CESifo WORKING PAPERS

10270
2023

February 2023

Trade and Regional Economic Development

Mathias Bühler

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10270*

Trade and Regional Economic Development

Abstract

A central argument for trade liberalization is that when the 'gains from trade' are shared, countries see large gains in economic development. In this paper, I empirically evaluate this argument and assess the impact of elite capture on regional development. Africa provides a unique study ground because the arbitrary placement of country borders during the colonial period partitioned hundreds of ethnic groups across borders. This partitioning is a source of variation in population heterogeneity and cross-country connectedness that is independent of economic considerations. Thus, African borders provide both a credible instrument for bilateral trade flows and enable the assignment of trade flows —and their impacts— to individuals. I find that while ethnic networks increase trade flows, increased trade activity decreases subnational economic development when measured by satellite data or individual wealth. I show that this counter-intuitive result comes from elite groups capturing the gains from trade, with detrimental impacts on trust and democratic progress in society.

Mathias Bühler
Ludwig-Maximilians-University
Ludwigstrasse 33
Germany – 80539 Munich
mathias.iwanowsky@econ.lmu.de

2nd February 2023
I benefited greatly from discussions with Alessandra Alloca, Konrad Burchardi, Dorothee Buhler, Davide Cantoni, Serena Cocciolo, Andrew Dickens, Masayuki Kudamatsu, Andreas Madestam, Martina Magli, 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. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged.

1 Introduction

Countries that open up for trade see large gains in economic development (Arkolakis et al., 2012). Yet, these 'gains from trade' can be unevenly distributed (Autor et al., 2013, 2014), especially when powerful interest groups exist. Such elite capture may be particularly detrimental to economic and societal development if dominant groups discriminate against others. Africa provides a unique study ground, because the arbitrary placement of country borders split some ethnic groups into multiple parts, but not others. This 'scramble for Africa' arguably contributed to the relative economic underperformance of, and ethnic favoritism in, African countries today (Alesina et al., 2016; Michalopoulos and Papaioannou, 2016; Clochard and Hollard, 2018; Dickens, 2018).

In this paper, I ask where the gains from trade accumulate and whether elite capture by dominant ethnic groups distort the welfare gains of trade policies. As trade flows (gravitate towards / are attracted by) the largest economy, I require exogenous variation in trade flows and calculate each individual's exposure to trade. I exploit the arbitrary placement of country borders creating cross-border networks as a shift in export intensity and use pre-colonial population shares of ethnic groups to assign trade exposure to individuals in a shift-share setting.

The analysis reveals three insights. First, larger ethnic networks between countries increase bilateral trade flows. Second, the resulting gains from trade are not shared equally as nighttime light data and individual survey data reveal a negative relationship between trade exposure and welfare. Third, the gains from trade accumulate in ethnic groups in political power, providing suggestive evidence for elite capture. This paper thus yields new insights on the distribution of the gains from trade under elite capture and casts a shadow on trade policies' impact on regional development.

These insights build on obtaining quasi-exogenous variation in trade flows that can be linked to individuals. To this purpose, I utilize the pre-colonial distribution of ethnic groups in all 46 continental African countries and use that not a single country border aligns with a border between ethnic groups. My comparison is thus strictly between countries that share borders and ethnic groups, and does not take into account countries that do not. Since networks are calculated only using split ethnic groups, the analysis also does not compare ethnic groups that are split to those that are not. Then, any (imaginary) shift in the border of two countries only affects the fraction of the population belonging to a specific ethnicity. Since these borders were drawn in 1884 without taking into consideration that these countries could become independent more than 60 years later, the population share of each ethnicity is essentially random, and so is the ethnic network across countries.

The empirical analysis unfolds in three parts. How ethnic networks impact trade flows, how trade flows impact individual welfare, and what are the channels. First, I estimate a gravity-type equation (Chaney, 2008), expanded to allow for heterogeneous populations within the importing and exporting country. Controlling for a large set of importer and exporter fixed effects then isolate the effect of the cross-border ethnic network. The ethnic network is constructed by calculating the population share of each ethnicity in each country using grid-cell population data prior to the country's independence. The product of population shares in each country then form the ethnic network of ethnicity e. It thus has an interpretation akin to classical search and matching models: It captures the probability of two random individuals from each country belonging to the same ethnic group.

The ethnic networks of all shared ethnic groups between the exporting and importing country are then aggregated and correlated with bilateral exports. In this dyadic structure, controlling for country fixed effects and determinants of the border isolates the impact of ethnic networks on trade flows. My results suggest that a 1% larger ethnic network between two countries increases trade flows by 0.18%. This result is robust to various specifications, definitions of the dependent variable, and a Poisson-Pseudo-Maximum-Likelihood (PPML) accounting for zeros in trade data.

Second, to capture each individual's exposure to trade flows, I utilize insights from the *shift*-*share*-instrument literature. I aggregate predicted bilateral exports to the country-of-origin level (*shift*) and interact this aggregate with the pre-colonial population share of the individual or ethnicity (*share*). Controlling for country$\times$year and country$\times$ethnicity fixed effects then isolates the variation that increased exports have on individual ethnic groups.

Both nighttime light data (2012-2020) at the ethnic-group level and georeferenced data at the individual level from the Afrobarometer (1999-2018) reveal a significant negative relationship between predicted trade exposure and regional economic development. A standard deviation increase in trade exposure decreases average luminosity by 3.2% and welfare by 11.4% of a standard deviation. Considering that trade exposure also significantly reduces trust and democratic values, policies increasing trade may have significant negative consequences for individuals.

Lastly, I ask where the gains from trade accumulate. I show how ethnic networks increase migration, which in turn increases trade flows. IV estimates suggest that a each percent increase in migration increases trade by 0.66%. As these effects are centered in the manufacturing sector, I show that increasing exports in the manufacturing sector also reduces regional development. This is consistent with the hypothesis that ethnic networks decrease information frictions across borders (Aker et al., 2014), but that industrial structures that exploit this advantage are not built in the ethnic homelands, but rather with the ethnicity in power.

To estimate the effects of elite capture, I use data on ethnic power relations in government Wimmer et al. (2009). Controlling for the average status of ethnic groups in each country, I show that ethnic groups who form ruling coalitions benefit from trade exposure and show higher values of welfare, trust and democracy. Discriminated ethnic groups, however, show significant reductions in welfare, trust and democratic values. Similar to Burgess et al. (2015), the findings indicate that the ruling elites capture the gains from trade, directly affecting livelihood and social cohesion of non-ruling ethnic groups. Given that many of these ethnic groups live around the border, and are thus connected to other countries creating trade, the evidence here suggests that they do not benefit from the trade they created.

These findings contribute to our understanding of Africa's long-run development and the important role its colonial history plays. In related work, Michalopoulos and Papaioannou (2016) show that ethnic groups split across country borders are poorer and lag behind non-split ethnic groups. Split ethnic groups were also less politically centralized in the pre-colonial period, which further emphasizes the fact that they exhibit lower levels of economic development today (Michalopoulos and Papaioannou, 2013). My findings suggest that split ethnic groups with large ethnic networks across borders benefit from increased trade activity, yet these gains from trade disproportionately accumulate with the ethnic groups that hold political power. This is suggestive of a mechanism that aligns with the insights of Dickens (2018), who documents evidence of ethnic favouritism within split groups throughout sub-Saharan Africa.[1] Overall, my results highlight a novel channel through which patterns of development have persisted throughout the African continent.

