	Australia	EAP
Costa Rica	LAC	Brazil	LAC	Bangladesh	SAS	Belgium	WEU	Cameroon	SSA	New Zealand	EAP
Dominican Republic	LAC	Chile	LAC	China	EAP	Bulgaria	CEU	Central Afr. Rep.	SSA		
El Salvador	LAC	Colombia	LAC	Hong Kong	EAP	Croatia	CEU	Egypt	MENA		
Guatemala	LAC	Ecuador	LAC	Indonesia	EAP	Denmark	SCA	Gambia	SSA		
Honduras	LAC	Guyana	LAC	Israel	MENA	Estonia	CEU	Ghana	SSA		
Jamaica	LAC	Paraguay	LAC	Japan	EAP	Finland	SCA	Lesotho	SSA		
Mexico	LAC	Peru	LAC	Jordan	MENA	France	WEU	Mali	SSA		
Nicaragua	LAC	Venezuela	LAC	Kazakhstan	CAS	Germany	WEU	Mauritius	SSA		
Panama	LAC			Korea	EAP	Greece	CEU	Senegal	SSA		
Trinidad and Tobago	LAC			Kyrgyzstan	CAS	Hungary	CEU	South Africa	SSA		
USA	NAM			Malaysia	EAP	Ireland	WEU	Swaziland	SSA		
				Nepal	SAS	Italy	WEU	Uganda	SSA		
				Pakistan	SAS	Latvia	CEU	Zambia	SSA		
				Philippines	EAP	Lithuania	CEU	Zimbabwe	SSA		
				Singapore	EAP	Luxembourg	WEU				
				Tajikistan	CAS	Netherlands	WEU				
				Thailand	EAP	Norway	SCA				
				Turkey	CAS	Poland	CEU				
				Vietnam	EAP	Portugal	WEU				
						Romania	CEU				
						Slovakia	CEU				
						Slovenia	CEU				
						Spain	WEU				
						Sweden	SCA				
						Ukraine	CEU				
						UK	WEU				

Notes: Regions: SSA: Sub-Saharan Africa, MENA: Middle East and North Africa, SAS: South Asia, CAS: Central Asia, EAP: East Asia and Pacific, LAC: Latin America and Caribbean, NAM: North America, CEU: Central Europe, WEU: Western Europe, SCA: Scandinavia.

Table A-2: ***Omitting Destination Countries***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	AUS	AUT	CAN	DNK	FIN	FRA	DEU	IRL	NLD	NZL	NOR	ESP	SWE	USA	UK
F_{ST} genetic distance$_{sd}$	–0.324**	–0.366***	–0.322***	–0.346***	–0.407***	–0.368***	–0.346***	–0.324***	–0.358***	–0.356***	–0.353***	–0.274**	–0.370***	–0.381***	–0.421***
	(0.133)	(0.130)	(0.123)	(0.126)	(0.118)	(0.136)	(0.129)	(0.123)	(0.130)	(0.118)	(0.135)	(0.112)	(0.135)	(0.140)	(0.117)
F_{ST} genetic distance$_{sd}$, squared	0.160***	0.177***	0.164***	0.175***	0.198***	0.179***	0.173***	0.167***	0.169***	0.173***	0.174***	0.150***	0.179***	0.164***	0.190***
	(0.040)	(0.036)	(0.037)	(0.035)	(0.027)	(0.038)	(0.036)	(0.034)	(0.036)	(0.032)	(0.036)	(0.032)	(0.037)	(0.040)	(0.034)
Control variables	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.653	0.667	0.655	0.665	0.679	0.683	0.669	0.678	0.668	0.666	0.676	0.698	0.671	0.673	0.665
Observations	1,020	1,053	1,021	1,021	1,027	1,018	1,025	1,058	1,018	1,030	1,021	1,037	1,043	1,018	1,018
Cluster	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Kleibergen-Paap F statistic	62.3	37.7	33.8	37.5	45.3	41.5	38.1	37.3	37.4	44.4	36.5	37.5	37.4	39.6	40.3

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables. The country in the column header is omitted as a destination country. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-3: ***Fixed Effects Specifications***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline		Destination		Source		Destination & source	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
F_{ST} genetic distance$_{sd}$	–0.240***	–0.363***	–0.066	–0.207	–0.183	–0.290	–0.101	–0.130
	(0.078)	(0.121)	(0.072)	(0.140)	(0.165)	(0.302)	(0.143)	(0.283)
F_{ST} genetic distance$_{sd}$, squared	0.135***	0.176***	0.085***	0.124***	0.074*	0.187***	0.062*	0.166***
	(0.025)	(0.034)	(0.019)	(0.036)	(0.035)	(0.070)	(0.030)	(0.062)
Control variables	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.671	0.669	0.718	0.716	0.814	0.804	0.820	0.810
Observations	1,102	1,102	1,102	1,102	1,102	1,102	1,102	1,102
Cluster	15	15	15	15	15	15	15	15
Kleibergen-Paap F statistic		42.0		35.8		2.6		2.0
Marginal effects for F_{ST} genetic distance at								
10th percentile	–0.217**	–0.332***	–0.050	–0.185	–0.170	–0.257	–0.090	–0.101
	(0.074)	(0.115)	(0.068)	(0.134)	(0.160)	(0.292)	(0.139)	(0.276)
50th percentile	0.087**	0.063	0.141***	0.092	–0.005	0.161	0.049	0.271
	(0.036)	(0.054)	(0.036)	(0.064)	(0.104)	(0.197)	(0.095)	(0.211)
90th percentile	0.456***	0.544***	0.374***	0.429***	0.196*	0.670***	0.218**	0.724***
	(0.068)	(0.079)	(0.047)	(0.067)	(0.098)	(0.222)	(0.095)	(0.240)

