The Effects of Migration and Ethnicity on African Economic Development

Mathias Iwanowsky*

April 18, 2018

Abstract

Migration between countries has been shown to have positive effects on economic outcomes such as trade by fostering economic and cultural integration. In Africa, where ethnic identification is reasonably strong, omitting ethnic links between countries likely introduces a considerable bias in the estimates. Following the literature, I use past migrant clusters as instruments to show that migration in 1990 led to more bilateral exports for neighboring countries in the period 1989–2014. To account for the ethnic heterogeneity of African countries, I generalize this approach and use pre-colonial ethnic linkages between of home- and foreign-countries as an instrument for migration. The results suggest a downward bias when not accounting for ethnic heterogeneity. I discuss potential concerns of pre-colonial ethnic linkages and find no evidence of omitted variable biases caused by similar languages, preferences, or conflict. Ethnic connections instead facilitate trade, especially for groups that are excluded from government coalitions. The results are consistent with a model of international trade where cross border connections decrease the fixed costs of exporting.

*I benefited greatly from discussions with Konrad Burchardi, Dorothee Bühler, Serena Cocciolo, Masayuki Kudamatsu, Andreas Madestam, Stelios Michalopolous, Nathan Nunn, Torsten Persson, Thorsten Rogall, Jakob Svensson and Anna Tompsett. Thanks to the participants of the MIT Political Economy Lunch, Harvard History Lunch, Brown Macro Lunch, Nordic International Trade Seminars, the Oxford development workshop, the UCL Enter Seminar and NEUDC.

1 Introduction

In recent years, researchers have started to scrutinize the historical determinants of economic underdevelopment in Africa. Until recently, African economies underperformed and the arbitrary partitioning of African ethnicities into states was identified to be a contributing factor (Alesina et al., 2016; Michalopoulos and Papaioannou, 2016; Clochard and Hollard, 2018). However, African states are now among the fastest growing nations, which suggests that a model of African growth based on the permanent adverse effects of divided ethnicities on institutions is incomplete.

In contrast to the literature on institutions, research on migration and trade suggests that more connections across borders might promote growth through various channels (Burchardi et al., 2016). However, as bilateral migration is potentially affected by bilateral trade and unobservable factors, causal identification is difficult in most settings. To achieve identification, the literature has resorted to using past migration flows or clusters of migration (Munshi, 2003; McKenzie and Rapoport, 2007). Especially in Africa, this masks considerable heterogeneity as ethnic identification still plays a large role. As migrants identify with their counterparts in the exporting country, this previously neglected population is likely to have a considerable impact on the effectiveness of migration.

Studying the effects of ethnic links across African countries entails many challenges. First, the data quality on African ethnicities is, at best, questionable. Different names of ethnicities, language barriers and historical ethnic conflicts constitute considerable obstacles to research. Second, conflicts or natural catastrophes might cause migration across borders leading to endogenous contemporaneous linkages. Third, if migration today is formed by the same preferences that shape cross-border trade, an omitted variable bias needs to be addressed.

To achieve identification, I employ a spatial identification strategy based on the dispersion of pre-colonial ethnicities and their division into separate countries by the formation of country borders. However, as borders are formed to reflect interests and populations are divided between countries, population shares and trade flows are endogenous to the location of the border. To estimate a causal effect on cross-border trade flows, I rely on the quasi-random formation of country borders in Africa. The use of geographical factors has recently attracted a considerable interest among economists due to their exogeneity to the individual (Nunn, 2008; Nunn and Wantchekon, 2011; Nunn and Puga, 2012; Michalopoulos and Papaioannou, 2013, 2014). I use a map of the pre-colonial distribution of ethnicities in Africa provided by Murdock (1959) to identify ethnic groups and their population shares as an instrument for migration and ethnic linkages today. Due to the randomness of country borders in Africa with respect to ethnicities, I argue that my approach identifies a causal effect of pre-colonial linkages on contemporaneous trade.

In this study, I exploit the exogenous placement of national boundaries with respect to historical ethnic homelands in Africa to evaluate whether ethnic connections across borders are associated with relatively higher rates of bilateral trade. Politically separated ethnic enclaves can be viewed as a network spanning multiple countries, which can constitute a barrier to, as well as an opportunity for, economic development. These networks may exclude non-members from economically profitable actions, but also overcome unfavorable institutions for members. In Africa in particular, members of the same ethnicity have been shown to share information or risks, bargain jointly for preferred policies, or be more productive (Bates, 2008; Fafchamps and Gubert, 2007; de la Sierra and Mutakumura, 2014; Hjort, 2014).

Using recent bilateral trade data and migration data combined with the historical distribution of ethnicities for 46 African countries, I find that migration and exports are positively correlated. I overcome the reversed causality problem by using past migration and pre-colonial ethnic connections between countries as instruments for current migration. Taking into account ethnic heterogeneity, the impact of migration is more than twice as large as previously estimated. The data provides no evidence of potential omitted variable biases from conflict, shared preferences, or linguistic similarity, and suggests a causal impact of the strength of cross-border ethnic connections on exports.

Further, I exploit the uniqueness of African countries and ethnicities to estimate where the gains from trade are located. Using nighttime light data, I show that ethnic connections lead to relative increases in wealth in more ethnically connected regions. Cross-border ethnic connections are likely to compensate for unfavorable institutions within and between African countries, potentially affecting economic development more positively than previously thought.

The magnitude of the effect is at the upper end of the estimates found in the literature for developed countries (Bandyopadhyay et al., 2008), which supports the hypothesis that ethnic linkages in developing countries are especially important due to high levels of corruption (Svensson, 2003; Dunlevy, 2006; Olken and Barron, 2009). I use data on government participation to show that ethnicities that are excluded from political participation or trust their government less, substitute governmental institutions and use their cross-border ethnic connections to facilitate exporting. Using historic variation in political centralization of ethnicities, the evidence presented here is most consistent with information sharing within networks and in line with recent findings on the positive effects of ethnic diversity on cellphone coverage (Clochard and Hollard, 2018).

The heterogeneity in and spatial distribution of ethnicities in African countries has been identified as one reason for the relative underdevelopment of African countries (Alesina et al., 2011; Michalopoulos and Papaioannou, 2014, 2016). The contemporaneous borders in Africa were drawn by European colonial powers in the late nineteenth century. As a result, ethnic borders and country borders rarely coincide, which increased ethnic heterogeneity both within and across countries such that the average African country features more than 10 ethnic groups. As most groups speak their own language, preserve their own historical culture, and potentially share a history of conflict, extractive institutions hindered economic development and harmed smaller ethnicities (Acemoglu et al., 2001; Burgess et al., 2015). I complement this view and suggest that the ethnic division into separate countries created linkages that may have alleviated the negative impact by fostering cross-country trade.

The effects of migration in the US and Canada, as well as Asia, have been extensively studied in the trade literature.[1] With the exception of Felbermayr et al. (2010) and Burchardi et al. (2016), however, most studies suffer from endogeneity concerns. If groups migrate, they use and benefit from clusters of existing migration to settle in their new country (Munshi, 2003; McKenzie and Rapoport, 2007; Battisti et al., 2016). Accordingly, if these clusters are formed by exporting firms hiring workers from the destination country, reverse causality issues arise. These endogeneity concerns are of particular importance in developing countries which account for the majority of bilateral migration. Here, cross border linkages may be more important due to considerable barriers to trade, but also harder to measure causally, such that credible empirical evidence using African countries is missing.[2] I extend the analysis of the importance of ethnic linkages for international trade (Bandyopadhyay et al., 2008; Felbermayr et al., 2010) by using exogenously placed borders and pre-independence ethnic heterogeneity to provide causal evidence on the importance of ethnic linkages across countries for bilateral trade flows in Africa. Additionally, by using both past migration stock and ethnic heterogeneity as instruments, I highlight the potential for biased estimates in highly heterogeneous population groups.

The literature on ethnic identification has highlighted several channels and effects of ethnic fractionalization in Africa. Eifert et al. (2010) show that ethnic identification plays an important role in voting, potentially contributing to post-electoral violence in some countries (Dercon and Gutierrez-Romero, 2012). The importance of ethnicity for trust among local market ven-

[1] Gould (1994); Dunlevy and Hutchinson (1999); Herander and Saavedra (2005); Dunlevy (2006); White (2007); Partridge and Furtan (2008); Burchardi et al. (2016) study the US and Canada and Rauch (1999); Rauch and Trindade (2002); Felbermayr et al. (2010); Felbermayr and Toubal (2012) study Asia.

[2] Other papers (Peri and Requena-Silvente, 2010; Felbermayr et al., 2010, e.g.) include some African countries but focus on the links to developed countries.

dors, price dispersion across borders, and public goods provision has been well documented (Fafchamps, 2003; Aker et al., 2014; Burgess et al., 2015). By focusing on price dispersion, these studies offer credible evidence at the micro level but at the cost of generality for the entire African continent. I contribute to the work on ethnic identification and investigate whether it shapes cross-border trade flows between 46 African countries.[3] However, my estimates reflect the effect of migration on formal trade which is likely to be smaller than the effect on informal trade across border regions between African countries.

In summary, this paper highlights a potential positive effect of ethnic division in Africa and potential channels through which these effects materialize. The paper is structured as follows. In Section 2, I discuss my empirical strategy and the data. I present the baseline estimates in Section 3 and show robustness in Section 4. I identify potential mechanisms in Section 5 and conclude the paper in Section 6.

2 Empirical strategy

In work on bilateral trade, the value of bilateral exports is modeled in gravity type equations. Here, the value of trade is correlated with the size of the exporter and importer economy as larger economies attract more trade flows. In this framework, adding a stock or flow of migrants estimates the impact of migration on bilateral trade. However, estimating the impact of migration on bilateral trade between two developed and two developing countries is distinctly different. While migrants from developed countries often identify themselves by their nationality, ethnic identification dominates nationality in many African countries. Second, emigration from developing countries is correlated with natural, political or economic factors, leading to severe endogeneity concerns. These distinct features of African countries require a generalization of the standard empirical approach as well as exogenous variation to identify a causal effect.

[3]One example of such cross-border solidarity was the temporary practice of Air Namibia, the major carrier of Namibia, having a stopover in Luanda (Angola) only to refuel due to disputes with the fuel supplier at their main airport. The airline is run by an ethnicity that has strong ties between the two countries and, hence, used its credibility in Angola to buy fuel (http://www.economist.com.na/headlines/2795-air-namibia-increases-frankfurt-flights and http://hannamibia.com/uploads/pdf/news/130305093441120.pdf). Additionally, the main supplier of jet fuel in Namibia, Engen, is South African and the ethnicity is only dispersed in Angola and Namibia.

2.1 Empirical framework: trade and migration

Estimating the impact of migration between developed countries, the literature uses gravity type equations derived from most theories of international trade. These gravity equations include a population stock or flow of migrants and take the form (Anderson, 1979):

$$\log(X_{ij,t}) = \beta \log(PS_{j,t}) + B_{ij,t} + \delta_i + \delta_j + \varepsilon_{ij,t} \tag{1}$$

Here the log of exports from the exporting country i to the importing country j $\log(X_{ij,t})$ is correlated with the population share of people from i in j ($PS_{j,t}$). Controlling for exporter (δ_i) and importer (δ_j) fixed effects and bilateral characteristics ($B_{ij,t}$), β identifies the effect of the population share on the log of exports. A larger $\beta > 0$ indicates a stronger response of trade flows to changes in the likelihood of an exporter from country i finding someone with her own nationality in country j.