The paper also contributes to the understanding of how ethnic networks across borders affect trade and development for developing countries. While most empirical research uses a gravity-type equation (Anderson, 1979; Chaney, 2008), only few explicitly allow for subnational groups to determine trade. Felbermayr et al. (2010) establish that ethnic networks are significant determinants of trade in a variety of countries outside of Africa. For Africa, Abramson et al. (2020) estimate that ethnic borders introduce trade frictions approximately one quarter of the size of national borders and Copeland (2018) finds significant effects of ethnic linkages on bilateral trade. Focusing on developing countries,

[1]More broadly, evidence of ethnic favoritism in African politics is well documented in the literature (Frank and Rainer, 2012; Burgess et al., 2015; Kramon and Posner, 2016).

Tansey and Touray (2009) estimates the gravity equation for African countries and Osabuohien et al. (2019) for countries of the Economic Community of West African States. Most closely related is Aker et al. (2014) who show that the border effect on price dispersion between Niger and Nigeria is smaller if the border lies within one ethnic homeland. This paper advances this research by providing a simple empirical and theoretical model to identify the impact of ethnic networks on trade and migration in a gravity-type setup.

Finally, I contribute to the recent discussion on the distributional effects of trade. While it is clear that liberalizing trade generates winners and losers, identifying them empirically was near impossible until the advent of firm-level data. Engel et al. (2021) provides an overview of the distributional effects of trade across regions and demographic groups over time. At the firm level, Baccini et al. (2017) highlight how preferential trade agreements increase trade disproportionately for large firms. This evidence is corroborated in the developing countries setting, where Dhingra and Tenreyro (2020) evaluate agribusinesses providing access to farmers and show that while businesses gained, farmers in villages that produced policy-affected crops saw reductions in consumption. Using the staggered implementation of the Africa Growth and Opportunity Act, Desmet and Gomes (2023) show that trade access increases income in general, but decreases it for remote ethnic groups. In contrast to existing studies focusing on tariff reductions, I provide evidence how trade flows differential affect groups based on their power status within government. Thus, my findings add to the academic and policy debates on the distributional impacts of trade policies.

This paper is structured as follows. Section two develops and estimates a gravity equation that allows for continuous ethnic network sizes between countries and discusses its identification assumptions. Section three extends this framework to regional development and discusses under which assumptions I recover an unbiased estimate of aggregated trade on regional development. Here, I present results on regional development and elite caputre. Section four then analyses how ethnic networks affect trade flows and regional development. Section five concludes.

2 Ethnic Networks and Trade

For the first part of the analysis, I develop a gravity-type equation that allows for heterogeneous ethnic groups across multiple country pairs. I exploit the quasi-exogenous placement of borders to estimate this equation using pre-colonial ethnicity shares in each exporting country —and their connections to the importing country— to highlight the importance of ethnic networks in shaping and directing export flows.

2.1 Trade in Africa: A gravity equation

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 and the distance between two countries, as larger and more geographically close economies attract more trade flows. In this framework, the addition of a stock or flow of migrants is typically used to estimate the impact of cross-country networks on bilateral trade.

However, estimating the impact of such networks on bilateral trade between two developed and two developing countries is distinctly different. While migrants from developed countries often identify with their nationality, ethnic identification is a main factor in many developing countries.[2] Second, emigration from developing countries is driven by natural catastrophes, political instability, or economic factors, leading to more severe endogeneity concerns. These features of developing countries require a generalization of the standard empirical approach as well as exogenous variation to obtain a valid first stage for the shift-share analysis.

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

$$\log(X_{cd,t}) = \beta \log(PS_{d,t}) + \Gamma_{cd,t} + \delta_c + \delta_d + \varepsilon_{cd,t} \tag{1}$$

Here the log of exports from the exporting country c to the importing country d, $\log(X_{cd,t})$, is correlated with the population share of people from country of origin c in destination country d ($PS_{d,t}$). Controlling for exporter (δ_c) and importer (δ_d) fixed effects and bilateral characteristics ($\Gamma_{cd,t}$), β identifies the effect of the population share on the log of exports. The elasticity $\beta > 0$ indicates that trade flows increase if the trading partners share a larger network.[3]

Implicitly, equation (1) assumes that migrants to the importing country d identify with the nationality of their country of origin c.[4] While approximately true in developed coun-

[2] In the Afrobarometer Round 7 (2016-2018) 57.65% of respondents reply that their ethnic identity is at least as important as their national identity. Only 33% identify themselves solely by their nationality.

[3] Estimating equation (1) yields an elasticity of 1.89. A one percent increase in the number of people from country c in country d increases trade by 1.89%. This estimate is ten times larger than my baseline estimate in Table 2.

[4] The underlying equation is of the form $X_{cd,t} = PS_{d,t}^{\beta} = (Pop_{d,t,c}/Pop_{d,t})^{\beta}$. Here, the population of migrants from country c in country d at time t is denominated by the population size of country d at time t. The implicit assumption is that all migrants form c identify with country c, and not with a subgroup e. That is, $(Pop_{c,t,c}/Pop_{c,t})^{\beta} \approx 1$. Combining these yields $X_{cd,t} = PS_{d,t}^{\beta} = (Pop_{d,t,c}/Pop_{d,t} \times Pop_{c,t,c}/Pop_{c,t})^{\beta}$, and allowing

tries, the population structures in developing countries are more diverse. African countries combine a multitude of ethnic groups, each with their own identity and separated into multiple countries. Thus, allowing for multiple ethnic groups (e) from the set of ethnic groups (E) in each country $e \in E_c \subseteq E$, the general form of equation (1) is given by:

$$\log(X_{cd,t}) = \beta \log\left(\sum_{e \in E_c \cap E_d}^{E} PS_{c,t,e} \times PS_{d,t,e}\right) + \Gamma_{cd,t} + \delta_c + \delta_d + \varepsilon_{cd,t} \tag{2}$$

where $PS_{c,t,e} \in [0,1]$ is the population of an ethnicity e that is prevalent in each country pair cd, relative to the population of country c 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 ethnic groups if the country is prosperous.

The formulation of equation (2) is supported by two observations. First, it is the empirical equivalent of an otherwise standard model of international trade (Melitz, 2003; Chaney, 2008) that adds an ethnicity-specific fixed cost capturing lower entry costs into an export market for ethnically connected firms.[5] Second, the interpretation is equivalent to the search and matching literature if an exporter from country c can export more cheaply if she finds an importer in country d that is of the same ethnicity. Aggregating each firm's exports then yields the gravity-type equation (2).[6] 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

for a single subgroup e yields $X_{cd,t} = ({}^{Pop_{d,t,e}}/_{Pop_{d,t}} \times {}^{Pop_{c,t,e}}/_{Pop_{c,t}})^{\beta}$. Equation (2) then follows from the aggregation of all subgroups e. Details in Appendix E.

[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 Appendix E, I show that equation (2) follows if firms face a fixed cost of exporting

$$PS_{c,e}^{-\eta} f_{cd}$$

with $\eta \in [0,1)$ providing concavity for the impact of fixed costs f_{cd} 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 Chaney (2008) model with $\eta = 0$.

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

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 its 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 not the true 'ethnicity' specific effect, as it does not account for the variability in the exporting country.

2.2 Empirical specification

To obtain a first stage estimate of bilateral trade-flows that is only driven by exogenous factors, I estimate the following equation:

$$\log(X_{cd,t}) = \beta \, \log \text{Ethnic Connections}_{cd} + \Gamma_{cd} + \delta_c + \delta_d + \delta_t + \varepsilon_{cd,t} \qquad (3)$$

Here the 'Ethnic Connections' is defined by the sum of all ethnic match probabilities for all ethnic groups that are prevalent in both countries $PS_{c,e} \times PS_{d,e}\ \forall e \in E_c \cap E_d$ and constitutes the measure of ethnic similarity between a country pair. Every regression follows the standard in the trade literature and includes exporter (δ_c) and importer (δ_d) fixed effects and, where applicable, includes exporter-importer pair characteristics (Γ_{cd}).[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.