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

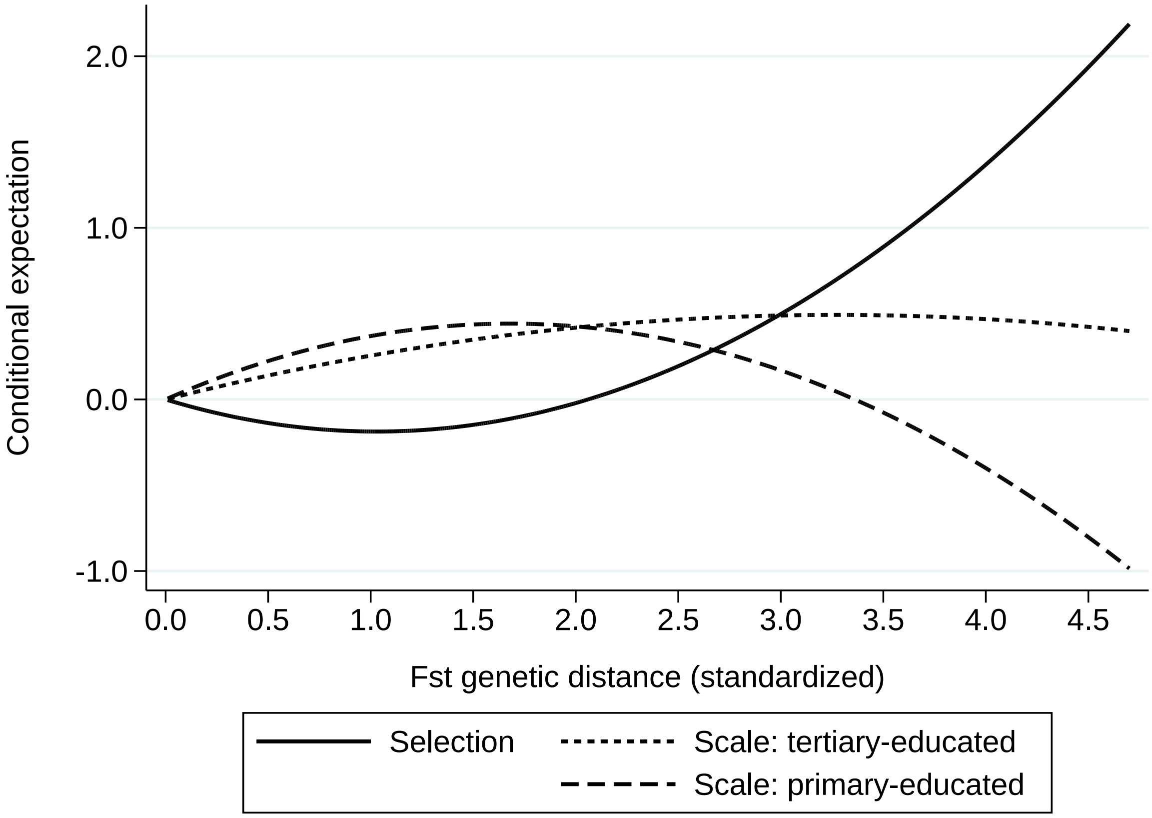
Table A-4: ***Omitting Sending Regions***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	SSA	MENA	SAS	CAS	EAP	LAC	NAM	CEU	WEU	SCA
F_{ST} genetic distance$_{sd}$	–0.507***	–0.368***	–0.339***	–0.290**	–0.347***	–0.444***	–0.270**	–0.415**	–0.231	–0.376***
	(0.148)	(0.116)	(0.131)	(0.131)	(0.118)	(0.099)	(0.118)	(0.163)	(0.156)	(0.131)
F_{ST} genetic distance$_{sd}$, squared	0.198**	0.174***	0.178***	0.160***	0.174***	0.187***	0.157***	0.185***	0.141***	0.178***
	(0.079)	(0.034)	(0.037)	(0.035)	(0.032)	(0.032)	(0.033)	(0.042)	(0.040)	(0.036)
Control variables	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.558	0.671	0.660	0.682	0.704	0.712	0.659	0.676	0.677	0.666
Observations	956	1,059	1,061	1,043	939	868	1,074	925	947	1,046
Cluster	15	15	15	15	15	15	15	15	15	15
Kleibergen-Paap F statistic	8.2	54.6	39.5	39.0	30.3	31.5	37.4	25.2	32.2	36.6

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables and Appendix Table A-1 for regions. The world region in the column header is omitted as a sending region. SSA: Sub-Saharan Africa, MENA: Middle East and North Africa, SAS: South Asia, CAS: Central Asia, EAP: East Asia and Pacific, LAC: Latin America and Caribbean, NAM: North America, CEU: Central Europe, WEU: Western Europe, SCA: Scandinavia. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

B Online Appendix: Further Results

Figure B-1: ***Conditional Expectation of Emigrant Selection***



Notes: The figure shows the conditional expectation of the migrant skill mix, i.e. $\mathbb{E}\left[\ln\left(E_{sd}^{H}/E_{sd}^{L}\right)-\ln\left(E_{s}^{H}/E_{s}^{L}\right)|\mathbf{X}\right]$, and the conditional expectation for the scale of emigration by education level, i.e., $\mathbb{E}\left[E_{sd}^{j}/E_{s}^{j}|\mathbf{X}\right]$ with $j \in$ {primary-educated, tertiary-educated}. To construct the figure, we use the estimates that condition on the full set of control variables from Columns (5), (6), and (7) of Table 3. Confidence intervals are omitted for expositional purposes.