Implicitly, equation (1) assumes that migrants to the importing country j have a population share of one in their exporting country i.[4] While approximately true in developed countries, the population structures in developing countries are more diverse. African countries combine a multitude of ethnicities, each with their own identity and separated into multiple countries. Thus, allowing for multiple ethnicities (e) from the set of ethnicities (E) in each country $e \in E_i \subseteq E$, the general form of equation (1) is given by

$$\log(X_{ij,t}) = \beta \log\left(\sum_{e\in E_i \cap E_j}^{E} PS_{i,t,e} \times PS_{j,t,e}\right) + B_{ij,t} + \delta_i + \delta_j + \varepsilon_{ij,t} \tag{2}$$

where $PS_{i,t,e} \in [0, 1]$ is the population of an ethnicity e that is prevalent in each country pair ij, relative to the population of country i at time t. This equation correlates bilateral exports to the probability of a co-ethnic relationship (match) when randomly drawing two individuals from each country. It captures the idea that it is easier to trade with someone from your own ethnicity, but does not exclude the possibility of trading with other ethnicities if the country is prosperous.

The specific formulation of equation (2) is supported by two factors. First, it is the empirical equivalent of an otherwise standard model of international trade (Melitz, 2003; Chaney, 2008), and when amending the cost function of the exporting firm by an ethnicity specific fixed cost that

[4]The underlying equation is of the form $X_{ij,t} = (PS_{i,t} \times PS_{j,t})^{\beta} L_{i,t} L_{j,t}$ where $PS_{i,t}$ reflects the share of people in the exporting country, which is unity in the case of equation (1), and raw population ($L_{i,t} L_{j,t}$).

captures lower entry costs into an export market for ethnically connected firms (Appendix A).[5] Second, the interpretation is equivalent to the search and matching literature if an exporter from country i can export more cheaply if she finds an importer in country j that is of the same ethnicity.[6] Aggregating each firm's exports then yields the gravity type equation (2). In the search and matching literature, a match is defined when two individuals with the same characteristics are drawn. Since these characteristics are stochastic, the likelihood of a match is given in probabilities. Here, characteristics are distributed along ethnic lines, and thus the fraction of the population representing an ethnicity in the importing country is equivalent to the likelihood that an exporting firm from the exporting country finds a match in the importing country. Then, the estimated β can be interpreted as an elasticity that captures the change in match probability of each ethnicity when the population changes on either side of the border.[7]

This interpretation is similar to the standard in equation (1), as both can be interpreted as a probability of drawing two connected people in each country. In equation (2), however, I incorporate the heterogeneous population structures in African countries and allow for a large amount of subgroups within two countries that are connected. Thus, using the standard empirical approach would identify a 'nationality' effect and overstate the true 'ethnicity' specific effect, as it does not account for the variability in the exporting country.

2.2 Empirical specification and data

The empirical equivalent I estimate throughout the paper is given by:

$$\log(X_{ij,t}) = \beta \ \log \text{Ethnic Match Probability}_{ij} + B_{ij,t} + \delta_i + \delta_j + \varepsilon_{ij,t} \tag{3}$$

[5] These costs can be lower information costs, more reliable information about market structures or bribes, and fewer cases of fraud between business partners. In the Appendix, I show that equation (2) follows if firms face a fixed cost of exporting

$$PS_{i,e}^{-\eta} f_{ij}$$

with $\eta \in [0, 1)$ providing concavity for the impact of fixed costs f_{ij} on the exporting firms' profits. These fixed costs represent costs of setting up a distribution network, informing about markets, administration and paying for permits. A similar model has been suggested by Krautheim (2012), and the model nests the established standard model of Chaney (2008) with $\eta = 0$.

[6] With bilateral trade data at the ethnicity level, this equation would be $X_{ij,e,t} = (PS_{i,t,e} \times PS_{j,t,e})^{\gamma}$ with γ being the elasticity. Aggregating to the exporter-importer pair yields $X_{ij,t} = \sum_{e \in E_i \cap E_j}^{E} (PS_{i,t,e} \times PS_{j,t,e})^{\gamma}$. As long as $\gamma \in [0, 1)$, the estimated coefficient β in equation (2) underestimates γ due to the concavity introduced by γ.

[7] Note that match probability is defined as the likelihood of randomly drawing two individuals from the same ethnicity. The probability that two randomly drawn individuals are not from the same ethnicity is non-zero, but is captured by the exporter and importer fixed effects in the trade equation (2).
I explore the possibility of inter-ethnic trade in Appendix A.1. By assuming an increasing cost of trade for ethnicities that are far away from each other, I confirm the baseline estimates for the entire sample of African countries.

Here the 'Ethnic Match Probability' is defined by the sum of all ethnic match probabilities for all ethnic groups that are prevalent in both countries $PS_{i,e} \times PS_{j,e}\ \forall e \in E_i \cap E_j$ and constitutes the measure of ethnic similarity between a country pair. Every regression follows the standard in the trade literature and includes exporter (δ_i) and importer (δ_j) fixed effects and, where applicable, includes exporter-importer pair characteristics ($B_{ij,t}$).[8] A positive point estimate, $\beta > 0$, suggests that a larger population on either side of the border for a connected ethnicity yields larger trade flows.

To identify the current population of each ethnicity in each county, I use the Ethnologue data set with estimates on ethnic populations around the world based on a variety of sources.[9] I obtain exogenous variation in ethnic shares using data containing the distribution of ethnic groups before colonialization. The geographic data provided by Murdock (1959) has been used to study the relationship between slavery and trust (Nunn and Wantchekon, 2011). Matching the spatial extent of every ethnicity with grid cell population data from the United Nations Environment Program in 1960, it approximates the population of every ethnicity in every country in 1960, a time when African countries gradually gained independence.[10] To my knowledge, this is the first paper that combines the Murdock data with population data in the context of international migration and trade.

Due to time invariant population figures, the variation at the country-pair level leads to a Moulton (1986) type problem of inconsistent standard errors. I collapse the data to the exporter-importer observation and report standard errors at the country-pair level.[11] In every country pair, every country is observed once as an exporter and once as an importer, to match the data to observed migration flows. For the dependent variable, the log of bilateral exports, I use UN comtrade data from the World Bank Integrated Trade Systems from 1989–2014.[12] Since the trade data does not capture unreported and informal trade, the literature has focused on price level differences (Aker et al., 2014). I use reported trade only to attempt to estimate the

[8] Exporter-importer pair characteristics include log(Length of border), log(Distance between country centroids), dummies for speaking the same language, number of ethnic connections between the country, sharing a colonial history and a dummy that indicates whether parts of the border are determined by a river or mountains.

[9] www.ethnologue.com Sources in the data vary in timing and quality.

[10] France retreated from most of its possessions in 1958–1962, Britain in 1957–1965 and Belgium in 1960–1962. The conclusions in this paper are qualitatively robust to very coarse information on population in 1900 contained in Murdock (1959), but due to its incompleteness and the noise I do not report it here.

[11] As this severely reduces the degrees of freedom and to weight observations by their informativeness, I show robustness to weighting every observation with the number of times I observe trade between that pair.

[12] In order to have a better match, I download import and export data and cross match imports and exports to generate reliable export measures. The results are robust with either inputs, but for sample-size reasons, I end up using the matched data.

effects for all countries, taking into account that the point estimates are likely lower bounds on the true extent of exports between countries.[13]

The final sample consists of 46 African countries in 91 country pairs with 182 exporter-importer relationships that share a border, observed over 26 years. Due to non-reported trade, the sample is reduced to 3,287 observations. Since the variation I intend to exploit is at the country-pair level, I follow the conservative choice and cluster the standard errors at this level.[14]

2.3 Identification strategy

The empirical approach in the trade literature uses flows or stocks of migrants and correlates these with bilateral exports. However, economic activity attracts trade and migration flows in a similar fashion, leading to problems of reversed causality. Additionally, borders are not set randomly and reflect spheres of influence and historical economic activity, such that the direction of a potential omitted variable bias is unclear. I use the historical dispersion of ethnic groups to address the issue of reversed causation and argue that, contrary to borders between European countries, borders between African countries are exogenously placed. Combined, the historical distribution of ethnicities and exogenous placement of borders allows me to identify a causal effect of cross-border ethnicities on bilateral trade between African countries.

In African countries, ethnic population shares are affected by a multitude of factors. Natural catastrophes, hunger, civil conflicts or past migration contribute to the dispersion of people around the continent. Even without accounting for ethnic heterogeneity, these factors are correlated with economic activity and threaten a causal identification of the ethnicity effect in equation (3). In addition, if people migrate following a trade route because it constitutes their best information about potential destinations, any factor that increases trade also increases migration, leading to a problem of reversed causality.

The standard approach in the literature uses past migration to instrument for current migration as it has been shown that migrants follow their networks and settle in clusters in the importing country (Munshi, 2003; McKenzie and Rapoport, 2007). This strategy solves the reversed causality problem if initial migrants were randomly placed in countries. For this ap-

[13]If the data is split up into reported or unreported trade, the true estimate will be $\beta = \left(\beta^{reported} X_{ij}^{reported} + \beta^{unreported} X_{ij}^{unreported}\right) / (X_{ij}^{reported} + X_{ij}^{unreported})$. As long as $\beta^{reported} \leq \beta^{unreported}$, I estimate a lower bound effect. Since unreported trade is much more dependent on trust, I argue that this condition is fulfilled.

[14]The final sample leaves out island territories such as Madagascar or São Tomé and Príncipe, as well as Sudan and South Sudan. In the robustness section, I show that coding the missing observations as zero and applying the standard estimating technique by Santos-Silva and Tenreyro (2006) does not change the results. Dyadic data has specific issues with standard errors, as errors can be correlated across country pairs over time. I explore two-way clustering for 46 exporting and 46 importing countries separately in the robustness section.

proach, I use data on bilateral migration at the country level dating back to 1960 to have some exogeneous variation before the time period of interest 1989–2014.[15] To specifically allow for ethnic heterogeneity and counteract any potentially remaining issues of reverse causality and omitted variable biases, I use the pre-colonial distribution of ethnic tribes in Africa (Murdock, 1959). Here, I combine the geographic location of each ethnicity with detailed grid cell population data from the United Nations Environment Program in 1960 to obtain population estimates of migrants and their home population at the time of independence.

Since the Murdock map shows the pre-colonial distribution of 833 ethnicities in Africa, strategic selective sorting into future countries is relatively unlikely. However, the population figures in Murdock (1959) are estimates combined from different sources and given by ethnicity, as opposed to by country, leading to potentially severe measurement error. Hence, I use detailed grid cell population data at a 4.5 km resolution in 1960 which yields a reliable population estimate for the ethnic homelands just prior to independence.

Having a reliable estimate of pre-independence population does not solve the issue of endogeneity. In European countries, borders reflect spheres of interest and were likely set to encompass a homogeneous population. Thus, the population shares of each group in each country are determined by the border, and governments might have had economic reasons to place a border. If a border between two governments was set to include a territory, this might reflect military considerations, but also the interest of having access to potential future markets. Then, population shares and bilateral trade flows are endogenous, and causal inference is problematic.