I obtain exogenous variation in ethnic shares using data containing the distribution of ethnic groups before colonialization. The geographic data provided by Murdock (1959)

[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 equation (2).
This model can be amended to allow for inter-ethnic trade. By assuming an increasing cost of trade for ethnic groups that are far away from each other, I confirm the baseline estimates for the entire sample of African countries.

[8]Exporter-importer pair characteristics include the log of the length of the border, log border fractionalization (Alesina et al., 2011), the log of the 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. I do not control for preferential trade agreements, as they are an outcome of ethnic connections (Table C.2). Controlling for this endogenous coregressors does not alter the point estimates. Conflict intensity is absorbed by country times year fixed effects in columns (3) and (4).

has been used extensively in the literature to study the relationship between slavery and trust (Nunn and Wantchekon, 2011), ethnic identification (Lowes et al., 2015) and conflict (Moscona et al., 2020) . 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.[9]

I estimate this equation at the country pair by year level.[10] For the dependent variable, the log of bilateral exports, I use UN comtrade data from the World Bank Integrated Trade Systems from 1990–2020. Since trade data do not capture unreported and informal trade, the literature has sometimes used price level differences instead (Aker et al., 2014). I use reported trade to estimate the effects for all countries, acknowledging that the point estimates are likely lower bounds on the true extent of exports between countries.[11]

The final sample consists of 46 African countries in 91 country pairs with 182 exporter-importer relationships that share a border. Due to unobserved trade the sample is further reduced to 169 observations from 1990–2020. Since the exploited variation is at the country-pair level, I follow the conservative choice and cluster the standard errors at this level. I report estimates using ordinary least squares due to the interpretative simplicity, but show the robustness using weighted least squares, a panel estimation with country specific year fixed effects, and the Poisson-Pseudo-Maximum-Likelihood estimator as suggested by Santos-Silva and Tenreyro (2006).

2.3 Identification strategy of bilateral exports

To estimate the effect of networks between countries on trade flows, the empirical approach usually uses flows or stocks of migrants. However, economic activity attracts

[9] 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.

[10] To avoid a Moulton (1986) type problem of inconsistent standard errors when the independent variable varies by country-pair and the dependent variable at the country-pair by year level, I also estimate equation (3) at the country-pair level. Here, every country is observed once as an exporter and once as an importer. 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. 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.

[11] If the data is split up into reported or unreported trade, the true estimate will be $\beta = \left(\beta^{reported} X_{cd}^{reported} + \beta^{unreported} X_{cd}^{unreported}\right) / (X_{cd}^{reported} + X_{cd}^{unreported})$. 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.

trade and migration flows similarly, leading to problems of reverse causality. In addition, borders are not set at random and instead reflect spheres of influence and historical economic activity, such that the direction of a potential omitted variable bias is unclear. To overcome the issue of reversed causation and omitted variable bias, I need to argue that (i) the local dispersion of ethnic groups and (ii) the borders between African countries are placed without the intention to increase trade, migration, or GDP in modern times.

First, in African countries, ethnic population shares are affected by a multitude of factors. Natural catastrophes, hunger, and civil conflicts contribute to the dispersion of people around the continent. Even without accounting for ethnic heterogeneity, these factors are often correlated with economic activity and threaten a causal identification of the network 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 reverse causality.

The standard approach in the literature uses past migration to instrument for networks as it has been shown that migrants follow their networks and settle in clusters in the importing country (Munshi, 2003; McKenzie and Rapoport, 2007, e.g.). This strategy solves the reverse causality problem if initial migrants were randomly placed in countries. In the African context, I allow for ethnic heterogeneity and counteract any potentially remaining issues of reverse causality and omitted variable biases by using the precolonial 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.

I show the precolonial distribution of 833 ethnic groups in Africa in Figure A.1. It is unlikely that selective sorting of ethnic groups occurred prior to independence as borders were decided by colonial powers and not ethnic groups. 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.

Second, contrary to European countries where borders reflect spheres of economic interests, African borders were drawn in 1884 at the Berlin conference. These borders do not reflect the interest of ethnic groups, but the interest of their colonial powers.[12] The

[12] 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 ethnic groups shows a similar pattern in 2008.

exogeneity of these borders has been extensively used in the 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 (Alesina et al., 2011). Where the geography was known and country borders could have been set to follow rivers or mountain ridges, the evidence in Figure A.2 still suggests no such pattern. Here, country borders, shown in black, rarely overlap with major rivers shown in blue.

I argue that these borders were arbitrarily drawn, do not reflect the interests of a specific ethnicity, and divide ethnic groups into more than one country. These split ethnic groups are likely to be different from other ethnic groups in terms of size or historical economic activity. In line with Michalopoulos and Papaioannou (2013, 2016), I show that the only determinant of an ethnic group being divided across two counties is its geographical size (Table 1). Using data on historical characteristics of tribes, I show that split ethnic groups 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 division into multiple countries (column 8).

To address concerns that these correlations influence the results, I only consider country borders where ethnic groups have been split when estimating the impact of ethnic networks on trade flows. I thus abstract from a comparison of influential ethnic groups with negligible ethnic groups and use a balanced sample across similar ethnic groups. Additionally, this procedure abstracts selection into whether countries share an ethnic group or are contiguous and focuses on the intensive margin only. When analyzing individual level data, these ethnic groups are added again.

2.4 Results at the bilateral level

The final sample to estimate the effect of ethnic networks on Trade consists of 46 neighboring countries between 1990 and 2020 that split 366 ethnic groups. All borders split at least one ethnicity, such that the comparison is strictly between countries that share ethnic groups and does not take into account countries that do not.[13] Since networks are calculated only using split ethnic groups, the analysis also does not compare ethnic groups

[13] Appendix E.1 relaxes this identification assumption and expands the theory to allow for inter-ethnic trade between all countries. The results are robust.

that are split to those that are not.

Then, any (imaginary) shift in the border of two countries only affects the fraction of the population belonging to a specific ethnicity $\in (0,1)$. Since these borders were drawn in 1884 without taking into considerations that these countries could become independent more than 60 years later, the population share of each ethnicity is essentially random, and so are the ethnic network across countries.

I present the reduced form effects from estimating equation (3) in Table 2.[14] Column (1) presents the raw correlations, controlling for exporter, importer, and year fixed effects. A 1% larger ethnic connection between two countries increases trade flows per capita by 0.176%. Supporting the exogeneity assumption of ethnic networks, the point estimate remains stable when adding country-pair controls (Column 2), and including exporter$\times$year and importer$\times$year fixed effects absorbing any variation from either exporting or importing country alone and isolating the variation to the country-pair (Column 3). In particular, these fixed effects capture population, gross domestic product, and conflict status for each country and year.

It is not clear that adding country-pair controls in columns (2)-(3) is the correct choice here. The exogenous process dividing countries and creating the ethnic connections between countries also determines all covariates. If the border is shifted, it changes the ethnic connections, but also linguistic similarity, border length, whether the border contains a river, mountain, and the distance between capitals. Thus, adding controls does not address an omitted variable bias but rather control for alternative channels through which the underlying exogenous process could have influenced trade flows. If border setting changed the linguistic similarity between two countries and this is the actual driver of exports, the point estimate on ethnic connections would decrease. However, as the point estimate remains stable (0.176 $\rightarrow$ 0.178) and the standard errors increase (0.052$\rightarrow$ 0.071), control variables only explain additional variation in exports, and not the impact of ethnic connections per se.[15]

Finally, I estimate equation (3) using a Poisson-Pseudo-Maximum-Likelihood (PPML) model in column four and report the same robust relationship. This standard estimator addresses the problem that trade flows are not normally distributed but follow a Poisson distribution, including zero trade. By assuming a Poisson distribution of realized trade values and estimating the relationship using a maximum likelihood, this estimator

[14]Reduced form results are presented as individual exposure to export flows is predicted in the next section.