Table B-1: ***Variable Definitions and Sources***

Variable	Definition	Source
F_{ST} genetic distance	Average variation in the frequencies of 120 alleles between two populations matched to countries according to majority populations.	Spolaore and Wacziarg (2009)
F_{ST} genetic distance, 1500	Average variation in the frequencies of 120 alleles between two populations matched to countries, with majority populations as of 1500.	Spolaore and Wacziarg (2009)
Migration data	Data on migrant stocks (aged 25 and above) in 2000 with skill levels.	Docquier et al. (2007)
Log geographic distance	Population-weighted great circle distance between large cities of the two countries.	Head et al. (2010)
Contiguous	Dummy equal to 1 if both countries share a common border.	Head et al. (2010)
Δ absolute latitude	Absolute value of the latitude of a country's approximate geodesic centroid.	Ashraf and Galor (2013)
Δ absolute longitude	Absolute value of the longitude of a country's approximate geodesic centroid.	Ashraf and Galor (2013)
Δ temperature	Difference in average monthly temperatures between sending and destination country, measured in degrees Celsius from 1961-1990.	Ashraf and Galor (2013)
Δ precipitation	Difference in average monthly precipitation between sending and destination country, measured in degrees Celsius from 1961-1990.	Ashraf and Galor (2013)
Levensthein language distance	Global percentage of dissimilarity in the pronunciation of words with the same meaning in two languages, the value is averaged over 40 words. Phonetic similarity is automatically evaluated by the Automatic Similarity Judgement program (ASJP).	Isphording and Otten (2013)
Cladistic language distance	Distance measure based on linguistic proximity that assess the number of common nodes two languages have. We use the weighted measure that accounts for linguistic heterogeneity. Linguistic trees are from Ethnologue and data on spoken languages from Fearon (2003).	Spolaore and Wacziarg (2016b)
Anglophone destination	Dummy equal to 1 if English is the first official language.	Own research
Common language	Dummy equal to 1 if destination and sending country share a language that is spoken by at least 9 per cent of the population.	Head et al. (2010)
Migrant networks	Ration of the stock of migrants from a sending country summed over all education levels to the residents in the sending country summed over all education levels in the year 1990.	Own calculations with the data from Docquier et al. (2007)
Δ 80/20 wage ratio	Difference in wage differences, i.e., between high- and low-skilled wages.	Grogger and Hanson (2011)
Δ GDP per capita	Difference in GDP per capita, not deflated. Calculated on means between 1980 and 2000.	Own calculations with data from Head et al. (2010)
Travel visa restriction	Dummy equal to 1 if visa restrictions on travel are imposed by the destination on a sending country.	Neumayer (2006)
Schengen pair	Dummy equal to 1 if both countries are signatories of the Schengen agreement.	Own research
Colony	Dummy equal to 1 if the countries have ever been in a colonial relationship.	Head et al. (2010)
Inflow foreigners	Inflow of foreign population from 216 sending countries in 1999, measured in 1000s.	Own calculation, data from the International Migration Dataset, OECD
Inflow asylum seekers	Inflow of asylum seekers from 216 sending countries in 1999, based on data provided by the United Nations High Commissioner for Refugees.	Own calculation, data from the International Migration Dataset, OECD
Δ years of schooling	Difference in the average years of schooling attained.	Barro and Lee (2013)
Δ share tertiary	Difference in the percentage of completed tertiary education in population.	Barro and Lee (2013)
Privileged access to work visas	Dummy equal to 1 if the destination country offers privileged access to work visas for citizens of the sending country.	Own research
Privileged access to citizenship	Dummy equal to 1 if the destination country offers privileged access to citizenship for citizens of the sending country.	Own research
Political freedom	Index between 1 and 7 measuring the degree of political freedom in the sending country. 1 is free, 7 is not free.	Freedom House Index 1999-2000
Average poverty rate, predicted	Prediction of the average poverty rate in sending countries by the regression of the average poverty rate (the share of population living on less than two US dollars per day) on the average share of employees in the agricultural sector. Both are averaged over the period 1980-2000. Calculation as by Belot and Hatton (2012).	Data from World Bank Development Indicators
Δ share agriculture	Difference in the value added as percentage of GDP of the agricultural sector between destination and sending country.	Own calculation, data from the WDI
Δ share industry	Difference in the value added as percentage of GDP of the industrial sector between destination and sending country.	Own calculation, data from the WDI
Δ share service	Difference in the value added as percentage of GDP of the service sector between destination and sending country.	Own calculation, data from the WDI
Δ share Protestants	Difference in percentage of the Protestant population.	Ashraf and Galor (2013)
Δ share Catholics	Difference in percentage of the Catholic population.	Ashraf and Galor (2013)
Δ share Muslims	Difference in percentage of the Muslim population.	Ashraf and Galor (2013)
Δ share other religions	Difference in percentage of the population belonging to any other religion or denomination than Catholic, Protestant or Muslim.	Ashraf and Galor (2013)
Δ distance Addis Ababa	Difference in migratory distance to East Africa. Calculated as the great circle distance from Addis Ababa in East Africa, Ethopia, to the capital of each country as long as possible over land and following specified waypoints. Measured in thousands of km.	Ashraf and Galor (2013)

Notes: Δ represents the simple difference between destination and sending country, that is, $\Delta X = X_d - X_s$.

Table B-2: ***Alternative Genetic Distance Measure***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Linear				Squared								
	OLS		IV		OLS		IV						
									Polynomials in first stage				
									3	4	5	6	6
Cavalli-Sforza et al. (1994)													
F_{ST} genetic distance$_{sd}$	0.234***		0.321***		–0.240***		–0.363***					–0.418***	
	(0.038)		(0.068)		(0.078)		(0.121)					(0.112)	
F_{ST} genetic distance$_{sd}$, squared					0.135***		0.176***					0.187***	
					(0.025)		(0.034)					(0.035)	
Pemberton et al. (2013)													
F_{ST} genetic distance$_{sd}$		0.264***		0.282***		–0.516***		–1.875**	–1.792***	–0.498***	–0.419**		–0.440**
		(0.037)		(0.095)		(0.150)		(0.861)	(0.641)	(0.178)	(0.186)		(0.214)
F_{ST} genetic distance$_{sd}$, squared						0.236***		0.687**	0.660***	0.206***	0.178***		0.186***
						(0.038)		(0.299)	(0.225)	(0.054)	(0.055)		(0.066)
Control variables	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.650	0.649	0.646	0.649	0.671	0.670	0.669	0.585	0.595	0.667	0.665	0.668	0.666
Observations	1,102	1,102	1,102	1,102	1,102	1,102	1,102	1,102	1,102	1,102	1,102	1,102	1,102
Cluster	15	15	15	15	15	15	15	15	15	15	15	15	15
Kleibergen-Paap F statistic			301.9	173.2			42.0	4.6	5.0	33.7	74.2	1,137.2	61.8
Hansen J statistic *(p-value)*									0.929	0.042	0.089	0.126	0.164
Marginal effects for F_{ST} genetic distance at													
10th percentile					–0.217**	–0.478***	–0.332***	–1.764**	–1.686***	–0.465***	–0.390**	–0.386***	–0.410**
					(0.074)	(0.144)	(0.115)	(0.812)	(0.605)	(0.170)	(0.177)	(0.106)	(0.204)
50th percentile					0.087**	–0.051	0.063	–0.520*	–0.492**	–0.092	–0.068	0.033	–0.073
					(0.036)	(0.078)	(0.054)	(0.279)	(0.209)	(0.091)	(0.094)	(0.047)	(0.099)
90th percentile					0.456***	0.819***	0.544***	2.014**	1.942***	0.668***	0.589***	0.543***	0.615***
					(0.068)	(0.076)	(0.079)	(0.842)	(0.643)	(0.159)	(0.156)	(0.090)	(0.184)

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables. Measures for genetic distance based on data from Cavalli-Sforza et al. (1994) at the country level are provided by Spolaore and Wacziarg (2009) and measures for genetic distance based on data from Pemberton et al. (2013) at the country level are provided by Spolaore and Wacziarg (2016a). Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table B-3: ***Correlations for Selected Variables***