In Africa, however, the contemporaneous borders between countries were drawn in 1884 at the Berlin conference. These borders do not reflect the interest of the ethnic groups, but the interest of the colonial powers.[16] The exogeneity of these borders has been extensively used in the recent literature on culture and development, price dispersion across borders as well as ethnic fractionalization (Alesina et al., 2011; Aker et al., 2014; Michalopoulos and Papaioannou, 2014). Most country borders today feature parts that follow either latitudinal or longitudinal lines since the exact geography of Africa was largely unknown at the Berlin conference.[17] Where the geography was known and country borders could have been set to follow rivers or mountain ridges, the evidence in Figure 2 still suggests no such pattern. Here, country borders, shown in black, rarely overlap with major rivers shown in blue.

[15] http://databank.worldbank.org/data/reports.aspx?source=global-bilateral-migration

[16] For example, Aker et al. (2014) argue that the border between Nigeria and Niger was set at the Berlin Conference in 1884-1885. It was not a border reflecting geographic features but rather the political interests of France and Britain. The border eventually emerged in 1906 and the resulting mixture of ethnicities shows a similar pattern in 2008.

[17] Alesina et al. (2011) show that 80% of African political borders follow either latitudinal or longitudinal lines.

I argue that these borders were arbitrarily drawn, split many ethnic groups into two countries, and do not reflect the interests of a specific ethnicity. In my data, all country borders between African countries divide at least one ethnicity. These split ethnic groups are likely to be different from other ethnicities in terms of size or historical economic activity. In line with Michalopoulos and Papaioannou (2013, 2016), I show in Table 1 that an ethnicity is more likely to be split if it is larger in terms of population or territory. However, population density as a measure of economic activity is negatively correlated with split ethnicities.

Supported by the evidence on historical behavior of ethnicities (columns 4-8), these correlations suggest that more widespread, more nomadic and less economically active ethnicities were split. Using data on historical characteristics of tribes, I show that split ethnicities were more likely to be nomadic (column 4), but neither the size of local communities nor historical institutions predict a future divide into more countries. Estimating all characteristics jointly to account for correlations between variables, the area an ethnicity covers in the Murdock data is the only determinant that robustly predicts the divide between countries (column 8).

However, to address concerns that these correlations influence the results, I only consider country borders where ethnicities have been split, and only consider ethnicities that are split at this border in the heterogeneity analysis. Thus, I abstract from a comparison of influential ethnicities with negligible ethnicities and use a balanced sample across similar ethnic groups. Additionally, this procedure abstracts from selection effects into having a shared ethnicity, and focuses on the intensive margin only.

I use historical information on ethnic dispersion to address the issue of omitted variable bias by conflicts, political effects, natural disasters or migration. The historical distribution of ethnicities mitigates the threat of reverse causality if migrants follow trade routes. Furthermore, the use of exogenous borders alleviates the threats posed by the endogenous formation of borders.

3 The impact of migration on exports

The positive effects of migration on bilateral trade between African countries have been highlighted in the literature to a great extent. In this section, I provide evidence for a positive effect of migration between developing countries on bilateral trade using two approaches. First, I document the effect of migration on trade using the standard approach with past bilateral migration. Second, introducing baseline ethnic heterogeneity in exporting and importing countries highlights a substantial downward bias as subsequent migration flows are likely to be correlated with initial conditions that affect trade and migration.

Effects on Exports The main results are presented in Table 2 where I report the endogeneous ordinary least squares, the first stage and reduced form using the instrument, and the point estimate from instrumenting. I estimate the impact of the stock of migrants in 1990 on the value of bilateral exports in the period 1989–2014 for the full sample of countries with trade flows. The ordinary least squares point estimate suggests that a 1% higher migrant stock increases bilateral exports by 0.139% (s.e. 0.024). For the standard approach to obtain a causal estimate, I use the migrant stock in 1980 or 1960 which are valid instruments according to the F-Statistics and similar reduced form point estimates. Instrumenting the migrant stock in 1990 with its past values in the last row of Table 2 suggests a small downward bias of the OLS as the IV point estimates range from 0.166 (s.e. 0.027) to 0.199 (s.e. 0.031).

The exclusion restriction in this specification featuring exporting- and importing-country fixed effects and country-pair controls requires that no unobserved country-pair characteristic affects both migration and exports. In standard trade theory, however, large initial income differences reflect productivity differences which increase exports from rich to poor and migration from poor to rich countries as the marginal product of labor is equalized in both countries. In this setting, more migration is negatively correlated with exports and implies a downward bias on the OLS and the IV, when past migrant stocks are used as an instrument.

In column (3), I repeat the exercise for the sample of bordering countries and show that while the point estimates are insignificant due to the lack of power, qualitative conclusions carry over to this narrower defined sample. I introduce ethnic heterogeneity in column (4) and use the constructed initial ethnic match probability as an instrument for migration flows between two bordering countries. Both the reduced form (0.192, s.e. 0.080) and the first stage (0.260, s.e. 0.093) suggest a valid instrument and the resulting IV point estimate suggests that a 1% increase in the stock of migrants increases trade flows by about 0.739% (s.e. 0.417).

The point estimates from the standard approach (column 3) are about 50% smaller than the estimates using the approach that incorporates ethnic heterogeneity to the African context (column 4). The results suggest that migration likely increases trade and ethnic identification plays an important role, even when observing cross-country trade. Similar to well-identified studies on price differences in a narrow setting, ethnic identification also seems to influence the value of exports (Aker et al., 2014).

I conclude this set of results by introducing ethnic heterogeneity in the endogenous variable. By constructing the ethnic match probability using contemporaneous data on ethnic populations in countries, the migration variable now contains information about ethnic heterogeneity (column 5). Here, the OLS suggests that a 1% higher ethnic population in the exporting or importing country increases bilateral exports by 0.232% (s.e. 0.123). Instrumenting the contemporaneous match probabilities with the pre-colonial match probabilities corroborates the results in column (4) with a similar point estimate (0.665, s.e. 0.279).

While the F-Statistic is reasonably low (7.813 and 12.200), the point estimate is unlikely to be affected by a weak instrument problem as this would bias the estimate towards the OLS. The point estimate in columns (4) and (5) is unaffected by the above violation of the exclusion restriction as the constructed ethnic match probability is determined using the pre-colonial distribution of ethnicities and country borders that quasi-randomly displaced parts of an ethnic group in another country. However, as the degree of ethnic connectivity between two neighboring countries may impact bilateral conflicts and politics, other violations of the exclusion restrictions are possible. I investigate this possibility in the mechanisms section, but focus on reduced form estimates in the remainder of the paper.

The results from Table 2 highlight the potential bias when ethnic heterogeneity is not taken into account. Columns (4) and (5) indicate that the elasticity to migration is about 0.7, which is about 2.5 times larger as compared to the results from the conventional approach in columns (1)–(3).

Effects on sub-national development The uniqueness of the African context allows me to infer on the spatial distribution of the gains from trade. As bilateral trade flows, especially in Africa, are usually between capitals or major cities, I aim at identifying whether ethnic homelands benefit from increased exports. However, as bilateral trade is part of the national GDP and hence hard to disentangle, I focus on sub-national gains from trade using nighttime light emissions.

Light emitted in a region has been found to be a valid proxy for regional GDP (Michalopoulos and Papaioannou, 2013; Henderson et al., 2012) and has frequently been used to study sub-national development in Africa.[18] I replicate these findings at the country level in Table 3. Average nighttime light emitted per country is a strong predictor of per capita GDP, controlling for population, country characteristics, and conflicts. Including country fixed effects (column 4), a 1% higher nighttime luminosity is correlated with a 0.472% (s.e. 0.054) higher per capita GDP. Additionally, including year fixed effects confines the variation to the country-year level and decreases the point estimate and increases the noise (column 5).

[18]In this study I use nighttime light data from NOAA available under `http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`. The recorded measure ranges from 0–63 and is available from 1992–2010.

The results established in the literature and Table 3 motivate to estimate the effects of exports on nighttime luminosity in Table 4. A 1% increase in exports is correlated with a 0.411% (s.e. 0.152) increase in nighttime luminosity of the ethnically connected regions of the exporter. To compare equally densely populated areas, I additionally control for population density and show the first-stage regressions of the ethnic match probabilities today and in the past in columns (2) and (3) for this sample. Both point estimates are less precisely estimated but suggest the same relationship as in Table 2. A larger ethnic match probability is positively correlated with larger exports and more nighttime lights in the ethnically connected region in the exporting country (columns 4 and 5). Instrumenting exports with the ethnic match probability (column 6) suggests that a one percent increase in exports increases sub-national GDP in the ethnic homeland by 1.6% (s.e. 0.635) over a mean of 0.213.[19] Following Henderson et al. (2012) and identifying the time dimension (column 7) yields the same conclusion with a lower point estimate of 0.618 (s.e. 0.200). Compared to the impact of total country exports on nighttime luminosity in the entire country (0.013, s.e. 0.006), the gains from exports induced by a larger ethnically connected population are overwhelmingly centered inside the ethnic homeland.[20]

The results from Table 2 and 4 indicate the importance of including population heterogeneity when studying the economic effects of migration. The impacts of migration on exports are estimated to be about 2.5 times larger than with the conventional approach. Exports are correlated with increased nighttime luminosity, and the spatial distribution of the gains from trade are likely concentrated within the ethnic homelands.

4 Instrument discussion

For the instrument to recuperate an unbiased estimate of the effect of migration, the exclusion restriction of the instrument needs to be fulfilled. However, as the exclusion restriction cannot be tested, I focus on the reduced form estimates for the remainder of the paper. In this section, I aim at showing that these reduced form estimates are not affected by specification choice, population figures, or the Murdock data itself. Having shown the stability of the instrument, I continue in the mechanisms section to test alternative hypotheses for the question of how ethnic connectivity affects exports.

[19] With a mean of 0.213 and a range of the nighttime light data of 0–63, there likely exists room for improvements in the border regions. The point estimate without controlling for population density and country-pair controls is 1.890 (s.e. 1.110) and 1.464 (s.e. 0.622), including country-pair controls without population density.

[20] The point estimate is from the regression: $\log \text{nighttime lights}_{i,t} = \beta \log \text{Total exports}_{i,t} + \delta_i + \delta_t + \varepsilon_i$ and is robust to including per capita GDP.

In Table 5 I present the reduced form estimates in various specifications. The point estimate from the baseline specification (column 1) is robust to including country-pair controls (columns 3 and 4) and adding time varying conflict controls (column 5). To increase the precision of the estimate and put less weight on exporter-importer-pairs with less informational content, I weight the baseline specification with the number of times trade is observed between the exporting and importing country in columns (2), (4), and (5). As expected, the point estimate is unaffected and noise is reduced, resulting in smaller standard errors.

Since standard errors are likely to be correlated within the country pair, I cluster the standard errors at the country-pair level in the baseline specification. However, as shocks might instead be correlated across countries, I report standard errors clustered at the exporting and importing country separately in brackets and conclude that the original clustering is more conservative.

In the remainder of Table 5, I estimate the baseline model using the method suggested by Santos-Silva and Tenreyro (2006). As trade data is likely to be Poisson distributed, I re-estimate the baseline specification using the Pseudo-Poisson-Maximum-Likelihood estimator and show that the point estimates are not statistically different to the baseline.