[15]In a variance-decomposition analysis, the additional variance explained by control variables and ethnic connections over the fixed effects in column (3) can be attributed to 1/3 to the ethnic connections and 2/3 to eleven control variables; again highlighting the relative importance of ethnic connections.

is more efficient than the standard ordinary least squares. I thus report robustness to this estimation technique throughout the paper.[16]

The exploited variation is highlighted in Figure 1. Here, I plot the residual variation in trade against the residual variation in ethnic connections as in column 3 of Table 2. Even neglecting the smallest and largest values, a clear positive relationship with sufficient variation emerges.

These results are highly robust. They do not depend on the choice of the dependent variable, as total exports and exports per capita yield similar results (Table C.1). Also, while countries with ethnic connections are more likely to sign preferential trade agreements, controlling for this outcome variable does not alter the estimated impact (Table C.2). The results are also robust to estimation at the country-pair level, aggregating the data to that cross-sectional level. The estimates are less precise, but similar in magnitude (Table C.3). In addition, the estimates are unaffected by outliers, as dropping individual countries from the sample does not impact the main estimates significantly (Figure B.1).

2.5 Results at the ethnicity level

In the first set of results in Table 2, I estimated the gravity equation at the country-pair level, exploiting variation in the cross-border ethnic networks across all ethnic groups. To empirically assess each ethnicity's contribution to exports, I now estimate the same equation at the country-pair by ethnicity level. Instead of aggregating all ethnic networks for each ethnicity, I estimate the impact of each ethnicity's network separately on the variation in export flows that remained unchanged from Table 2:

$$\log(X_{cd,t}) = \beta \log\left(PS_{c,e} \times PS_{d,e}\right) + \Gamma_{cd} + \delta_{c,t} + \delta_{d,t} + \delta_{e,c} + \varepsilon_{cde,t} \tag{4}$$

Here, I predict export flows between two countries $\log(X_{cd,t})$ by the ethnic connections between every ethnicity pair $(PS_{c,e} \times PS_{d,e})$, controlling for pre-defined country-pair characteristics Γ_{cd} and country$\times$year fixed effects $(\delta_{c,t}, \delta_{d,t})$ absorbing any time-varying variation in either country.[17] I include ethnicity$\times$country fixed effects $\delta_{e,c}$ to absorb any variation coming from the ethnicity in the exporting country. The identifying assumption remains unchanged; conditional on fixed effects, the network size of ethnic groups $(PS_{c,e} \times PS_{d,e})$ is essentially random.

[16] I only consider reported zero trade as actual 'zero' trade in column (4) as missing trade data could be due to non-reporting, non-trading, or data issues. Including missing trade as 'zero' trade does not change the point estimates.

[17] I do not control for preferential trade agreements as they are an outcome of ethnic connections. Controlling for preferential trade agreements does not alter the point estimates (Table C.2).

The idea behind this exercise is to predict each ethnic groups' contribution to shaping trade flows. Since the value of exports does not vary between ethnic groups of the same country pair, $\hat{\beta}$ will only estimate the average elasticity between export flows and all ethnic groups. When predicting trade flows $\widehat{X_{cde,t}}$, this elasticity $\hat{\beta}$ is then multiplied with each ethnicity's network $(PS_{c,e} \times PS_{d,e})$ to generate variation in each ethnicity's contribution to exports. Larger ethnic groups will have larger export flows and smaller ethnic groups smaller, governed by their network size $\times$ the average elasticity $\hat{\beta}$.

I present the results of estimating equation (4) in Table 3. Countries whose ethnic connections are one percent larger, trade on average 0.013% more. While this estimate is a magnitude smaller than the estimate in Table 2, it is the result of estimating the same equation at a higher level of resolution. If each ethnicity has 1% more connections, these effects accumulate to the effect estimated at the country-pair level.

Controlling for ethnicity fixed effects (column 2) and ethnicity $\times$ country fixed effects increases the point estimate of the OLS to 0.055. In the most demanding specification of column (3), I only exploit variation between two countries, as the effect of a larger population share in the exporting country is absorbed by the fixed effects. Again, the results are robust to various definitions of the dependent variable and estimating equation (4) using PPML.

To assign export flows to each ethnicity, I predict $\widehat{X_{cde,t}}$ using the empirical specification (4) estimated in columns (3) and (4). Taking into account fixed effects, the predicted values then capture the expected value of bilateral exports without ethnic connections $\hat{\delta}_{c,t}, \hat{\delta}_{d,t}$ and the predicted export flows contributed by the population share of ethnicity e in the exporting country $\hat{\delta}_{e,c}$. Thus, the only variation between two country-pairs and ethnicity-pairs then arises from the strength of ethnic connections, and its impact on export flows $\hat{\beta}$.

The remaining variation is significant and shown in Figure 2. I plot realized trade against predicted trade for all trading partners of Tanzania in Figure 2a and observe significant variation across all country-pairs. Figure 2b plots predicted trade for Tanzania's eight neighboring countries. Within each country, every ethnic group has different predicted trade flows: ethnic groups that have a strong connection to Burundi will have larger predicted trade volumes, whereas ethnic groups that have a weak (or no) ethnic connection to Burundi have smaller predicted trade volumes, only governed by Tanzania's and Burundi's economies.[18]

A final step then aggregates these predicted trade flows to the country-by-ethnicity

[18] In standard trade models, this multilateral resistance term and captures how much each country trade's with each other country solely because of their economic power.

level as bilateral export flows cannot be attributed to single regions.

3 Trade and regional development

To estimate the relationship between trade and regional development, I require variation in each individual's exposure to trade. In this section, I discuss the estimation strategy and identification assumption, before presenting the results on regional development.

3.1 Empirical specification

Since the object of interest, regional development, is measured at the country-by-ethnicity level, I estimate the following equation:

$$Y_{ce,t} = \beta Trade\ Exposure_{ce,t} + \delta_{c,t} + \delta_{e,c} + \varepsilon_{ce,t} \quad (5)$$

Where $Y_{ce,t}$ captures regional development as either satellite data capturing luminosity (Elvidge et al., 2021) or individual welfare from the Afrobarometer surveys. *Trade Exposure*$_{ce,t}$ captures the exposure of ethnicity e in country c to exports per capita. $\delta_{c,t}$ and $\delta_{e,c}$ are country-by-year fixed effects and ethnicity-by-country fixed effects to capture national level changes and the level impact of ethnic groups in each country. These fixed effects mirror those in equation (4), as in they capture the total export volume, gross domestic product, and population of country c in time t as well as the general impact on regional development of ethnicity e in country c.[19]

Trade Exposure can be calculated in two —a priori— equally reasonable procedures following the shift-share literature. The key difference lies in the construction of the *shift*; it is either constructed as an ethnicity-level aggregate or a national-level aggregate. In the main tables of this paper, I chose to first predict each ethnicity's impact on trade flows by predicting $\widehat{X_{cde,t}}$ from an ethnicity-level gravity equation (4) before aggregating bilateral trade flows to its d-neigboring country to the country-by-ethnicity level $\widehat{X_{ce,t}}$. This object is a shifter in trade flows constructed from exogenous variation in network strength across two countries and fixed effects. However, even though $\widehat{X_{ce,t}}$ now varies at the country-by-ethnicity level, its prediction still contains all trade flows between country c's other trading partners. I thus control for country-by-year fixed effects $\delta_{c,t}$ to control for aggregate trade intensity and isolate the variation coming from the variation in network strength.