Variables	F_{ST} genetic distance	Levensthein language distance	Common language	Log distance	Contiguity	Colony	Δ share Catholics	Δ share Protes-tants	Δ share Muslims	Δ share other religions	Δ distance Addis Ababa
F_{ST} genetic distance	1										
Levensthein language distance	0.2501	1									
Common language	0.1870	-0.4384	1								
Log distance	0.4333	0.0672	0.1788	1							
Contiguity	-0.1581	-0.1649	0.0637	-0.3945	1						
Colony	0.0459	-0.3291	0.3985	0.0036	0.1051	1					
Δ share Catholics	0.0728	0.0226	0.0645	-0.0353	-0.0097	0.0465	1				
Δ share Protestants	0.0745	0.2405	-0.1876	0.0635	-0.0947	-0.1749	-0.5649	1			
Δ share Muslims	-0.0747	-0.2074	-0.0148	-0.0800	0.0824	0.0174	-0.3157	-0.1549	1		
Δ share other religions	-0.1214	-0.1326	0.1257	0.0391	0.0528	0.1121	-0.4973	-0.1992	-0.1607	1	
Δ distance Addis Ababa	0.0241	0.1149	0.1783	-0.0326	0.0119	-0.0941	0.2815	-0.1502	-0.2066	-0.0612	1

Notes: Δ represents the simple difference between destination and sending country, that is, $\Delta X = X_d - X_s$.

Table B-4: ***Nonlinearities in Geographical Distance***

	(1)	(2)	(3)	(4)
F_{ST} genetic distance$_{sd}$	–0.363***	–0.387***	–0.368***	–0.361***
	(0.121)	(0.116)	(0.125)	(0.131)
F_{ST} genetic distance$_{sd}$, squared	0.176***	0.180***	0.177***	0.175***
	(0.034)	(0.034)	(0.036)	(0.037)
Log geographic distance$_{sd}$	–0.005			
	(0.048)			
Geographic distance$_{sd}$		0.009	–0.041	–0.096
		(0.038)	(0.144)	(0.246)
Geographic distance$_{sd}$, squared			0.013	0.047
			(0.030)	(0.131)
Geographic distance$_{sd}$, cubic				–0.006
				(0.021)
Control variables	yes	yes	yes	yes
R-squared	0.669	0.668	0.668	0.668
Observations	1,102	1,102	1,102	1,102
Cluster	15	15	15	15
Kleibergen-Paap F statistic	42.0	37.1	45.7	39.8
Marginal effects for F_{ST} genetic distance at				
10th percentile	–0.332***	–0.355***	–0.337***	–0.330***
	(0.115)	(0.111)	(0.119)	(0.125)
50th percentile	0.063	0.049	0.060	0.062
	(0.054)	(0.051)	(0.050)	(0.053)
90th percentile	0.544***	0.541***	0.542***	0.540***
	(0.079)	(0.080)	(0.078)	(0.077)

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions. Control variables: Contiguous, Δ absolute latitude, Δ absolute longitude, Δ temperature, Δ precipitation, Δ 80/20 wage ratio, Δ GDP per capita, Levensthein language distance, anglophone destination, common language, migrant networks, travel visa restriction, Schengen pair, colony, log inflow foreigners, log inflow asylum-seekers, Δ years of schooling, Δ share tertiary. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table B-5: ***Further (Potential) Mechanisms and Robustness Checks***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F_{ST} genetic distance$_{sd}$	–0.363***	–0.311***	–0.319***	–0.325***	–0.303***	–0.316***	–0.338***
	(0.121)	(0.108)	(0.095)	(0.092)	(0.102)	(0.094)	(0.103)
F_{ST} genetic distance$_{sd}$, squared	0.176***	0.165***	0.159***	0.161***	0.145***	0.159***	0.168***
	(0.034)	(0.029)	(0.028)	(0.027)	(0.032)	(0.030)	(0.033)
Privileged access to work visas$_{sd}$		0.143*	0.190**	0.202**	0.221***	0.256***	0.256***
		(0.082)	(0.079)	(0.081)	(0.082)	(0.079)	(0.080)
Privileged access to citizenship$_{sd}$		0.087	0.074	0.068	–0.050	–0.057	–0.037
		(0.095)	(0.104)	(0.109)	(0.092)	(0.085)	(0.080)
Average poverty rate$_s$, predicted			0.006***	0.005***	0.005***	0.001	0.003
			(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Political freedom$_s$				0.032*	0.025	0.050***	0.050***
				(0.017)	(0.015)	(0.014)	(0.014)
Δ share Protestants$_{sd}$					–0.004***	–0.004***	–0.004**
					(0.001)	(0.001)	(0.002)
Δ share Muslims$_{sd}$					–0.001	–0.001	–0.002
					(0.002)	(0.002)	(0.002)
Δ share other religion$_{sd}$					–0.001	–0.001	–0.002*
					(0.001)	(0.001)	(0.001)
Δ share industry$_{sd}$						0.021***	0.019***
						(0.004)	(0.005)
Δ share service$_{sd}$						0.012***	0.011***
						(0.004)	(0.004)
Δ distance Addis Ababa$_{sd}$							–0.115
							(0.090)
Control variables	yes	yes	yes	yes	yes	yes	yes
R-squared	0.669	0.670	0.639	0.639	0.653	0.663	0.663
Observations	1,102	1,102	1,029	1,009	994	942	942
Cluster	15	15	15	15	15	15	15
Kleibergen-Paap F statistic	42.0	35.0	49.5	51.3	47.2	58.0	59.3
Marginal effects for F_{ST} genetic distance at							
10th percentile	–0.332***	–0.282***	–0.291***	–0.297***	–0.278***	–0.288***	–0.308***
	(0.115)	(0.104)	(0.091)	(0.088)	(0.098)	(0.090)	(0.099)
50th percentile	0.063	0.087	0.066	0.064	0.046	0.067	0.067
	(0.054)	(0.056)	(0.059)	(0.057)	(0.069)	(0.068)	(0.072)
90th percentile	0.544***	–0.537***	0.500***	0.503***	0.441***	0.500***	0.525***
	(0.079)	(0.075)	(0.089)	(0.090)	(0.111)	(0.112)	(0.119)

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

C Online Appendix: Missing Data

We do not observe the migrant stock for all country pairs, either because of missing data or because the migrant stock of the sending country in the destination country is so small that it is not reported in the statistics. Out of 1,260 potential country pairs, we do not observe 158 pairs because of these missing values. The destination country with the greatest lack of information is Ireland (40 sending countries), followed by Austria (35), Sweden (25), and Spain (19). In Appendix Table A-2, we do not find that omitting these countries leads to significant changes in the results. However, missing country pairs are not similar to included country pairs. For example, the genetic distance is much larger for the omitted country pairs (see Table C-1).