To identify which country-pair characteristics are predicted by the ethnic match probability, I run the baseline specification using the characteristic as the dependent variable and report the point estimates in Table 6. Country pairs with a larger ethnic match probability share longer borders and the centroids of these countries are closer together. Moreover, these countries have boundaries that are less random as measured by the border fractionalization index of Alesina et al. (2011).[21] However, other relevant characteristics such as the number of ethnic connections, the judicial language, or colonial histories seems to confound the estimate. Estimating all country-pair controls jointly, the F-Statistic on joint significance is 1.572 and only the length of the shared border remains significant at the 10% level.

In the remainder of Table 6, I estimate whether the instrument predicts conflict incidence or severity. Contrary to the literature on ethnic fractionalization within a country, co-ethnic membership across countries does not increase conflict incidences or their severity.

In Table 7, I control for various measures of the not ethnically connected population to rule out the possibility that alternative measures of population affect the precision of the instrument. Including the log population and the non-ethnic match probability in column (4), the point es-

[21] The fractionalization index is derived from a regression varying the size of boxes needed to cover the border: $\log(\text{square count}) = \alpha - \beta \log(squaresize)$. Intuitively, the number of squares needed to cover a straight line can be approximated by square size^{-1}. Any deviation will lead to a number $-1 - \delta$ and a higher fractionalization index.

timate is virtually identical to the baseline reported in column (1). However, as the population data or the underlying Murdock-maps are potentially incorrectly drawn close to country borders, there may exist ethnicities that are incorrectly coded as split between countries. I drop ethnicities with a population share of less than 1% in columns (5)–(7) and show that while the point estimates increase, they are not statistically different from the baseline.

As a last robustness test, I draw upon the Geo-Referencing of Ethnic Groups (GREG) from the *Atlas Narodov Mira* created by Russian scientists and digitized by Weidmann et al. (2010). It shows the geographic dispersion of ethnicities around the world in 1960 and has been used to study ethnic inequality before. While the two ethnic match probabilities are strongly correlated, the Murdock map captures more ethnicities at an earlier point in time and, hence, more likely to fulfill the orthogonality assumption. However, as the GREG map is closer to independence, it is likely that it captures the ethnic composition more accurately. I repeat the specification tests as well as the IV estimations in Table 8. As the input is closer in time to the independence of countries and features fewer ethnic groups, the point estimates are larger and as robust as the original Murdock input. Furthermore, the IV estimations on exports (column 6) and night time luminosity (column 7) show the same magnitude as the baseline, supporting my initial conclusions.

5 Mechanisms

A causal link between migration and exports or economic outcomes is questionable if the initial ethnic match probability has other impacts on conflict, GDP, or government coalitions which, in turn, affect exports or nighttime luminosity. However, two identification decisions support such an interpretation. First, I restrict my analysis to exclusively split ethnicities and, hence, violations of the exclusion restriction must come from an intensive rather than an extensive margin. For example, it has been shown that split ethnicities are more likely to face conflicts and are worse of economically. Here, a violation of the identification assumption would require that larger split ethnicities are differently affected than smaller split ethnicities, and that this difference is correlated with exports. Second, as I include exporter and importer fixed effects in a cross-country regression, the identification relies solely on country-pair variation. For example, the history of conflicts within each country is captured by these fixed effects, leaving country-pair induced conflict variation as the only confounder.

In this section, I provide evidence on how co-ethnic matches can affect bilateral trade by testing four hypotheses that could explain the results. The first hypothesis considers the equator belt where many ethnicities were economically active before independence. In that case, the instrument would only confirm pre-existing trade patterns in sectors that have been trading long before independence. The second hypothesis concerns the effects of conflict in Africa. If cross-border ethnic links reduce conflict incidences, it might raise economic activity and explain the results. The third hypothesis is that ethnic match probabilities only capture similarities in languages which are larger between ethnically connected countries. Hence, it is possible that the main impact of migration on trade is via reducing language barriers. Then omitting linguistic similarity constitutes an omitted variable bias. The fourth hypothesis concerns the literature on the economic effects of institutions in Africa. It is possible that ethnicities with historically stronger institutions are more likely to participate in contemporaneous governments and instead of using their network, use government institutions to facilitate trade.

Pre-existing trade patterns I approach the first hypothesis of trade in preference goods that predates independence from three angles. First, I show that the effect of ethnic matches is stable across all sectors. Then, I document that these ethnic matches increase the amount of goods traded as well as the number of sectors in which they are traded. In the last step, I demonstrate that no country from the equator belt has a large impact on the estimation. Combined, these results indicate that co-ethnic connections do not only reflect trade in preference goods and likely increase the flow of information across countries. This hypothesis is supported by the evidence on the extensive margins of trade, as more sectors are actively trading.

The effect on increased trade might only capture pre-existing trade patterns for habitual goods or document a similarity of preferences for certain goods. Such preference goods were likely traded already before the independence of countries and, hence, the external validity of the impact of migration on exports would be limited. As it is likely that these goods are concentrated in the agricultural sector and not in other sectors, I evaluate the reduced form impact in all sectors. Since Figure 4 shows no differential impact of the instrument in the agricultural sector as compared to other sectors, a preference driven story is not supported by the data at hand. Also, if preference goods are more likely to be traded undetected, the point estimates presented here are lower bounds of the true impact of cross-border ethnicities on economic outcomes.

In contrast, the largest impacts are found in industrial sectors where information and trust are more important than preferences. If these co-ethnic connections do facilitate trade via decreasing information costs, exports should be more diverse in areas with larger ethnic match probabilities. In Table 9, I provide evidence in favor of decreasing information costs using data disaggregated into two-digit industries (SIC-2) for 1989–2014 and four-digit industries (SIC-4) for 2010–2014. Country pairs with larger ethnic match probabilities have more sectors actively trading. Compared to the relevant mean, we observe about 2-3% more sectors actively trading, suggesting that ethnic connections across countries in Africa increase trade at the extensive margin as well.

As the densely populated equator belt contains many ethnicities across many countries which were economically active before colonization, the instrument might reflect this initial economic activity. In Figure 3, I omit individual countries from the baseline specification to estimate their impact in the regression. The low point estimates for Angola and Zambia indicate that these countries have larger effects than the remaining countries. However, since the country borders for Angola are mostly straight lines and neither country belongs to the belt, Figure 3 provides no evidence in favor of a hypothesis based on pre-existing trade patterns.

Since the ethnic match probability affects all sectors and increases the number of sectors trading, a hypothesis based on pre-existing trade routes based on preferences is unlikely. This conclusion is supported by the result that no country from a densely populated area with a rich history of trade affects the point estimates significantly.

Conflict In the second hypothesis, I test whether an increased ethnic connectedness is associated with less conflict, which, in turn, increases the economic activity in Table 10. In addition to the null result on country wide conflict severity when testing for balance (Table 6), I use geocoded conflict data from the Uppsala Conflict Data Program and identify conflict intensity inside the homelands of cross-border ethnicities. Including conflict incidence or conflict severity in the exporter country, importer country, or both jointly, does not affect the point estimate throughout Table 10 and all interaction effects are insignificant, suggesting no heterogeneous effects. Additionally, I find no evidence that the ethnic match probability is predictive for any conflict measure (last row). In sum, the evidence from Table 10 suggests no direct channel of ethnicities affecting conflict and hence exports.

Linguistic similarity In the third hypothesis, I argue that two more ethnically connected countries are likely to have more similar languages and cultures. Then, an increased ethnic match probability might just reflect countries that are trading more because of similarity, and not because of migration or ethnicities. To obtain a measure beyond sharing a judicial language, I use data from Spolaore and Wacziarg (2015) that captures the linguistic similarity between a subset of country pairs in Africa as a measure between zero and one. In this subset, the point estimate is largely unchanged from the baseline with controls (Table 12, column 1), and a larger ethnic match probability is associated with a larger linguistic similarity (column 2). However, including the level effect and the interaction with the ethnic match probability does not affect the point estimate. Again, the interaction is insignificant, suggesting no heterogeneous effects of linguistic similarity.

Combined, the positive effects on the extensive margin of trade and the null-effect on conflict and linguistic similarity suggest that ethnic connections between countries are likely to work by providing information and increasing trust between business partners. However, as government institutions should alleviate problems of mistrust between business partners, the question remains whether these government institutions complement or substitute ethnic connections across country borders.

Government participation and institutions It is entirely possible that governments build on their ethnic connections to foster trade. However, it is equally likely that ethnicities trade with their counterparts in other countries, as they are actively discriminated against or excluded from power. I use data from the Ethnic Power Relations data set to identify the political status of cross-border ethnicities in Africa in Table 12 (Wimmer et al., 2009).[22] Unfortunately, the sample is severely reduced, but the main effect is robust to any definition of political status (columns 2–4). While the political status variable ranges from being discriminated to being the dominant ethnicity (column 2), I vary the definitions to increase power and show robustness (columns 3 and 4). In the raw data, the impact of being discriminated or being a senior partner is lower than the effect of being a junior party in the omitted category. This suggests that at the one end, ethnicities at least need some economic freedom, and at the other end, ethnic groups use the tools of the government when they are at their disposal. The point estimates suggest that ethnic groups with considerable political power have a 60% lower impact on exports than their counterparts without power (column 4). The decreased point estimate could imply that ethnicities with political power use it to foster trade with other countries, or trade more within their own country. As the former is picked up by the country fixed effects, and the latter is unobservable, I cannot disentangle the two. However, it is clear that ethnicities that are not an influential part of government coalitions are likely to have a large effect on trade.

Whether ethnicities are part of governmental coalitions is likely influenced by their pre-colonial institutions. If an ethnic group had the organizational structure to manage cities and a political system, it is likely to take part in politics and coalitions. I use data from Michalopoulos and Papaioannou (2013) on pre-colonial ethnic institutions and show that while the point estimate is not affected when controlling for such institutions, the interaction effects on political

[22] I exclusively focus on cross-border ethnicities, as I am interested in their political status, and not the status of an ethnicity that is not split, or irrelevant for this country pair.

centralization confirm the results from contemporaneous governmental coalitions (Table 13). Historically, more politically centralized ethnicities have a smaller effect on contemporaneous exports, likely because they are more likely to participate in governments.[23] In turn, if these ethnic groups then use government institutions to foster their economic development, this will explain the decreased impact on trade when they are part of government coalitions.

Country pairs with many cross-border ethnic groups are likely to have similar preferences and hence, their governments might work to form preferential trade agreements which are conducive to trade. Indeed, larger ethnic match probabilities are associated with more preferential trade agreements (Table 14, column 1) which, in turn, facilitate trade (column 2). However, including the endogeneous formation of preferential trade agreements (column 4), or interacting it with the instrument, does not affect the point estimate of ethnic match probabilities. The insignificant interaction suggests no heterogeneous effects of preferential trade agreements, which is consistent with the results on government coalitions, indicating that ethnic groups inside ruling coalitions are more likely to rely on institutions, rather than their connections, to facilitate trade.[24]

Consistent with a model of international trade, where ethnic migration decreases the fixed cost of exporting, a higher likelihood of ethnic matches across country pairs is associated with more trading along the intensive and extensive margin. Every sector of trade benefits, consistent with a hypothesis where ethnic networks across countries decrease information costs or increase the enforcement of cross country contracts. When these ethnicities are part of government coalitions, evidence suggests that institutions act as substitutes to such networks and become more important.

6 Conclusion

In this paper, I provide two pieces of evidence that add to our understanding of African economic development. First, I document that the standard approach to estimating the impact of migration on exports is biased when ignoring ethnic heterogeneity. Deriving a simple model of ethnic connectedness across neighboring countries, I document the positive impacts on exports and economic development using nighttime light data.