[19]A potential concern is differential population density across ethnic groups. Ethnicity-by-country fixed effect capture everything that is constant within the period of observation, and thus population density.

An alternative way to construct the *shift* is from an national-level aggregate. Here, $\widehat{X_{cd,t}}$ is predicted from the standard gravity equation (2), before aggregating bilateral trade flows to its d-neigboring country to the country level $\widehat{X_{c,t}}$. This procedure exploits the same exogenous variation in network strength across two countries, but yields less precise estimates as the variation comes only from the aggregate. In Appendix D, I replicate the entire analysis using this *shift*; the conclusions are unaffected as the exploited variation has not fundamentally changed.

To attribute the *shift* to each ethnicity, I follow the shift-share literature and multiply it with the share in population of ethnicity e. Following Borusyak et al. (2022), I cluster the standard error at the level of variation, which after aggregating the data comes from each ethnicity and country.[20]

3.2 Identification assumptions

In this shift-share setting, exogeneity from the individual's perspective comes from two sources: predicted exports, shifted by quasi-exogenous cross-border networks, and second, that ethnic groups are quasi randomly placed in different countries. The ethnicity's or individual's decision cannot impact trade flows or their population shares in the home country, with the aim to increase one's development.

This assumption is likely fulfilled for the shifter in exports: the cross-border network size. Conditional on exporting and importing country fixed effects, the size of the ethnic network across the border $\left(\sum_{e\in E_c\cap E_d}^{E} PS_{c,t,e} \times PS_{d,t,e}\right)$ is arguably as good as random if country borders were set without taking the distribution of ethnic groups into account. This assumption is best illustrated at the example of a straight border. If the border is a straight line, it slices apart ethnic borders which had been determined by agricultural suitability, political economy, or unobserved variables. Then, a slight shift in the straight line redistributes people from one country to the other.[21]

[20]Borusyak et al. (2022) suggest to cluster standard errors either at the level of the *shift* or the *share*. In this case, the variation comes from ethnic groups being split by country borders at the country-pair level. It is therefore nontrivial at which level to cluster. I use ethnicity and country as this is closest to the exploited variation.
Alternatively, one could argue that the shifter is the exogenous part of the shift-share and thus the level of variation, in which case the standard errors need to be clustered at the country-year level. The standard errors are, however, almost identical and thus not reported.

[21]In the empirical analysis I do assume that both home- and foreign ethnic populations have the same impact on trade flows. However, Appendix E shows that theory predicts the foreign ethnic populations impact to be smaller. Its exponent ranges from zero to one and is determined by the elasticity of substitution between goods, the degree of firm heterogeneity and the importance of ethnic connections decreasing the fixed costs of exporting. I estimate this model in Table E.1 and show the relationship remains unchanged.

It is however likely that the population shares themselves correlate with social standing, power structures, income or other unobservables that might influence economic activity. Minority populations are often discriminated against and countries with a large number of ethnic groups, each with a low fraction of the population, are often more unstable. Yet, the variation I exploit in equation (5) is net of these confounding factors as country×ethnicity fixed effects absorb their impact on regional economic development. In particular, they absorb whether an ethnicity is split or not, which is a powerful determinant of political power (Michalopoulos and Papaioannou, 2016).

As is standard in the shift-share literature, the variation exploited is the interaction: quasi-exogenous increases in exports (shift) interacted with population shares. I thus use the exogenous process determining the borders of African countries in two 'stages': I exploit cross-border variation in network size ($PS_{c,e} \times PS_{d,e}$) to obtain variation in the size of exports and variation in population shares ($PS_{c,e}$) to assign these to individuals. The coefficient on trade exposure can thus be interpreted, conditional on this ethnic group's status in society, as how does an increase in trade affects its economic and social standing.

3.3 Trade and nighttime lights

In Table 4, I present results of estimating equation (5) on nighttime light data using Version 2.1 of the VIIRS nighttime light series (Elvidge et al., 2021). Available from 2012–2020, this data is an update to the older nighttime light series and solves many of its original problems (Elvidge et al., 1997).[22] On average, ethnic homelands show a luminosity of 0.184 with only 3.3% of pixels (500m×500m) actively lit.

Ethnic homelands predicted to be more exposed to trade show consistently lower nighttime luminosity, and thus regional development. As the distribution is heavily skewed, I show robustness using both the continuous measure of nighttime luminosity (columns 1 and 2) and the fraction of pixels lit (columns 3 and 4). Increasing trade exposure by one standard deviation decreases average luminosity by 3.2% and the fraction of lid pixels by 4.9%.[23] As the point estimate derives itself from the interaction of trade flows and population shares, two interpretations of the point estimate arise. First, in terms of trade flows, a one standard deviation increase in Trade flows decreases average luminosity by 0.6%

[22]See for instance Michalopoulos and Papaioannou (2013). Since the combination of both would lead to a discontinuous jump in nighttime luminosity and both data have different resolutions, I use the VIIRS as it is closer in time to most observations in the Afrobarometer data.

[23]Mean trade exposure is -0.1 and a standard deviation is 0.365 on the logarithmic scale. The effect is calculated as $-3.2\% = -0.0086 \times {}^{0.365}/_{-0.1}$. The effects are slightly larger when trade is measured in levels at -6.3%. For the fraction of pixels lid, I calculate $-4.9\% = \frac{-0.0043 \times 0.365 + 0.032}{0.032} - 1$, where 0.032 is the average fraction of pixels lid.

and the fraction of lid pixels by 26.6%. Second, in terms of population, a one standard deviation increase in the population share of ethnicity e decreases average luminosity by 2.2% and the fraction of lit pixels by 1.3%.

This result cannot be explained by time-varying country factors such as population density or ethnic-level confounders such as pre-colonial distributions (Michalopoulos and Papaioannou, 2013) or the ethnic group being split (Michalopoulos and Papaioannou, 2016); country $\times$ year and ethnic $\times$ country fixed effects absorb these confounders completely. These results are also not driven by outliers as dropping countries individually does not alter the estimate significantly (Figure B.3-B.4). The results are also robust to alternative measures of bilateral exports (Table C.4). Finally, using the dyadic prediction of trade (3) instead of ethnic-level trade prediction yields slightly larger point estimates (Table D.2).

Table 4 thus provides evidence that economic outcomess accruing to the ethnic groups are not proportional to their exported share in international trade.To understand this discrepancy, I analyze the impact of trade flows at an even finer level of aggregation.

3.4 Trade and individual welfare

In the final analysis on regional development, I relate ethnicity-level trade exposure to household welfare, trust, and democratic values. *Trade Exposure* is measured as before; it is the interaction of country-by-ethnicity volumes of trade with the population share of ethnicity e. Individual-level outcomes are obtained from the Afrobarometer Rounds 1-7 georeferenced data (BenYishay et al., 2017). As seen in Figure 3, I link these two data by the precolonial distribution of ethnic groups from Murdock (1959).

This procedure has several advantages. First, the distribution of ethnic groups is determined prior to data collection of trade and individual outcomes. Second, since the geographic distribution is also used to obtain exogenous variation in trade flows, it lends itself naturally as a basis for both data. Lastly, the assignment is not based on recorded ethnic groups in the Afrobarometer, which could be subject to a classification error, misreported, or not collected.