In Table C-2, we provide further analyses to verify the robustness of the results regarding missing data. We bring the 158 country pairs back into the sample by imputing different values for the migrant skill mix. First, we transfer information from the migrant skill mix in 1990. This gives us information for nine missing values. Second, we impute values for the remaining 149 pairs using different approaches. In Column (2) of Table C-2, we use all covariates (excluding genetic distance) to predict the migrant skill mix. In Column (3), we use the destination-country-specific regional average of the migrant skill mix for imputation.[37] Both robustness checks yield similar results. In Columns (4) to (6), we impute destination-country-specific percentiles of the migrant skill mix. This exercise yields bounds on the baseline effect. Given that the genetic distance of the excluded country pairs is larger compared to the included country pairs, we expect that a migrant skill mix that is biased towards high-skilled migrants better fits the data. In line with this, we find that imputing the 90th percentile (Column (6)) performs much better than imputing the 50th percentile (Column (5)) or the 10th percentile (Column (4)). However, even in the scenario that all missing country pairs are drawn from the destination-country-specific 10th migrant skill mix percentile, we still find a substantial nonlinearity that is not very different from the baseline effect.

[37] Compare Appendix Table A-1 for the categorization of sending countries to regions.

Table C-1: ***Comparison of Missing Data***

	(1)	(2)	(3)	(4)
Variable	Averages		Difference	p-value
	Included	Excluded		
Fst genetic distance	1.252	2.072	-0.820***	0.000
Fst genetic distance, 1500	1.522	2.187	-0.665***	0.000
Log geographic distance	8.541	8.947	-0.406***	0.000
Contiguous	0.030	0.000	0.030	0.853
Δ absolute latitude	-9.274	-28.928	19.654***	0.000
Δ absolute longitude	18.738	31.166	-12.428***	0.000
Δ temperature	-8.545	-12.950	4.404***	0.000
Δ precipitation	-27.232	-34.401	7.170	0.162
Δ 80/20 wage ratio	22.632	21.120	1.512*	0.070
Δ GDP per capita	15,856	18,996	-3,139***	0.003
Anglophone destination	0.406	0.361	0.045	0.786
Common language	0.166	0.133	0.033	0.839
Levensthein language distance	86.732	91.432	-4.700**	0.018
Migrant networks	0.004	0.000	0.004	0.981
Travel visa restriction	0.477	0.646	-0.168	0.309
Schengen pair	0.159	0.032	0.127	0.436
Colony	0.064	0.000	0.064	0.690
Log inflow foreigners	11.286	10.703	0.584***	0.002
Log inflow asylum-seekers	9.670	9.201	0.469**	0.011
Δ years of schooling	2.637	3.801	-1.164***	0.000
Δ share tertiary	6.116	8.737	-2.620***	0.000
Observations	1,102	158		

Notes: Table shows averages for main explanatory variables by sample status. *Difference* reports the simple difference in sample means. The *p-value* reports p-values from a two-sample *t*-test. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table C-2: ***Sensitivity to Missing Data***

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Covar	Region	10th	50th	90th
F_{ST} genetic distance$_{sd}$	–0.363***	–0.267*	–0.409**	–0.374	–0.345*	–0.358**
	(0.121)	(0.141)	(0.164)	(0.234)	(0.197)	(0.160)
F_{ST} genetic distance$_{sd}$, squared	0.176***	0.135***	0.181***	0.129**	0.142***	0.164***
	(0.034)	(0.038)	(0.044)	(0.064)	(0.054)	(0.045)
Control variables	yes	yes	yes	yes	yes	yes
R-squared	0.669	0.670	0.662	0.460	0.577	0.630
Observations	1,102	1,260	1,260	1,260	1,260	1,260
Cluster	15	15	15	15	15	15
Kleibergen-Paap F statistic	42.0	28.4	28.4	28.4	28.4	28.4
Marginal effects for F_{ST} genetic distance at						
10th percentile	–0.332***	–0.241*	–0.375**	–0.350	–0.318*	–0.327**
	(0.115)	(0.135)	(0.156)	(0.222)	(0.188)	(0.152)
50th percentile	0.063	0.049	0.015	–0.071	–0.014	0.027
	(0.054)	(0.070)	(0.074)	(0.105)	(0.090)	(0.072)
90th percentile	0.544***	0.466***	0.574***	0.328**	0.424***	0.534***
	(0.079)	(0.096)	(0.101)	(0.146)	(0.130)	(0.111)

Notes: The dependent variable is the (standardized) migrant skill mix in 2000. See Table 3 for variable definitions and the list of control variables. The migrant skill mix and genetic distance is standardized over the set of countries in the specific sample. *Covar* uses imputed migrant skill mix that is predicted based on the full set of covariates. *Region* uses imputed migrant skill mix based on destination-specific averages by sending region. *10th* uses imputed migrant skill mix based on destination-specific 10th percentile. *50th* uses imputed migrant skill mix based on destination-specific 50th percentile. *90th* uses imputed migrant skill mix based on destination-specific 90th percentile. Robust standard errors in parentheses clustered at the destination country level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