[23]Cross referencing the data from Wimmer et al. (2009) and Michalopoulos and Papaioannou (2013) only leads to 103 ethnicities in 15 countries. More historically centralized ethnicities are 50% more likely to have gained power in the years 1989–2010. The results are not shown.

[24]The same conclusion is drawn when using the number of border crossings per country pair as an indicator of the 'willingness to trade'. While increasing trade, the point estimate of the instrument remains unchanged and I find no evidence of an heterogeneous effect.

The second result then concerns the reduced form effect of how ethnic connectedness between countries increases exports. I provide evidence against hypotheses based on preferences, conflict, and linguistic similarity, and show that the most likely mechanism is government exclusion. Ethnicities which are excluded from government participation show the strongest impacts on trade. I argue that ethnicities divert their economic activity to other countries when they are being discriminated against. In line with research on price dispersion (Aker et al., 2014), I argue in favor of information being transmitted and higher within-group trust that facilitates transactions.

In light of the vast amount of research done on the negative development outcomes of ethnicities (Alesina et al., 2016; Michalopoulos and Papaioannou, 2016), this paper provides evidence for a more nuanced view on ethnicities in Africa.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–1401.

Aker, J. C., M. W. Klein, S. A. O'Connell, and M. Yang (2014). Borders, ethnicity and trade. *Journal of Development Economics 107*(1), 1–16.

Alesina, A., W. Easterly, and J. Matuszeski (2011). Artificial states. *Journal of the European Economic Association 25*, 246–277.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016). Ethnic inequality. *Journal of Political Economy*.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review 69*(1), 106–116.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science 293*(5536), 1818–1820.

Bandyopadhyay, S., C. c. Coughlin, and H. J. Wall (2008). Ethnic networks and US exports. *Review of International Economics 16*, 199–213.

Bates, R. H. (2008). *When Things Fell Apart: State Failure in Late-Century Africa.* Cambridge University Press.

Battisti, M., G. Peri, and A. Romiti (2016). Dynamic effects of co-ethnic networks on immigrants' economic success. *NBER Working Paper* (22389).

Burchardi, K., T. Chaney, and T. Hassan (2016). Migrants, ancestors, and investments. *Working Paper*.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Pardo i Miguel (2015). The value of democracy: Evidence from road building in Kenya. *American Economic Review 105*(6), 1817–1851.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–1721.

Chaney, T. (2016). Liquidity constrained exporters. *Journal of Economic Dynamics and Control 72*, 141–154.

Clochard, G. and G. Hollard (2018). Africa's growth tragedy, 20 years on. *Working Paper*.

de la Sierra, R. S. and J.-P. Z. Mutakumura (2014). On the impact of states: Contract enforceability and ethnic institutions in East Congo. *Working Paper*.

Dercon, S. and R. Gutierrez-Romero (2012). Triggers and characteristics of the 2007 Kenyan electoral violence. *World Development 40*(4), 731–744.

Dunlevy, J. A. (2006). The influence of corruption and language on the protrade effect of immigrants: Evidence from the American states. *Review of Economics and Statistics 88*(1), 182–186.

Dunlevy, J. A. and W. K. Hutchinson (1999). The impact of immigration on American import trade in the late nineteenth and ealry twentieth centuries. *Journal of Economic History 59*, 1043–1062.

Eifert, B., E. Miguel, and D. N. Posner (2010). Political competition and ethnic identification in Africa. *American Journal of Political Science 54*(2), 494–510.

Fafchamps, M. (2003). Ethnicity and networks in African trade. *Contributions to Economic Analysis and Policy 2*(1).

Fafchamps, M. and F. Gubert (2007). Risk sharing and network formation. *American Economic Review 97*(2), 75–79.

Felbermayr, G. J., B. Jung, and F. Toubal (2010). Ethnic networks, information, and international trade: Revisiting the evidence. *Annals of Economics and Statistics 98*(4), 41–70.

Felbermayr, G. J. and F. Toubal (2012). Revisiting the trade-migration nexus: Evidence from new OECD data. *World Development 40*(5), 928–937.

Gould, D. M. (1994). Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. *Review of Economics and Statistics 76*, 302–316.

Grossman, G. M., E. Helpman, and A. Szeidl (2006). Optimal integration strategies for the multinational firm. *Journal of International Economics 70*(1), 216–238.

Henderson, J., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Herander, M. G. and L. A. Saavedra (2005). Exports and the structure of immigrant-based networks: The role of geographic proximity. *Review of Economics and Statistics 87*, 323–335.

Hjort, J. (2014). Ethnic divisions and production in firms. *Quarterly Journal of Economics 129*(4), 1899–1946.

Krautheim, S. (2012). Heterogenous firms, exporter networks and the effect of distance on international trade. *Journal of International Economics 87*(1), 27–35.

McKenzie, D. and H. Rapoport (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics 84*, 1–24.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Michalopoulos, S. and E. Papaioannou (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica 81*(1), 113–152.

Michalopoulos, S. and E. Papaioannou (2014). National institutions and subnational development in Africa. *Quarterly Journal of Economics 129*(1), 151–213.

Michalopoulos, S. and E. Papaioannou (2016). The long-run effects of the scramble for Africa. *American Economic Review 106*(7), 1802–1848.

Moulton, B. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*(3), 385–397.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the U.S. labor market. *Quarterly Journal of Economics 118*(2), 549–599.

Murdock, G. (1959). *Africa: Its Peoples and Their Culture History*. New York: McGraw-Hill.

Nunn, N. (2008). The long-term effects of Africa's slave trades. *Quarterly Journals of Economics 123*(1), 139–176.

Nunn, N. and D. Puga (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics 94*(1), 20–36.

Nunn, N. and L. Wantchekon (2011). The slave trade and the origins of mistrust in Africa. *American Economic Review 101*(7), 3221–3252.

Olken, B. and P. Barron (2009). The simple economics of extortion: Evidence from trucking in Aceh. *Journal of Political Economy 117*(3), 417–452.

Partridge, J. and H. Furtan (2008). Immigration wave effects on Canada's trade flows. *Canadian Public Policy 34*(2), 193–214.

Peri, G. and F. Requena-Silvente (2010). The trade creation effect of immigrants: Evidence from the remarkable case of Spain. *Canadian Journal of Economics 43*(4), 1433–1459.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics 48*, 7–35.

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

Santos-Silva, J. and S. Tenreyro (2006). The log of gravity. *Review of Economics and Statistics 88*(4), 641–658.

Spolaore, E. and R. Wacziarg (2015). Ancestry, language and culture. *Working Paper*.

Svensson, J. (2003). Who must pay bribes and how much? Evidence from a cross section of firms. *Quarterly Journal of Economics 118*(1), 207–230.

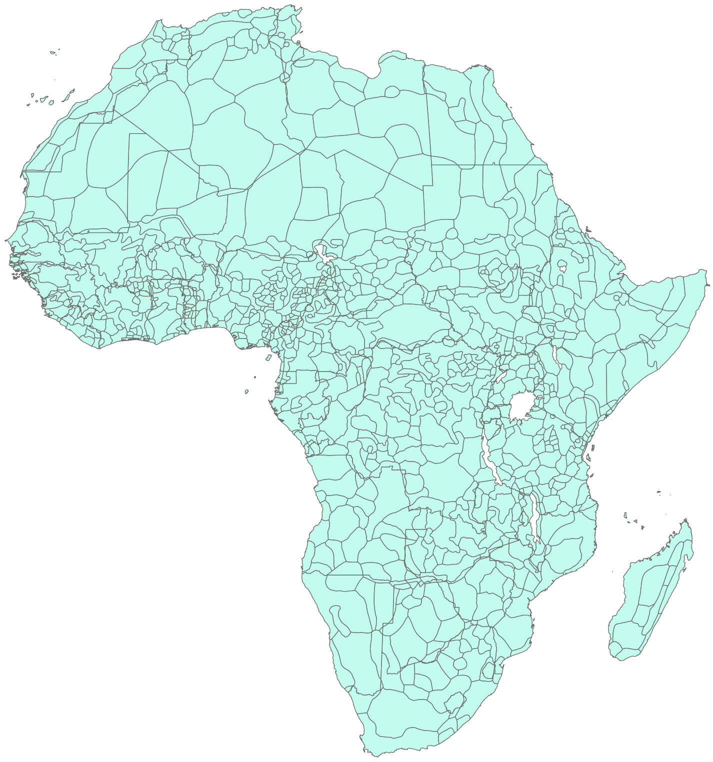
Weidmann, N. B., J. K. Rød, and L.-E. Cedermann (2010). Representing ethnic groups in space: A new dataset. *Journal of Peace Research 47*(4).

White, R. (2007). Immigrant-trade links, transplanted home bias and network effects. *Applied Economics 39*, 839–852.

Wimmer, A., L.-E. Cederman, and B. Min (2009). Ethnic politics and armed conflict. A configurational analysis of a new global dataset. *American Sociological Review 74*(2), 316–337.

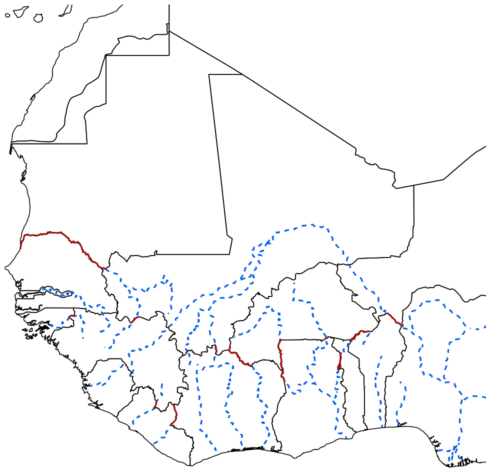
Appendix : Figures

Figure 1: Input data



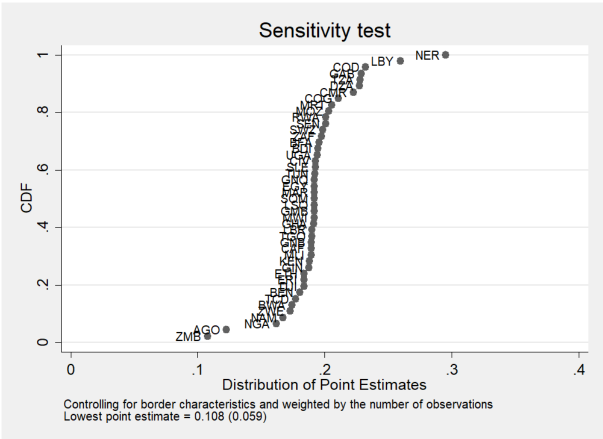
Distribution of ethnicities before colonization (Murdock, 1959).

Figure 2: Rivers as confounders



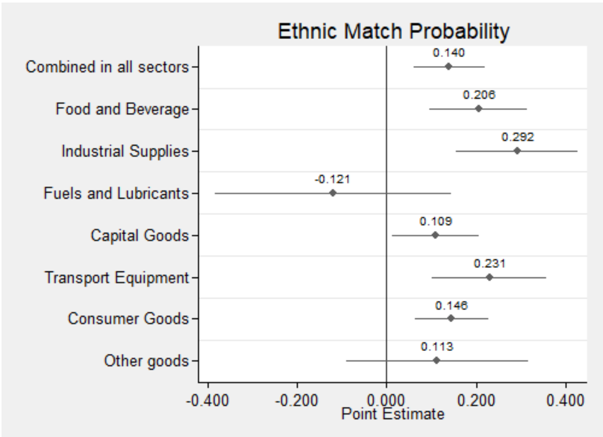
Country borders without rivers (black) and rivers that constitute country borders (red)

Figure 3: Sensitivity Analysis



Point estimates from leaving out individual countries from the baseline specification.