Thus, assigning each individual in the Afrobarometer an ethnicity from the precolonial distribution of ethnic groups, I link individuals to trade exposure at the country-by-ethnicity level and estimate the following equation:

$$Y_{cei,t} = \beta Trade\ Exposure_{ce,t} + \delta_{c,t} + \delta_{e,c} + X_i + \varepsilon_{ce,t} \tag{6}$$

Where individual i of ethnicity e in country c is exposed to its country and ethnicity's trade per capita in year t via $\textit{Trade Exposure}_{ce,t}$, controlling for country-by-year fixed effects $\delta_{c,t}$ and ethnicity-by-country fixed effects $\delta_{e,c}$ to capture national level changes and the level impact of ethnic groups in each country. In addition, I control for age, sex, urbanity, and education in X_i. Standard errors remain clustered at the level of variation, the country-by-ethnicity level, to avoid a Moulton (1986)-type problem.

One drawback of the Afrobarometer survey data is that it does not contain a measure for income $Y_{cei,t}$. To overcome this problem, I identify nine questions that cover topics of individual welfare, eleven that cover trust, and three that capture democratic values throughout all rounds. I present the variables and their raw correlations with trade exposure in Table C.5.

As these variables only share a common component that no single variable captures, each of the point estimates would be subject to concerns regarding multiple hypothesis testing. To address these concerns, I conduct two exercises. In the main Tables, I follow Anderson (2008) and standardize each question within the categories and sum the standardized outcomes, weighting each question by the inverse of the covariance matrix of the standardized outcomes. By accounting for the covariance between individual questions, I obtain a more accurate measure than alternative procedures that use an equally weighted average. The indices address concerns of multiple hypothesis testing and aggregate changes in preferences that individual questions only measure imperfectly. Second, I report p-values adjusted for multiple hypothesis testing in Table C.5.

Table 5 presents the results. Both predicting trade using the OLS or PPML method confirm earlier results on nighttime luminosity; household welfare decreases with trade exposure. A one standard deviation increase in trade exposure decreases household welfare by 11.4% of a standard deviation, trust by 9% of a standard deviation and democratic values by 28.9% of a standard deviation.

Again, these results are not affected by outliers as dropping countries individually does not alter the estimate significantly (Figure B.5-B.7). The results are robust to predicting exports at the dyadic level instead (Table D.3). Finally, the results are robust to using total exports or exports per capita instead (Table C.6).

3.5 Elite capture of the gains from trade

Where do the gains from trade accumulate? I argue that ethnic networks decrease information frictions across borders, but that the industrial structures that exploit this advantage and produce export-goods are not built in the ethnic homelands, but with the ethnicity that is in power.

To assess such elite capture of gains from trade, I use the ethnic power relations data and identify ethnic groups that are discriminated against, not in power, or in political control. I then link each individual in the Afrobarometer data to their ethnicity's power status and conduct a heterogeneity analysis in Table 6.

The results provide evidence in favor of elite capture. Individuals from ethnic groups that are in political control benefit from trade exposure, while other groups lose. A one standard deviation increase in trade exposure increases household welfare by 65% of a standard deviation. The effect on trust and democratic values remain positive for this group, while the other groups show decreasing levels with increased trade exposure.

This difference is particularly stark when exposed to increased manufacturing exports. If the group is not in power, increased trade exposure decreases welfare, trust and democratic values significantly. If, on the other hand, the group leads the country, increased trade exposure increases welfare, trust and democratic values.

Ethnic groups belonging to cross-border ethnic networks are, by construction, at the border of countries and are less likely to be in power of an entire country (Table A.1). However, even though these ethnic groups help bridge the gap between two countries and increase trade, the gains from trade are concentrated among the group that is in power. This likely explains the negative impacts on trust and democratic values. Being left behind by the elites that govern the country, they lose trust and faith in democratic progress.[24]

4 Mechanisms

In this section, I provide evidence consistent with the hypothesis that ethnic networks decrease information frictions across borders, but that the industrial structures that exploit this advantage and produce export-goods are not built in the ethnic homelands, but with the ethnicity that is in power. I show that ethnic networks increase migration, which in turn increases trade flows. These trade flows are centered in the agricultural and manufacturing sector. In contrast to exports in the agricultural sector, exports in the

[24] A similar effect can be seen when analyzing DHS data. The available wealth index captures stable assets and is constructed as a composite measure of a household's cumulative living standard. It is calculated using easy-to-collect data on a household's ownership of selected assets, such as televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities, combined as a principle component. It is thus a measure of fixed wealth. It also oversamples urban communities, at least relative to the Afrobarometer.
On average, trade exposure is positively related to wealth, but broken down by initial wealth categories, Appendix Figure C.1 shows how the poorest parts of society do not benefit from trade, and only the richest quintile (DHS Definition), benefit.

manufacturing sector are more sensitive to information frictions and do reduce regional development; thus supporting the hypothesis of elite capture of the gains from trade.

First, I document how ethnic networks increase migration across borders. A large literature has shown that migration shapes and directs trade flows (Gould, 1994; Rauch and Trindade, 2002; Dunlevy, 2006; White, 2007; Bandyopadhyay et al., 2008; Peri and Requena-Silvente, 2010; Felbermayr and Toubal, 2012). This extends to subnational ethnic identification in the case of developed countries (Felbermayr et al., 2010) as well as for the US bilateral trade, where Parsons and Vézina (2018) use Vietnamese refugees and Cohen et al. (2017) Japanese internment camps to estimate the elasticity between migration and US Trade.

One way migration can affect trade flows between nations is by reducing information frictions (Allen, 2014; Steinwender, 2018), another is by trading preference goods across borders. In Africa, preference goods could be agricultural products that are traditionally produced in one part but not the other and thus need to be traded. It is however unlikely that aggregate statistics capture such small scale trade, especially when country borders across ethnic groups are not enforced due to a lack of state capacity. If ethnic networks reduce information frictions, we would observe higher levels of migration and exports in sectors that benefit from reducing information frictions between markets.

I begin by estimating the elasticity between migration flows and aggregate trade flows in Table 7. Odd columns report the first stage correlation between ethnic connections and migration in 1960, and even columns report the two-stage least-squares estimate on trade when instrumenting migration with ethnic connections. A 1% larger share of ethnic groups split between two countries increases migration in 1960 by 0.34–0.26%, depending on the extent of control variables added. The F-statistics on the relevance of the instrument range from 29.98–10.57 in the most demanding specification.[25]

Instrumenting migration with ethnic connections, I consistently estimate an elasticity between 0.52 and 0.69. This suggests that policies increasing migration by 1% induce an increase in export flows by 0.69% in the most demanding specification with exporter$\times$year and importer$\times$year fixed effects absorbing any variation from either exporting or importing country. Importantly, not only migration-facilitating policies increase trade, also disproportionate population growth of the migrant community, in either the exporting or importing country, increases trade flows.

Second, I show that increased exports are concentrated in the manufacturing sector that is likely more senstive to information frictions. If ethnic networks only capture (pre-

[25] Estimations for each decade of migration, as well as pooled across all decades, are found in C.7. Results do not differ.

existing) trade in preference goods, the effects would be concentrated in the agricultural sector. If ethnic networks, however, also transmit information about local market conditions, they should have a larger impact in sectors that benefit more from reducing information frictions. By having more individuals from your own group, frictions in communication are lower and trust is higher, increasing exports. Indeed, Figure 4 confirms this pattern as the OLS estimate is significant only in the agricultural and manufacturing sector. Since these estimates are the reduced form effect of ethnic networks on trade flows and ethnic networks consistently predict migration in the first stage, the second stage mechanically implies that migration increases exports especially in the manufacturing sector.