D Online Appendix: Migration Policies

Table D-1: ***Migration Policies by Destination Country: Overview***

Country	Policy	Regulation by nationality		Regulation by skills	
		Yes/No	Nationality	Yes/No	Skill group
Australia	Citizenship	No	–	No	–
	Work Visa	Yes	New Zealand	Yes	Skilled workers
Austria	Citizenship	No	–	No	–
	Work Visa	Yes	EU countries, Switzerland	Yes	Qualified third-country workers
Canada	Citizenship	No	–	No	–
	Work Visa	(Yes)	For **temporary** visits of business people under Free Trade Agreements.	Yes	Skilled workers
Denmark	Citizenship	Yes	Nordic countries	No	–
	Work Visa	Yes	Nordic countries, EU/EEA and Swiss citizens, special rights for economically active Turkish nationals.	Yes	Professions with a current shortage, skilled workers
Finland	Citizenship	Yes	Nordic countries	No	–
	Work Visa	Yes	Nordic countries, EU/EEA countries, Switzerland, Liechtenstein	Yes	Certain professions
France	Citizenship	No	–	No	–
	Work Visa	Yes	EU/EEA countries, Switzerland	Yes	Skills or talents
Germany	Citizenship	Yes	Ethnic German resettlers	No	–
	Work Visa	Yes	EU/EEA countries, Switzerland, Jewish immigrants	Yes	Qualified professionals, high-qualified workers
Ireland	Citizenship	No	–	No	–
	Work Visa	Yes	EU/EEA countries, Switzerland	Yes	Strategically important skills
Netherlands	Citizenship	No	–	No	–
	Work Visa	Yes	EU/EEA countries, Switzerland	Yes	Skilled workers and certain professions
New Zealand	Citizenship	Yes	Samoa, Cook Islands, Niue or Tokelau	No	–
	Work Visa	No	–	Yes	High-skilled workers, highly-skilled young persons, certain professions
Norway	Citizenship	Yes	Nordic countries	No	–
	Work Visa	Yes	Nordic countries, EU/EEA countries, Switzerland	Yes	Skilled workers and certain professions
Spain	Citizenship	Yes	Ibero-American countries, Andorra, the Philippines, Equatorial Guinea, and Portugal, and persons of Sephardic origin	No	–
	Work Visa	Yes	EU/EEA countries, Switzerland	No	–
Sweden	Citizenship	Yes	Nordic countries	No	–
	Work Visa	Yes	Nordic countries, EU/EEA countries, Switzerland (18 to 30-year-olds for one year from Australia, Canada, Chile, New Zealand and South Korea)	Yes	Certain professions
United Kingdom	Citizenship	Yes	British Nationals from overseas territories or Hong Kong and Gibraltar, Commonwealth citizens under conditions regarding **ancestry**	No	–
	Work Visa	Yes	EU/EEA countries, Switzerland, Commonwealth citizens under conditions regarding **ancestry**	Yes	Point-system based on occupation categories
United States	Citizenship	No	–	Yes	Non-US citizen members of US armed forces, Green Card holders
	Work Visa	No	–	Yes	Job skills, certain professions

Notes: The table summarizes citizenship and work visa migration policy regulations by nationality and skill group in each of the destination countries according to authors' categorizations based on country-specific information in Table D-2. Last update: March 2017.

Table D-2: ***Migration Policies by Destination Country: Details***

Country	Policy	Regulation by nationality	Regulation by skills	Sources
Australia	Citizenship	No.	No.	`https://www.border.gov.au/Citizenship`
Department of Immigration and Border Protection	Work Visa	Yes: Special Category Visa (SCV) for **New Zealand citizens**, technically temporary but treated like permanent visas.	**Skilled visa** via SkillSelect, taking into account occupation, work experience, education, language skills.	`http://www.australia.gov.au/information-and-services/immigration-and-visas` and `https://www.border.gov.au/Trav/Work/Skil`
	Comments	Skilled Independent visa based on points-tested skilled workers without sponsor (family, employer, government).		`http://www.border.gov.au/Trav/Visa-1/189-`
Austria	Citizenship	No.	No.	`http://www.migration.gv.at/en/living-and-working-in-austria/integration-and-citizenship/citizenship/`
Bundesministerium für Arbeit, Soziales, und Konsumentenschutz; Bundesministerium für Inneres, Bundesministerium für Europa, Integration und Äußeres	Work Visa	Yes: **Free movement for EEA** (transitional requirements for Croatia) and Swiss citizens or their spouses conditional on being employed or able to support themselves.	Yes: Red-White-Red Card (since 2011) uses a point system for **qualified third-country workers** and their families, aimed at: very highly qualified workers, shortage occupations, key workers and graduates of Austrian higher education.	`http://www.migration.gv.at/en/types-of-immigration/mobility-within-the-eu/` and `http://www.migration.gv.at/en/types-of-immigration/permanent-immigration/`
	Comments	Blue Card from the EU is applicable for third-country nationals.		`http://www.migration.gv.at/en/types-of-immigration/permanent-immigration/eubluecard/`
Canada	Citizenship	No.	No.	`http://www.cic.gc.ca/english/citizenship/become-eligibility.asp`
Immigration, Refugees and Citizenship Canada	Work Visa	Yes: Eased entry (for temporary work visits) for business people who are **citizens under NAFTA, other Free Trade Agreements, General Agreement on Trade in Services**.	Yes: Special immigration programs for: **skilled workers**, Quebec-selected skilled workers, business founders , investors, self-employed persons in cultural, athletics or farming, sponsored by relatives, nominated by a Canadian province or territory, by graduating from a Canadian school, for caregivers and refugees.	`http://www.cic.gc.ca/english/work/special-business.asp` and `http://www.cic.gc.ca/english/immigrate/apply.asp`
	Comments	For skilled workers (entering under Express Entry) the Comprehensive Ranking System uses a point system that looks at skills, work experience, language ability, education and other factors.		`http://www.cic.gc.ca/english/express-entry/criteria-crs.asp`
Denmark	Citizenship	Yes: **Nordic countries (Denmark, Finland, Iceland, Norway, Sweden)**	No.	`http://uim.dk/arbejdsomrader/statsborgerskab/tidligere-danske-statsborgere-og-nordiske-statsborgere-1`
The Danish Agency for International Recruitment and Integration	Work Visa	Yes: **Nordic countries (Denmark, Finland, Iceland, Norway, Sweden), EU/EEA and Swiss citizens, special rights for economically active Turkish nationals.**	Yes: Under the positive list, **professions with a current shortage** can enter, the pay limit scheme eases entry for persons with an annual pay above the limit, the fast-track scheme is available for certified employers, favorable conditions for (guest) researchers, start-up founders, Master or PhD students, special individual qualifications (artists, athletes, chefs) herdsmen and farm managers, employees on moveable oil rigs or drillships.	`https://www.nyidanmark.dk/en-us/coming_to_dk/work/work.htm`
	Comments	Further (short-term) work permits are for certain professions available, e.g. trainees, 'fitters' for use of technical systems. Denmark does not participate in the EU Blue Card System.		`https://www.nyidanmark.dk/en-us/coming_to_dk/work/work.htm`

Last update: March 2017. *(continued on next page)*

Table D-2 (continued)