Figure 4: Heterogeneity across sectors



Identifying the impact of the ethnic match probabilities on bilateral trade in various sectors.

Appendix : Tables

Table 1: Determinants of being divided: Historical characteristics of Ethnic groups in Murdock (1959)

	Tribe is divided between two or more countries							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log Population in 1960	0.041***	0.008						0.015
	(0.013)	(0.015)						(0.021)
	[0.011]	[0.011]						[0.017]
log Ethnic Area		0.109						0.138***
		(0.019)						(0.022)
		[0.013]						[0.016]
log Population Density			−0.031**				−0.050***	
			(0.015)				(0.021)	
			[0.011]				[0.014]	
Cities				−0.087			−0.084	−0.046
				(0.055)			(0.059)	(0.060)
				[0.050]			[0.051]	[0.049]
Mean Size of Local Communities					0.013		0.020*	0.004
					(0.012)		(0.011)	(0.011)
					[0.011]		[0.011]	[0.011]
Political Centralization						0.036	0.038	−0.072
						(0.055)	(0.053)	(0.050)
						[0.051]	[0.051]	[0.051]
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	833	833	833	441	441	441	441	441
Adjusted R-squared	0.022	0.086	0.014	0.017	0.014	0.011	0.038	0.134

Every column shows the point estimate from a regression on the probability of an ethnicity being divided between two or more countries. Geographic Controls include latitude, longitude, and their product. log Population in 1960 taken from UNEP SIOUX grid cell data. log Ethnic Area is the total expansion area of an ethnicity as given by the Murdock map. Data in columns (4)–(8) taken from Michalopoulos and Papaioannou (2013) and coded as follows. 'Cities': If at least one ethnicity that crosses the border historically had permanent or complex settlements. 'Political Centralization' If at least one ethnicity that crosses the border historically had a jurisdictional level beyond the local level: Centralized Tribe≥2. 'Centralized Tribe' is the count variable of jurisdictional level beyond the local level (range: 0-3). Standard errors corrected for spatial correlation within 500km shown in parenthesis. Lower cutoffs decrease the standard errors to the robust standard errors level shown in brackets. Symbols reflect the significance level for spatially corrected standard errors: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: The effect of Migration on Bilateral Trade

	log Exports				
	Full sample of African countries		Sample of bordering African countries		
	(1)	(2)	(3)	(4)	(5)
Endogenous variable:	log Stock Migrants 1990				log Ethnic Match Probability, Today
Ordinary Least Squares	0.139***	0.139***	0.147	0.147	
	(0.024)	(0.024)	(0.147)	(0.147)	
					0.232*
					(0.123)
Instrument:	log Stock Migrants 1980	log Stock Migrants 1960	log Stock Migrants 1960	log Ethnic Match Probability	log Ethnic Match Probability
Reduced Form	0.153***	0.140***	0.142	0.192**	0.192**
	(0.025)	(0.023)	(0.095)	(0.080)	(0.080)
First Stage	0.922***	0.704***	0.544***	0.260***	0.289***
	(0.016)	(0.025)	(0.122)	(0.093)	(0.083)
IV Estimate	0.166***	0.199***	0.260	0.739*	0.665**
	(0.027)	(0.031)	(0.157)	(0.417)	(0.279)
Country-pair controls	Yes	Yes	Yes	Yes	Yes
Observations	1902	1902	168	168	168
F-Stat	3532.121	790.770	19.892	7.813	12.200

The samples consist of 45 countries that trade with every other country (columns (1)–(2)) or only their neighbor (columns (3)–(6)). Country-pair controls for the full sample are: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. In the border sample I additionally control for the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Nighttime lights as a predictor for per capita GDP

	log GDP, per capita				
	(1)	(2)	(3)	(4)	(5)
log Average Nighttime lights	0.380***	0.557***	0.553***	0.472***	0.162
	(0.057)	(0.051)	(0.053)	(0.054)	(0.162)
# Ethnicities		−0.003	−0.003		
		(0.008)	(0.008)		
Ruggedness		0.012	0.013		
		(0.055)	(0.055)		
log country area		0.575***	0.574***		
		(0.064)	(0.066)		
log population in 1960		−0.608***	−0.608***		
		(0.098)	(0.099)		
log # Conflicts		0.016	0.018	−0.037**	−0.019
		(0.039)	(0.040)	(0.016)	(0.015)
log # Civilian Casualties		−0.026	−0.021	0.007	0.001
		(0.027)	(0.026)	(0.012)	(0.011)
Country Fixed Effects				Yes	Yes
Year Fixed Effects			Yes		Yes
Observations	893	874	874	893	893

Average Nighttime lights calculated in the period 1992–2010. # determined by the amount of tribes from the Murdock map in a country. Ruggedness taken from Nunn and Puga (2012). Conflicts and Casualties taken from the Uppsala Conflict Database. Standard errors clustered at the country level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The effect of bilateral trade on the economic activity of bordering ethnicities

	log Nighttime lights	log Exports		log Nighttime lights			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log Exports	0.411***					1.664**	0.618***
	(0.152)					(0.635)	(0.200)
log Ethnic Match Probability, Today		0.640		0.812*			
		(0.462)		(0.457)			
log Ethnic Match Probability			0.184**		0.307***		
			(0.079)		(0.110)		
Estimation	OLS	OLS	FS	OLS	RF	IV	IV
Population density in the exporter and importer country	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter×Year fixed effects							Yes
Exporter×Year fixed effects							Yes
Importer ×Year fixed effects							Yes
Observations	168	168	168	168	168	168	2,983
F-Stat						5.428	7.348

Nighttime lights calculated from the period 1992–2010 in the ethnically connected region in the exporting country. I control for log(0.01+population density in 1960) in all specifications following Michalopoulos and Papaioannou (2013). Mean luminosity in the sample is 0.214. Column (1) shows the point estimate from a regression of exports on nighttime lights and column (2) from the regression of the ethnic match probabilities today on nighttime lights. Column (3) is the first-stage estimate using the ethnic match probabilities from the Murdock map to instrument exports via the ethnic match probability today. Columns (4) and (5) then show the reduced forms of the ethnic match probabilities on nighttime lights and column (6) the instrumental variable estimate. In Column (7) I estimate the time varying version and include country×year fixed effects to account for all country-specific variables that might change over time. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The effect of ethnic matches on bilateral exports: Sensitivity to specification

	log Exports					Exports	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log Ethnic Match Probability	0.135*	0.135***	0.192**	0.175***	0.201***	0.118*	0.127**
	(0.073)	(0.050)	(0.080)	(0.049)	(0.047)	(0.067)	(0.062)
	[0.061]	[0.046]	[0.060]	[0.042]	[0.040]		
Weighted by Number of Observations		Yes		Yes	Yes		Yes
Country-pair controls			Yes	Yes	Yes	Yes	Yes
Conflict Controls					Yes		
PPML estimation						PPML	PPML
Observations	168	3,287	168	3,287	3,287	168	3,287

Every column shows the point estimate from a regression of ethnic match probabilities on exports. In columns (1) and (2) I estimate the unrestricted, plain model with exporter- and importer-country fixed effects. In columns (3)–(7) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. In column (5) I include time varying conflict controls that include the logged amount of conflict, civilian casualties, total deaths, and unknown deaths. In columns (6) and (7) I estimate the Pseudo-Poisson-Maximum-Likelihood method as suggested in Santos-Silva and Tenreyro (2006). In columns (2), (4), (5), and (7), I weight every observation with the amount of positive trade observed in the time span 1989–2014 to put more weight on observations with more informational content. Standard errors clustered at the country-pair level shown in parenthesis and two-way clustered standard errors for the exporter and importer country shown in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Instrument Validity: Covariate Checks

	Mean	Std. Dev.	Observations	β	s.e.
Country-pair controls:					
log Border length	6.405	0.715	168	0.114***	(0.037)
log Distance Centroids	6.052	1.805	168	−0.148*	(0.088)
Border with River	0.601	0.491	168	0.022	(0.022)
# ethnic connections	3.548	1.999	168	0.114	(0.115)
Same Judicial Language	0.161	0.368	168	0.005	(0.029)
Shared Colonial History	0.411	0.493	168	−0.004	(0.018)
log Border Fractionalization	−0.004	0.026	168	0.002**	(0.001)
Border with mountain top $\geq$ 1000m	0.536	0.500	168	0.021	(0.029)
Border with mountain top $\geq$ 2000m	0.173	0.379	168	0.010	(0.017)
			F-Stat on joint significance:		1.572
Conflict controls:					
log # conflicts	6.039	1.760	168	0.043	(0.047)
lnCivillian	7.366	2.322	168	0.043	(0.065)
lnDeath	7.156	2.918	168	0.033	(0.083)
lnUnknown	6.731	2.672	168	0.030	(0.066)
		F-Stat on joint significance (incl. border country-pair controls):			1.765

I report β from the regression of the ethnic match probability on the variable in the first column. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Robustness to various population measures and cutoffs

	log Exports						
		Including non ethnic population			Excluding minorities in the country pair		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log Ethnic Match Probability	0.192^{**}	0.196^{**}	0.189^{**}	0.194^{**}	0.268^{*}	0.262^{*}	0.249^{*}
	(0.080)	(0.085)	(0.083)	(0.091)	(0.151)	(0.150)	(0.128)
log Non-Ethnic Population (Exporting Country)		0.161		0.163			
		(0.139)		(0.142)			
log Non-Ethnic Population (Importing Country)		−0.193		−0.192			
		(0.129)		(0.129)			
log Non-Ethnic Match Probability			−0.061	−0.041			
			(0.180)	(0.206)			
Country-pair controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cutoff in Exporter Country:					≥1%		≥1%
Cutoff in Importer Country:						≥1%	≥1%
Observations	168	168	168	168	168	168	168
Country pairs with ethnic connections:					152	154	125

Every column shows the point estimate from a regression of ethnic match probabilities on exports. In columns (1)–(7) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. In columns (2) and (4) I include the log population in the exporting and importing country that is not ethnically connected between the countries. In columns (3) and (4) I construct the Non-Ethnic Match Probability in the same way I construct the main explanatory variable. In columns (5)–(7) I exclude ethnicities who make up less than 1% of the population in the exporting or importing country. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Different input variable: Geo-Referencing of Ethnic Groups (1960)

	log Exports						log Nighttime lights
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log Ethnic Match Probability, 1960	0.353^{***}	0.297^{***}	0.301^{***}	0.246^{***}	0.389^{***}		
	(0.131)	(0.088)	(0.107)	(0.076)	(0.087)		
log Ethnic Match Probability, Today						0.831^{**}	
						(0.361)	
log Exports							0.346^{**}
							(0.133)
Weighted by Number of Observations		Yes		Yes	Yes		
Country-pair controls			Yes	Yes	Yes	Yes	Yes
Conflict Controls					Yes		
IV estimation						IV	IV
Controlling for population density							Yes
Observations	164	3,201	164	3,201	3,201	164	164
F-Stat						9.665	6.543