In Table 8, I analyze how these sectoral differences affect regional development. In Panel A, I show that while agricultural exports do not impact nighttime luminosity, manufacturing exports decrease the fraction of lit pixels significantly. This pattern continues when looking at individual exposure to agricultural or manufacturing exports in Panel B. The effect sizes are not statistically different from the effect of aggregate exports (cf. Table 5, suggesting that a large part of the negative impact of trade exposure on individual welfare is due to increasing manufacturing exports.

In Panels C and D, I show that both agricultural and manufacturing exports decrease trust and democratic values for affected individuals. The differential impact of agricultural and manufacturing exports on regional development and preferences might be explained by the different nature of production. If ethnic networks increase agricultural exports, the production of these goods is likely still in the ethnic homelands. And since fields emit no light and regional farmers benefit from this production, regional development is not negatively impacted by increasing trade. Manufacturing in contrast can be located everywhere, is capital intensive, and only benefits a select few. In this case, powerful interest groups could utilize the ethnic networks to increase trade, but locate production in their vicinity.

5 Conclusion

General trade theory predicts that the welfare impact of trade is positively affected by trade openness (Arkolakis et al., 2012). Across most trade models, the unanimous policy recommendation is thus to liberalize trade; a recommendation eagerly adopted by lenders and multinational organizations: Trade policies are a significant part of 95% bilateral and multinational agreements of African countries today (Egger and Larch, 2008).

Empirical evidence for these countries, however, has been hard to identify. In this paper I show that the opposite is true. Across all African countries, individuals are worse off when trade increases and their ethnic group is not in power. Policies should thus adapt and ensure that the 'gains from trade' are more equally shared and do not disproportionately benefit individual groups. The resulting loss in trust and democratic values indicate that future economic and societal development might suffer, increasing the chance for conflict (Desmet and Gomes, 2023).

References

Abramson, S. F., B. Cooley, and B. Lacina (2020). How wide is the ethnic border. *Working Paper*.

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*.

Allen, T. (2014). Information frictions in trade. *Econometrica 82*(6), 2041–2083.

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

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012, February). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

Autor, D. H., D. Dorn, and G. H. Hanson (2013, October). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review 103*(6), 2121–68.

Autor, D. H., D. Dorn, G. H. Hanson, and J. Song (2014, 09). Trade Adjustment: Worker-Level Evidence *. *The Quarterly Journal of Economics 129*(4), 1799–1860.

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

Baccini, L., P. M. Pinto, and S. Weymouth (2017). The distributional consequences of preferential trade liberalization: Firm-level evidence. *International Organization 71*(2), 373–395.

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

BenYishay, A., R. Rotberg, J. Wells, Z. Lv, S. Goodman, L. Kovacevic, and D. Runfola (2017). Geocoding afrobarometer rounds 1 - 7: Methodology & data quality. Technical report, AidData. Available online at http://geo.aiddata.org.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies 89*(1), 181–213.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró 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*.

Cohen, L., U. G. Gurun, and C. Malloy (2017). Resident networks and corporate connections: Evidence from world war ii internment camps. *The Journal of Finance 72*(1), 207–248.

Copeland, C. (2018). Bridging new divides: The impact of international ethnic linkages on bilateral trade in Africa. *Working Paper*.

Desmet, K. and J. F. Gomes (2023, January). Ethnic remoteness reduces the peace dividend from trade access. Working Paper 30862, National Bureau of Economic Research.

Dhingra, S. and S. Tenreyro (2020, February). The rise of agribusiness and the distributional consequences of policies on intermediated trade. CEP Discussion Papers dp1677, Centre for Economic Performance, LSE.

Dickens, A. (2018). Ethnolinguistic favoritism in African politics. *American Economic Journal: Applied Economics 10*(3), 370–402.

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.

Egger, P. H. and M. Larch (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics* (2), 384–399.

Elvidge, C., M. Zhizhin, G. T., and T. J. Hsu FC (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing 13*(5), 922.

Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis (1997). Mapping city lights with nighttime data from the dmsp operational linescan system. *Photogrammetric Engineering and Remote Sensing 63*(6), 727–734.

Engel, J., D. Kokas, G. Lopez-Acevedo, and M. Maliszewska (Eds.) (2021). *The Distributional Impacts of Trade : Empirical Innovations, Analytical Tools, and Policy Responses.* Washington, DC: World Bank.

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.

Frank, R. and I. Rainer (2012). Does the leader's ethnicity matter? ethnic favoritism, education, and health in sub-saharan africa. *The American Political Science Review 106*(2), 294–325.

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.

Kramon, E. and D. N. Posner (2016). Ethnic favoritism in education in kenya. *Quarterly Journal of Political Science 11*(1), 1–58.

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

Lowes, S., N. Nunn, J. A. Robinson, and J. Weigel (2015, May). Understanding ethnic identity in Africa: Evidence from the implicit association test (iat). *American Economic Review 105*(5), 340–45.

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.

Moscona, J., N. Nunn, and J. A. Robinson (2020). Segmentary lineage organization and conflict in sub-saharan Africa. *Econometrica 88*(5), 1999–2036.

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

Osabuohien, E. S., U. R. Efobi, J. T. Odebiyi, O. O. Fayomi, and A. O. Salami (2019). Bilateral trade performance in west Africa: A gravity model estimation. *African Development Review 31*(1), 1–14.

Parsons, C. and P.-L. Vézina (2018). Migrant networks and trade: The Vietnamese boat people as a natural experiment. *The Economic Journal 128*(612), F210–F234.

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. 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*.

Steinwender, C. (2018, March). Real effects of information frictions: When the states and the kingdom became united. *American Economic Review 108*(3), 657–96.

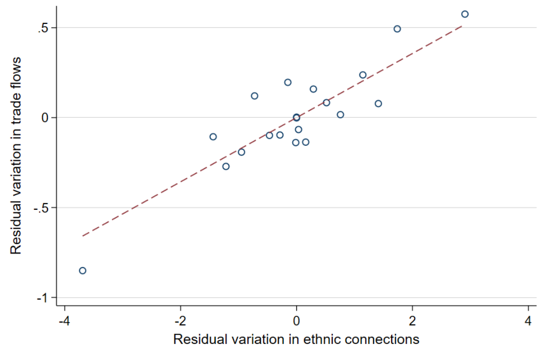
Tansey, M. and A. Touray (2009, 03). The gravity model of trade applied to Africa. *International Business & Economics Research Journal Volume 9*.

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.

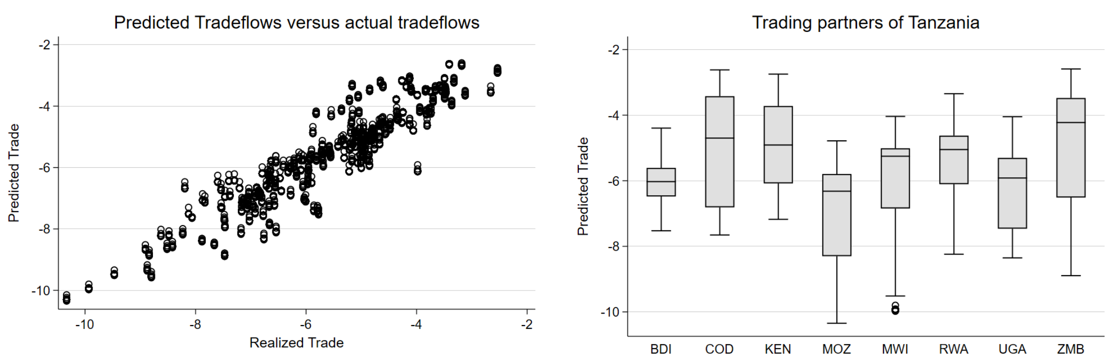
6 Figures

Figure 1: Identification: Plotting residualized variables:



Notes: In this binned scatter plot, I plot the residualized variation in exports per gdp per capita (y-axis) against the residualized varation in ethnic connections (x-axis) using equation (4). The slope coefficient is shown in Table 2 column 3.