Country	Policy	Regulation		Sources
		by nationality	by skills	
Finland	Citizenship	Yes: **Nordic countries (Denmark, Finland, Iceland, Norway, Sweden)**	No.	http://www.migri.fi/finnish_citizenship/applying_for_citizenship
Ministry of the Interior	Work Visa	Yes: **Nordic countries (Denmark, Finland, Iceland, Norway, Sweden), EU/EEA and Swiss or Liechtenstein citizens** can stay and work freely for up to three month, need to register afterward.	Normally: Residence permit for an employed person is required (job offer required). No residence permit for employed persons is required for **certain professions** (specialists, researchers, religious or non-profit associations' employees, athletes or coaches, middle or top management, persons with degree or qualification in Finland, work in science, culture or arts, international organizations' employees, mass media employees, for preparing a company's location in Finland.	http://www.migri.fi/information_elsewhere/eu_and_nordic_citizens and http://www.migri.fi/working_in_finland/an_employee_and_work http://www.migri.fi/working_in_finland/an_employee_and_work/residence_permit_for_an_employed_person
	Comments	Right to work without a residence permit for a list of professions for up to 90 days. EU Blue Card is applicable for third-country nationals.		http://www.migri.fi/working_in_finland/right_to_work_without_a_residence_permit
France	Citizenship	No.	No.	http://www.immigration.interieur.gouv.fr/Accueil-et-accompagnement/La-nationalite-francaise
Ministère de l'Intérieur	Work Visa	Yes: Free movement for **EU/EEA and Swiss citizens**	3 year residence permits for **skills or talents** who make a significant and lasting contribution to the (economic, intellectual, scientific, cultural, humanitarian or athletic) development of France, most countries of the ZSP (Priority Solidarity Zone) need to commit to return after 6 years.	http://www.immigration.interieur.gouv.fr/fr/Immigration
	Comments	Blue Card from the EU is applicable for third-country nationals.		http://www.migration.gv.at/en/types-of-immigration/permanent-immigration/eubluecard/
Germany	Citizenship	Yes: **Ethnic German resettlers**, recognition implies automatic German nationality.	No.	http://www.bamf.de/EN/Willkommen/Einbuergerung/einbuergerung-node.html
Federal Office for Migration and Refugees	Work Visa	Yes: Free movement for **EU/EEA and Swiss citizen**. Special acceptance rules **Jewish immigrants**, for victims of National Socialist persecution.	Residence title required, with special conditions for third-country citizens who are qualified professionals, high-qualified workers, researchers or self-employed with secure funding and/or economic interest or regional need. Permanent residence permits after five years and if further conditions are fulfilled, exceptions for graduates of German higher education, highly-qualified foreigners with a specific job offer, EU-Blue card holders (after 33 month's employment and further conditions), self-employed with sucessfully established business after 3 years.	http://www.bamf.de/EN/Migration/Arbeiten/arbeiten-node.html and http://www.bamf.de/EN/Migration/Arbeiten/Daueraufenthalt/daueraufenthalt-node.html
	Comments	Blue Card from the EU is applicable for third-country nationals.		http://www.bamf.de/EN/Migration/Arbeiten/arbeiten-node.html

(continued on next page)

Table D-2 (continued)

Country	Policy	Regulation		Sources
		by nationality	by skills	
Ireland	Citizenship	No.	No.	http://www.inis.gov.ie/en/INIS/Pages/citizenship
Department of Justice and Equality, Department of Jobs, Enterprise and Innovation	Work Visa	Yes: **EU/EEA and Swiss citizens**	All non-EEA nationals need permission to remain and work. Critical Skills Employment Permit (replaced the Green Card) aims at permanent residence and facilitates entry for occupations that are defined with respect to **strategically important skills** (e.g. natural and social science or engineering professionals, ICT or health professionals, health and social services managers, nurses or midwives, health associate professionals, business, research and administrative professionals, quality and regulatory professionals, sales, marketing and related associate. Further: Sport and cultural employment permit, intra-company transfer employment permit, reactivation employment permit (if status of employment permit was lost but through no fault of their own).	https://www.djei.ie/en/What-We-Do/Jobs-Workplace-and-Skills/Employment-Permits/Employment-Permit-Eligibility/Highly-Skilled-Eligible-Occupations-List/
	Comment	Dependent/Partner/Spouse Employment Permit for the family of Critical Skills Employment Permit holders available. Does not participate in the EU Blue Card System.		https://www.djei.ie/en/What-We-Do/Jobs-Workplace-and-Skills/Employment-Permits/Permit-Types/
Netherlands	Citizenship	No.	No.	https://ind.nl/en/dutch-citizenship/Pages/Naturalisation.aspx
Immigration and Naturalisation Service (Ministry of Security and Justice)	Work Visa	Yes: **EU/EEA and Swiss citizens** are free on the Dutch labor market, **until 2017 Japanese nationals** have also been free.	Yes: Residence permit needed, depending on type of work (e.g. Asian catering industry, spiritual counselor, cross border service providers, cross border workers, seasonal workers), application is made by the employer. Special conditions for **highly skilled migrants** (need an employer), start-up entrepreneurs (need a Dutch supervisor), self-employed persons (need to serve a special interest for the Dutch economy, assessed with a scoring system), residence permit for a 'orientation year highly educated persons' after graduation, work experience as trainee or apprentice, intra corporate transferees or scientific researchers.	https://ind.nl/en/work
	Comment	Blue Card from the EU is applicable for third-country nationals.		https://www.apply.eu/BlueCard/EUcountries.php
New Zealand	Citizenship	Yes: for people born in **Samoa, Cook Islands, Niue or Tokelau** special conditions apply.	No.	https://www.govt.nz/browse/nz-passports-and-citizenship/
The Department of Internal Affairs	Work Visa	No	Yes: **Essential skills** work visa for up to 5 years (no suitable New Zealanders for the job), Work Exchange visa for up to 12 month (restricted by a quota, work exchange scheme needs to be approved and accomodation provided), **Skilled workers from China**, Special work category visa (job offer required, limited to **certain professions**) for Philippines , Vietnam, Indonesia; Silver Fern Job Search Work Visa for up to 9 month (for highly skilled young people, 300 places a year only), Religious Worker Work Visa for up to 2 years.	https://www.govt.nz/browse/immigration-and-visas/applying-for-a-work-visa/
	Comment	Visa waivers for short-time stays applies for over 40 countries. Working holiday visa for over 40 countries, conditional on age.		https://www.govt.nz/browse/immigration-and-visas/applying-for-a-work-visa/

(continued on next page)

Table D-2 (continued)