Every column shows the point estimate from a regression of ethnic match probabilities on exports. In columns (1) and (2) I estimate the unrestricted, plain model with exporter- and importer-country fixed effects. In columns (3)–(7) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. In column (5) I include time varying conflict controls that include the logged amount of conflict, civilian casualties, total deaths, and unknown deaths. In columns (2), (4), and (5), I weight every observation with the amount of positive trade observed in the time span 1989–2014 to put more weight on observations with more informational content. In column (7) I control for population density in the exporter and importer country separately as suggested by Henderson et al. (2012). Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Extensive Margin of Trade: More sectors active in Trading

	# Sectors trading			
	SIC-2, 1989-2014		SIC-4, 2000-2014	
	(1)	(2)	(3)	(4)
log Ethnic Match Probability	0.244***	0.027***	9.705**	0.147***
	(0.060)	(0.008)	(4.037)	(0.033)
# of Ethnic Connections	0.017	0.005	1.685	0.007
	(0.110)	(0.011)	(5.062)	(0.022)
Country-pair controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Poisson Estimation		Yes		Yes
Mean # Active Sectors	11.417	11.417	267.016	267.016
Observations	3,287	3,287	925	925

Every column shows the point estimate from a regression of ethnic match probabilities on exports. In columns (1)–(4) I include year fixed effects and the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. In columns (2) and (4) I account for the count structure of the data and use a poisson maximum likelihood estimation. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Heterogeneous Effects: Conflict and the effect of ethnic matches on bilateral exports

	log Exports								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log Ethnic Match Probability	0.149*	0.183*	0.158	0.204**	0.170	0.186	0.153**	0.206***	0.170**
	(0.085)	(0.092)	(0.096)	(0.102)	(0.104)	(0.115)	(0.075)	(0.077)	(0.072)
× Exporter Conflicts	0.019		0.023						
	(0.029)		(0.025)						
× Importer Conflicts		0.004	−0.008						
		(0.037)	(0.037)						
× Exporter Deaths				−0.004		−0.011			
				(0.022)		(0.019)			
× Importer Deaths					0.006	0.012			
					(0.026)	(0.027)			
× Exporter Civilian							0.018		0.020
							(0.018)		(0.018)
× Importer Civilian								−0.007	−0.011
								(0.023)	(0.019)
Level effect included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-pair controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	168	168	168	168	168	168	168	168	168
Exogeneity T-statistic:	−0.423	−0.213		−1.422	−1.213		0.330	0.436	

Every column shows the point estimate from a regression of ethnic match probabilities on exports. In columns (1)–(3) I include the level and interaction effect of conflict incidence inside the ethnic homeland in the exporting- and importing country as defined by the Murdock maps. In columns (4)–(6) I include the level and interaction effect of the amount of deaths inside the ethnic homeland in the exporting- and importing country as defined by the Murdock maps. In columns (7)–(9) I include the level and interaction effect of the amount of civilians dead inside the ethnic homeland in the exporting- and importing country as defined by the Murdock maps. In the last row I present the t-statistic from the regression of ethnic match probability on the relevant conflict variable in the exporting- and importing country. In columns (1)–(7) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Heterogeneous Effects: Linguistic Similarity and the effect of ethnic matches on bilateral exports

	log Exports	Linguistic Similarity	log Exports	
	(1)	(2)	(3)	(4)
log Ethnic Match Probability	0.226***	0.004*	0.222***	0.170**
	(0.075)	(0.002)	(0.075)	(0.067)
Linguistic Similarity			2.356	13.695
			(7.272)	(8.648)
Linguistic Similarity × log Ethnic Match Probability				1.009
				(0.609)
Country-pair controls	Yes		Yes	Yes
Observations	137	137	137	137

Every column shows the point estimate from a regression of ethnic match probabilities on exports. Linguistic Similarity is defined as 1-weighted distance of language as in Spolaore and Wacziarg (2015). Its mean is 0.155 with a standard deviation of 0.107. In columns (1), (3), and (4) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Heterogeneous Effects: Government Participation and the effect of ethnic matches on bilateral exports

	log Exports			
	(1)	(2)	(3)	(4)
log Ethnic Match Probability	0.454***	0.659***	0.676***	0.510***
	(0.112)	(0.178)	(0.177)	(0.142)
× Discriminated		−0.549***		
		(0.197)		
× Irrelevant		−0.120		
		(0.425)		
× Powerless		−0.355**		
		(0.175)		
× Regional Autonomy		−0.200		
		(0.417)		
× Senior Partner		−0.727***		
		(0.166)		
× Dominant Ethnicity		0.147		
		(0.425)		
× Discriminated			−0.430**	
			(0.206)	
× Minority or Powerless			−0.335**	
			(0.166)	
× Senior Partner, Regional Autonomy, or Dominant			−0.524***	
			(0.175)	
× Senior Partner, Regional Autonomy, or Dominant				−0.294***
				(0.108)
Country-pair controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,282	1,282	1,284	1,284

Every column shows the point estimate from a regression of ethnic match probabilities on exports. Here I match the Murdock ethnicities to the ones reported in the ethnic power relations data set (Wimmer et al., 2009) and only use county-pairs where I could match at least one ethnicity. In column (1) I estimate the baseline model in this sub-sample. Columns (3)–(4) are variations of Column (2) with broader definitions of government relations to increase power. The omitted category in columns (2) and (3) is 'Junior Partner' and in columns (4) 'Discriminated', 'Irrelevant', 'Powerless', and 'Junior Partner' In columns (1)–(4) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Heterogeneous Effects: Historical institutions and the effect of ethnic matches on bilateral exports

	log Exports				
	(1)	(2)	(3)	(4)	(5)
log Ethnic Match Probability	0.182*	0.183*	0.197**	0.249***	0.242**
	(0.093)	(0.092)	(0.088)	(0.092)	(0.099)
Ethnicities had Cities	0.262	0.243			0.297
	(0.690)	(1.166)			(1.145)
× Ethnicities had Cities		−0.003			0.026
		(0.189)			(0.194)
Political Centralization			−0.178	−1.882**	−1.884**
			(0.513)	(0.877)	(0.899)
× Political Centralization				−0.212**	−0.207**
				(0.093)	(0.099)
Country-pair controls	Yes	Yes	Yes	Yes	Yes
Observations	164	164	164	164	164

Every column shows the point estimate from a regression of ethnic match probabilities on exports. Data taken from Michalopoulos and Papaioannou (2013) and coded as follows. 'Ethnicities had Cities': If at least one ethnicity that crosses the border historically had permanent or complex settlements. 'Political Centralization' If at least one ethnicity that crosses the border historically had a jurisdictional level beyond the local level: Centralized Tribe$\geq$2. 'Centralized Tribe' is the count variable of jurisdictional level beyond the local level (range: 0-3). In columns (1)–(5) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Heterogeneous Effects: Preferential Trade Agreements

	PTA	log Exports			
	(1)	(2)	(3)	(4)	(5)
log Ethnic Match Probability	0.018*		0.222***	0.218***	0.196***
	(0.009)		(0.045)	(0.044)	(0.038)
Preferential Trade Agreement		0.403*		0.262	1.014*
		(0.235)		(0.200)	(0.538)
×Preferential Trade Agreement					0.116
					(0.073)
Country-pair controls	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	3,012	3,012	3,012	3,012	3,012

Every column shows the point estimate from a regression of ethnic match probabilities on exports using the panel dimension of the data. In columns (1)–(5) I include the following country-pair controls: Whether the country pair shares a border, the same judicial language or a common colonial tie. Distance between the country pair and the amount of ethnic links between the country pair are included. Characteristics of the border feature the length of the border as well as whether the border contains a river or a mountain top above 1000 or 2000 meters. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A Technical Appendix

In this section, I derive a model of international trade with firm and ethnic heterogeneity to provide a motivation for the main estimation equation (2). My framework draws on Chaney (2008) and nests the standard within while being tractable.

The economy consists of N countries which contain a subset $e \in E$ of predefined ethnicities. Not every ethnicity is present in every country. Furthermore, every economy produces a homogeneous composite good q_0, as well as horizontally differentiated goods $q(\omega)$. Any firm of ethnicity $e \in E$ producing a heterogeneous good $\omega \in \Omega$ from country $i \in N$, uses its ethnic counterpart $e' \in E$ in country $j \in N$ to maximize the expected profits from selling in market $j \in N$ according to:

$$\pi_{ij,ee'}(\omega) = p_{ij}(\omega)q_{ij}(\omega) - c_{ij,ee'}(\omega) \tag{4}$$

Where the price of a good $p_{ij}(\omega)$ is country specific, as is the demand for a good $q_{ij}(\omega)$.[25] $\tau_{ij} > 1$ represent variable trade costs, denoted as "iceberg trade costs". A firm needs to produce τ_{ij} goods in order to sell one unit in country j. The cost of producing a good $c_{ij,ee'}(\omega)$ is assumed to be ethnic dependent in home e and foreign e' and of the form:

$$c_{ij,ee'}(\omega) = \frac{\tau_{ij}}{\varphi} q_{ij}(\omega) + \left(\frac{L_{j,e'}}{L_j}\right)^{-\eta} f_{ij} \tag{5}$$

Here, φ denotes productivity which every firm draws from a Pareto distribution $G(\varphi) = 1 - \varphi^{-\gamma}$.[26] γ represents the degree of firm heterogeneity, with increasing values denoting decreasing firm heterogeneity. Firms learn about their productivity when drawing from $G(\varphi)$ and, subsequently, decide to pay country pair specific fixed costs f_{ij} in order to serve market j.[27] These fixed costs are mitigated by the fraction of the population in country j that is of the same ethnicity $e' = e \in E$ as the owner of the firm.[28] I call the effect of the fraction $\left(\frac{L_{j,e'}}{L_j}\right)^{-\eta}$ the network effect of ethnic ties. This fraction lies within the unit interval and raised to the power

[25] Although Aker et al. (2014) show that ethnicities affect the prices between two countries, I assume that this is a result of a supply or demand shock. However, including a demand shock here would create a simple demand shift in the gravitiy equation. Alternatively, one could divide the product space into goods consumed by ethnicities which would yield a result similar to including different sectors.

[26] Following the literature standard I use the Pareto distribution as it mirrors the empirical distributions well (Axtell, 2001) and is notational convenient.

[27] The cost of producing a good are wages times $c_{ij,ee'}(\omega)$. Due to the production in the freely traded homogeneous good q_0 wages in both sectors are normalized to unity to simplify the expressions. Furthermore, since there are infinitely many possible firms of each ethnicity, I can characterize the costs of producing variety ω simply by the ethnicity and the productivity of the firm φ.

[28] A similar approach has been undertaken by Krautheim (2012) where the fraction is the number of domestic firms active in the destination market. In the following, I assume that every ethnicity has at least one member in every country. I can relax this assumption and assume that there is an additional fixed cost to pay when dealing with non co-ethnic members. The results are robust.

of $\eta \in \left[0, \frac{\sigma-1}{\gamma}\right)$ that gives the importance of ethnic networks in decreasing the fixed costs of exporting. It can be interpreted as a decreased costs of acquiring information about the market structure in the destination country or market demand. Alternatively, its interpretation permits lower payments to government officials because of ethnic ties or it serves a proxy for the general trust-worthiness of a society. Empirical evidence by Grossman et al. (2006) suggests that factors like cultural distance and institutional development are particular relevant for the fixed cost of exporting. Ethnic networks should then be beneficial when firms try to circumvent bureaucratic hurdles. The larger the hurdles, the larger should be the impact of ethnic networks.

In every country, households maximize their utility according to:

$$U = q_0^{1-\mu} \left(\int_{\omega \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}\mu} \tag{6}$$

That is, they consume a freely traded homogeneous good q_0 and consume every available variety of the heterogeneous good ω. The share of income spent on the heterogeneous good is given by μ and the elasticity of substitution is given by $\sigma > 1$. Standard results lead to a pricing of $p_{ij}(\varphi) = \frac{\sigma}{\sigma-1} \frac{\tau_{ij}}{\varphi}$ and a demand:

$$q_{ij}(\varphi) = p_{ij}(\varphi)^{-\sigma} P_j^{\sigma-1} \mu \left(1 + \frac{\Pi}{L}\right) L_j. \tag{7}$$

Here, $\left(1 + \frac{\Pi}{L}\right) L_j$ denotes the fraction of world capital Π and labor L income that belongs to country j.[29] Hereof, a fraction μ is spend on heterogeneous goods. Combining the profit function, pricing and demand yield the ethnicity dependent productivity cutoff above which firms start to export due to non-negative profits $\pi_{ij,ee'} \geq 0$:

$$\varphi^*_{ij,ee'} = \left(\frac{\sigma}{\sigma-1}\right) \frac{\tau_{ij}}{P_j} \left[\frac{\mu}{\sigma}\left(1 + \frac{\Pi}{L}\right) L_j\right]^{\frac{1}{1-\sigma}} \left(\frac{L_{j,e'}}{L_j}\right)^{\frac{\eta}{1-\sigma}} f_{ij}^{\frac{1}{\sigma-1}} \tag{8}$$

The price index P_j can be solved explicitly by summing all prices from all exporting countries together, taking their productivity cutoffs into account.[30] Then, the productivity cutoff can be expressed in terms of primitives:

$$\varphi^*_{ij,ee'} = \left[\frac{\gamma}{\gamma - (\sigma-1)}\right]^{\frac{1}{\gamma}} \left[\frac{\mu}{\sigma}\left(1 + \frac{\Pi}{L}\right)\right]^{-\frac{1}{\gamma}} L_j^{\frac{\eta-1}{\gamma}} \frac{\tau_{ij}}{\theta_j} f_{ij}^{\frac{1}{\sigma-1}} (L_{j,e'})^{\frac{\eta}{1-\sigma}} \tag{9}$$

[29] Due to the sector that produces the homogeneous goods, wages are driven down to unity.

[30] $P_j = \left(\sum_{k=1}^N L_k \sum_{e \in E} \int_{\varphi^*_{kj,ee'}}^{\infty} \left(\frac{\sigma}{\sigma-1} \frac{\tau_{kj}}{\varphi}\right)^{1-\sigma} dG(\varphi) \right)^{\frac{1}{1-\sigma}}$.

As in Chaney (2008), the total foreign population decreases the cutoff due to market size effects $L_j^{\frac{\eta-1}{\gamma}}$. This effect is dampened by $\frac{\eta}{\gamma}$ because the ethnic population has a stronger effect on the cutoff than the total population.[31] θ denotes the multilateral resistance term that approximates how distant a market is in comparison to all other markets.[32] Equation (9) suggests that much of the ethnic network effect will work through the extensive margin of trade. If the fixed costs of exporting are higher due to corruption, the cutoff for ethnically connected and non-connected firms increases, but to a lesser extent for the former group.[33]

In order to obtain a testable equation, I aggregate individual demand[34] to an network extended gravity equation:

$$X_{ij} = \mu \left(1 + \frac{\Pi}{L}\right) L_j f_{ij}^{\frac{\sigma-1-\gamma}{\sigma-1}} \left(\frac{\tau_{ij}}{\theta_j}\right)^{-\gamma} \sum_{e \in E_i \cap E_j} L_{i,e}(L_{j,e'})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}} \tag{10}$$

Total exports between any pair of countries increase in market size $\mu\left(1+\frac{\Pi}{L}\right)L_j$ and multilateral resistance θ and decrease in variable trade cost τ_{ij} and fixed costs f_{ij}. The network term is increasing the total trade flows since $\nu \equiv \frac{\eta(\sigma-1-\gamma)}{1-\sigma} \in [0,1)$ in order to obtain interior solutions for the system of equations.[35] If the number of ethnicities is greater then the number of countries, the system of equations is under-identified and individual parameters in ν cannot be identified. A way around is to assume specific values for ν and conduct sensitivity analyses. Specifically, if ν takes on the value one, the ethnic network variable leads to a search and matching interpretation and gives the likelihood that two randomly selected firms from both countries are of the same ethnicity, when controlling for population size.

The introduction of ethnic heterogeneity in the framework of Melitz (2003) and Chaney (2008) introduced a second source of heterogeneity that creates a particular feature regarding

[31] The original cutoff in Chaney (2008) can be recovered by setting $\eta = 0$. The effect of the foreign ethnic population is greater since $\frac{\eta}{\gamma} < \frac{\eta}{\sigma-1}$ due to the assumption $\gamma > \sigma - 1$ that guarantees interior solutions.

[32] $\theta_j = \left[\sum_{k=1}^{N} f_{kj}^{\frac{\sigma-1-\gamma}{\sigma-1}} \tau_{kj}^{-\gamma} \sum_{e\in E} L_{k,e}(L_{j,e})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}}\right]^{-\frac{1}{\gamma}}$. A popular example is the comparison between Portugal and Spain with New Zealand and Australia. Similar in terms of GDP, the latter trade relatively more with each other due to their distance to all other markets in the world.

[33] Putting it differently, in a world where all the fixed cost consist of corruption and trust, the ethnic networks are paramount to exporting. We should observe only ethnically connected firms. A similar exercise can be done by changing the cost function into a part which is ethnic dependent (trust and corruption) and a part that is non ethnic dependent. Then ethnic networks do not matter when there is no ethnic dependent fixed costs, but matter a lot when there is no non ethnic dependent fixed cost.

[34] $X_{ij} = L_i \sum_{e'=e\in E} \frac{L_{i,e}}{L_i} \int_{\varphi^*_{ij,ee'}}^{\infty} dG(\varphi)$, where $\frac{L_{i,e}}{L_i}$ is the ethnic fraction in country i. An alternative summation would be to include the non ethnic population in foreign and their cutoffs: $X_{ij} = L_i \left[\sum_{e\in E_i\cap E_j} \frac{L_{i,e}}{L_i} \int_{\varphi^*_{ij,ee'}}^{\infty} dG(\varphi) + \sum_{e'\neq E_i\cap E_j} \frac{L_{i,e}}{L_i} \int_{\varphi^*_{ij,ee'}}^{\infty} dG(\varphi)\right]$. The second term would be condensed to the part in Chaney (2008).

[35] I further require that $\gamma > (\sigma - 1)$ and $\eta < \frac{(\sigma-1)}{\gamma}$ to guarantee an interior solution.

export decisions. Firms owned by an ethnic minority might first export to other markets and only later serve their home market. This feature is similar to capital-constraint firms that cannot export in Chaney (2016) and implies imperfect selection into exporting. Firms that export might have lower productivity than firms that do not and, thus, create welfare losses.

The empirical equivalent of this equation is given by:

$$\log(X_{ij,t}) = \beta \log\left(\sum_{e\in i\cap j}^{E} L_{i,e}(L_{j,e'})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}}\right) + B_{ij,t} + \delta_i + \delta_j + \varepsilon_{ij,t} \tag{11}$$

Since the importer and exporter fixed effect also capture population in each country and $(L_j \times L_i)^{-1} = -\log L_j - \log L_i$ one can rewrite the equation as:

$$\log(X_{ij,t}) = \beta \log\left(\sum_{e\in i\cap j}^{E} \frac{L_{i,e}}{L_i} \times \frac{(L_{j,e'})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}}}{L_j}\right) + B_{ij,t} + \delta_i + \delta_j + \varepsilon_{ij,t} \tag{12}$$

which as $\frac{\eta(\sigma-1-\gamma)}{1-\sigma} \to 1$ approaches equation (2). This equation can be interpreted as a search and matching model, where the population in the importing country has to incur a penalty, thus needs a larger population to have the same effect on trade as the exporting population.

A.1 Inter-ethnic Trade

So far I assumed that connections can only exist within ethnicities and neglected the possibilities of inter-ethnic connections. Here, I relax this initial assumption and assume that every ethnicity has an implicit (weak) ranking of every other ethnicity. Then, for every ethnicity I can order the other ethnicities according to the cost they have to incur in order to conduct business with them. This cost is similar to the fixed costs discussed earlier, in the sense that it reflects learning costs between ethnicities. Therefore, I assume there exists a matrix $F_{E\times E}$ that reflects this ordering between every possible combination of ethnicities. The cost of producing and exporting are then given by:

$$c_{ij,ee'}(\varphi) = \frac{\tau_{ij}}{\varphi} q_{ij}(\varphi) + \left(\frac{L_{j,e'}}{L_j}\right)^{-\eta} f_{ij} f_{ij,ee'} \tag{13}$$

with $f_{ij,ee'}$ being an element from $F_{E\times E}$. Here bilateral fixed costs are disentangled from ethnic specific cost. Every firm has to incur bilateral fixed costs to set up the firm, but also have to invest in ethnic relations in order to mitigate the additional ethnic specific fixed costs.[36] The gravity equation is then given by:

$$X_{ij} = L_j \mu \left(1 + \frac{\Pi}{L}\right) f_{ij}^{1-\frac{\gamma}{\sigma-1}} \left(\frac{\tau_{ij}}{\theta_j}\right)^{-\gamma} \sum_{e \in E \cap E'} L_{i,e} (L_{j,e'})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}} f_{ij,ee'}^{1-\frac{\gamma}{\sigma-1}} \tag{14}$$

Now, the effect of ethnic match probabilities is not only measured within ethnicities, but also between ethnicities. If the fixed costs of creating ties between ethnicities are low enough, this specification should fit the data better. Combining the findings on the extensive margin formulation and the ethnic specific fixed costs, ethnicities have a two fold effect on trade flows. They increase the number of firms exporting in distrustful environments by affecting the extensive margin. However, trade volumes between two countries are negatively affected by the ethnic specific fixed costs. Then if these fixed costs represent trust or corruption issues, the above model puts a strong emphasis on reducing corruption and increase trust among ethnicities.

Table 15: Inter ethnic networks: Using the distance between ethnicities to proxy for the cost it takes to create trust

	log Exports							
	Border Sample				Entire Africa			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log Stock Migration 1990	0.096			0.436^{*}	0.139^{***}			0.692^{***}
	(0.126)			(0.228)	(0.024)			(0.023)
log Distance weighted Match Probability		1.192^{**}	2.734^{***}			2.067^{***}	2.986^{***}	
		(0.593)	(0.617)			(0.071)	(0.079)	
Country-pair controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV estimation				IV				IV
Observations	168	168	168	168	1,902	1,902	1,902	1,902
F-Test				19.651				1421.488

Country-pair controls included: Distance between the country pair and the amount of ethnic links between the country pair are included. I additionally control for sharing a colonial history, same language, or a border. Standard errors clustered at the country-pair level shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[36] The basic model is a special case of this case where the off diagonal elements of $F_{E \times E}$ are assumed to be so high that only within ethnicity connections can occur.