Country	Policy	Regulation		Sources
		by nationality	by skills	
Norway	Citizenship	Yes: **Nordic countries** (Denmark, Finland, Iceland, Norway, Sweden)	No.	https://www.udi.no/en/want-to-apply/citizenship/
The Norwegian Directorate of Immigration (UDI)	Work Visa	Yes: **Nordic countries, EU/EEA and Swiss citizens**	All non-EEA/EU citizens need a residence permit to work in Norway and normally need to have found a concrete job offer. Different residence permits for work for **skilled workers**, ethnic cooks, seasonal worker, self-employed persons, persons employees of humanitarian, non-profit or religious organization.	https://www.udi.no/en/want-to-apply/work-immigration/
	Comment	No residence permit is required for a stay below three months for certain occupations.		https://www.udi.no/en/word-definitions/
Spain	Citizenship	Yes: Condition of residence prior to naturalisation is reduced from 10 to 2 years for nationals of **Ibero-American countries, Andorra, the Philippines, Equatorial Guinea, and Portugal, and persons of Sephardic origin.**	No.	http://www.exteriores.gob.es/Portal/en/ServiciosAlCiudadano/InformacionParaExtranjeros/Paginas/Nacionalidad.aspx
Ministry of Foreign Affairs and Cooperation	Work Visa	Yes: **EU/EEA and Swiss citizens**	Third country nationals need a residence and a work permit, residence permit can be temporary (up to 5 years) or permanent. No special treatment for skill groups, everybody is required to proof sufficient means of subsistence.	http://www.exteriores.gob.es/Portal/en/ServiciosAlCiudadano/InformacionParaExtranjeros/Paginas/Residir.aspx
	Comment	Blue Card from the EU is applicable for third-country nationals.		https://www.apply.eu/BlueCard/EUcountries.php.
Sweden	Citizenship	Yes: **Nordic countries** (Denmark, Finland, Iceland, Norway, Sweden)	No.	https://www.migrationsverket.se/English/Private-individuals/Becoming-a-Swedish-citizen.html
Swedish Migration Agency	Work Visa	Yes: Free movement for **Nordic countries, EU/EEA and Swiss citizens**. Australia, Canada, Chile, New Zealand and South Korea: 18 to 30-year-olds can live and work in Sweden for up to a year.	Work and/or residence permit required for third country nationals. Some industries are subject to stronger controls for a work permit (Cleaning, hotellery, restaurants, construction, trade, agriculture and forestry, automobile repair, service and staffing), special occupations are performer, au pair, berry picker, visiting researcher, athlete or trainer, trainees.	https://www.migrationsverket.se/English/Private-individuals/Working-in-Sweden.html
	Comment	Blue Card from the EU is applicable for third-country nationals.		https://www.apply.eu/BlueCard/EUcountries.php

(continued on next page)

Table D-2 (continued)

Country	Policy	Regulation		Sources
		by nationality	by skills	
United Kingdom	Citizenship	Yes: **British nationals,** i.e. British Overseas citizens (citizens of the United Kingdom and Colonies (CUKC) on 31 December 1982), British subjects (since 1983 very few people qualify for this, does not comprise Commonwealth countries any more) , British protected persons or British Nationals Overseas (connected with Hong Kong before 1 July 1997), can acquire citizenship by descent. Citizens from **Gibraltar and Hong Kong** have separate regulations. UK Ancestry Visa for **Commonwealth** citizens with at least one grandparent born in the UK (for up to 5 years, after that time and conditional on continuous residence) can apply for permanent settlement.	No.	`https://www.gov.uk/browse/citizenship` and `https://www.gov.uk/government/publications/application-to-register-as-a-british-citizen-form-bos`
UK Visas and Immigration, part of Home Office	Work Visa	Yes: **EU/EEA and Swiss citizens** (valid for the time of UK leaving the EU), transitional regulations for Croatia, special regulations for **Turkish businesspersons or workers**. **Commonwealth citizens**: Right to abode (live and work) in the UK if they had the right before 1983 or if parents or spouse has the right to abode. UK Ancestry Visa for Commonwealth citizens with at least one grandparent born in the UK (for up to 5 years, after that time and conditional on continuous residence, one can apply for permanent settlement.)	For people outside EEA and Switzerland, employers need a sponsor license to employ them. Work visa are based on a **points system with special tiers for different occupations**: Tier 1: Entrepreneurs, exceptional talents, graduate entrepreneurs, investors,general. Tier 2: General, intra-company, minister of religion, sportsperson, priority services. Tier 5: Temporary worker - charity worker, temporary worker - creative and sporting, government authorized exchange, international agreements, youth mobility.	`https://www.gov.uk/guidance/immigration-rules/immigration-rules-part-6a-the-points-based-system` and `https://www.gov.uk/browse/visas-immigration/work-visas`
	Comment	*British overseas territories citizen* can hold a British passport (and get consular assistance) but have no automatic right to live or work in the UK and are not seen as a UK national by the EU. (None of the 14 British Overseas Territories is in our country sample.)		`https://www.gov.uk/types-of-british-nationality/british-overseas-territories-citizen`

(continued on next page)

Table D-2 (continued)

Country	Policy	Regulation		Sources
		by nationality	by skills	
United States	Citizenship	No.	Yes: **non-US citizen members of US armed forces** and their families might be eligible for citizenship through military services. **Green card holders** or through parents.	https://www.uscis.gov/us-citizenship/citizenship-through-naturalization/path-us-citizenship
U.S. Citizenship and Immigration Services	Work Visa	No (only in combination with special jobs)	Temporary worker visa is restricted to activities for which the non-immigrant visa was issued, students can be allowed to work. Permanent Worker Visa Preference Categories specify which professions or skill groups can immigrate based on their **job skills** (highest preference is for persons of extraordinary ability in the sciences, arts, education, business, or athletics; outstanding professors or researchers; and multinational executives and managers.). Green card (permanent residency) availability through a job for: permanent employment offer by certified employer, through investments that creates new US jobs, through self petition (for aliens of extraordinary ability), **special jobs** (Afghan/Iraqi Translater, Broadcaster, International Organization Employee, Iraqi who assisted the US, NATO-6 Nonimmigrant, Panama Canal employee, physician of national interest or religious worker).	https://www.uscis.gov/green-card/green-card-through-job and https://www.uscis.gov/working-united-states/working-us
	Comment	For shorter work-related travel: Temporary visitor for business visa (B-1 visa) and Visa Waiver Program.		https://www.uscis.gov/working-united-states/temporary-visitors-business