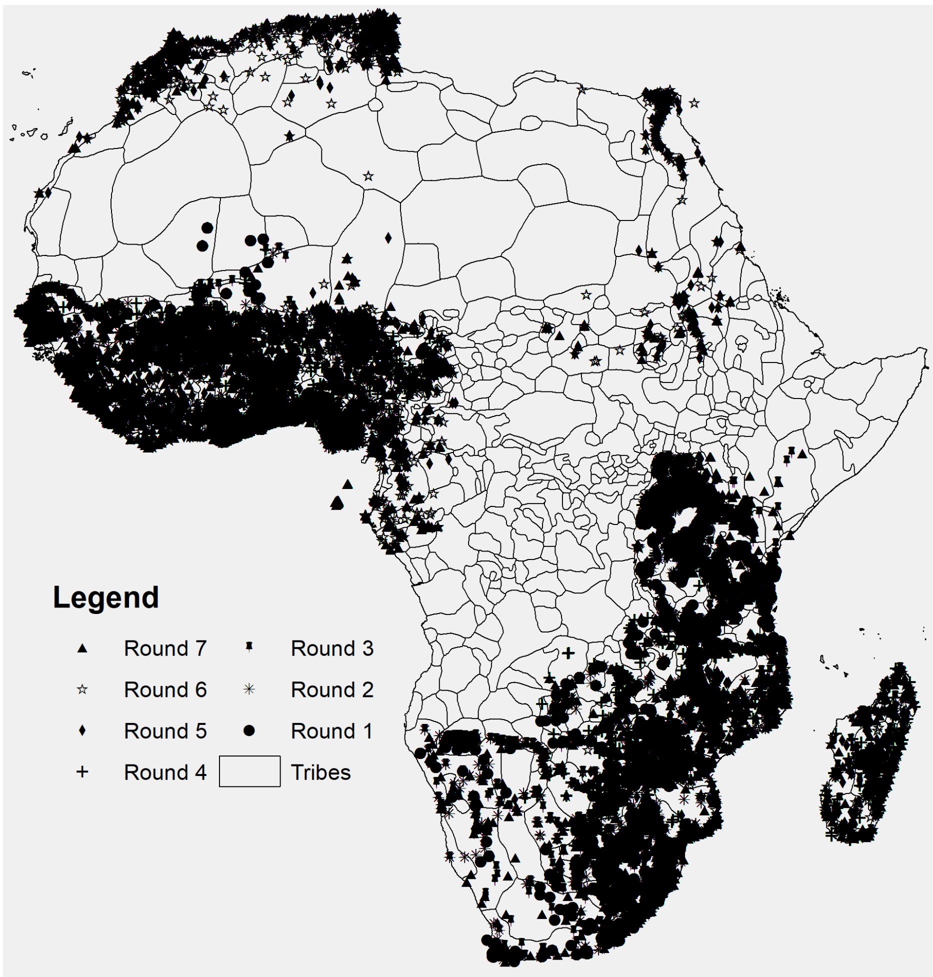
Figure 2: Predicted exports from Tanzania to its neighbors



(a) Predicted against realized trade flows **(b)** Predicted trade flows per importer

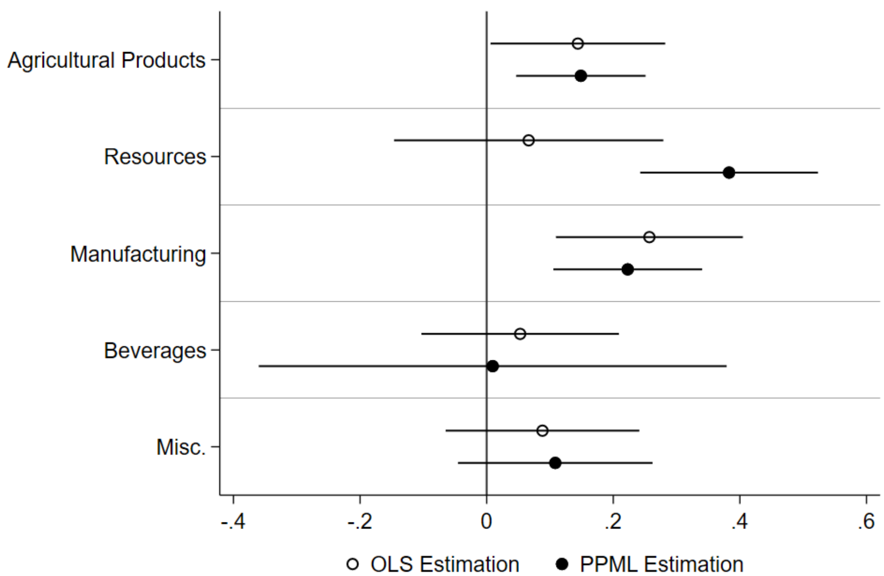
Notes: The left panel plots realized trade against predicted trade for Tanzania to all its neighbors. Predicted trade is defined as $\log(\widehat{X_{cde,t}})$ from equation (4). The right panel expands upon this and plots the box plots of variation for every trading partner. The variation between $\log(\widehat{X_{cde,t}})$ and $\log(\widehat{X_{cd,t}})$ comes from the size of the ethnic network of ethnicity e in Tanzania and the trading partner, as well as yearly variation.

Figure 3: Individuals in the Afrobarometer and the pre-colonial distribution of ethnic groups.



Notes: This maps plots every survey respondent in the Afrobarometer Rounds 1-7 within the ethnic homelands, as defined by the Murdock map.

Figure 4: Sectoral differences in the importance of ethnic networks



Notes: This figure estimates separate OLS and PPML models for the value of sectoral exports per capita per GDP.

7 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$\geq$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: Ethnic connections and trade flows

	Exports per capita			
	(1) OLS	(2) OLS	(3) OLS	(4) PPML
Ethnic connections	0.176^{***}	0.186^{***}	0.178^{**}	0.181^{***}
	(0.052)	(0.047)	(0.071)	(0.061)
	[0.039]	[0.056]	[0.056]	[0.063]
Country-pair controls		Yes	Yes	Yes
Exporter and importer $\times$ year fixed effects			Yes	Yes
Observations	4,195	4,195	4,195	4,198

In this table, I show that ethnic connections predict bilateral exports between countries. *Ethnic connections* are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. The main dependent variable are the logarithm of bilateral exports per current capita in the years 1992–2018. I add the following country-pair controls in columns (2)-(4): log length of the border, log distance between capitals, a binary variable indicating whether parts of the border is a river, a mountain above 1,000 or 2,000 meter. The number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, log border fractionalization (Alesina et al., 2011) and Linguistic and genetic distance $\in [0, 1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). In column (3) and (4) I add importer$\times$year and exporter$\times$year fixed effects to capture all time-varying variables at the country level. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. OLS and PPML denote the estimation method. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Ethnic connections and trade flows:
Estimation at the ethnicity level

	Exports per capita			
	(1) OLS	(2) OLS	(3) OLS	(4) PPML
Ethnic connections	0.013^{***}	0.032^{***}	0.055^{***}	0.014^{***}
	(0.004)	(0.008)	(0.015)	(0.005)
	[0.008]	[0.015]	[0.024]	[0.007]
Country-pair controls	Yes	Yes	Yes	Yes
Exporter and importer × year fixed effects	Yes	Yes	Yes	Yes
Ethnic fixed effects		Yes	Yes	Yes
Ethnic × exporter fixed effects			Yes	Yes
N	135,314	135,314	135,314	135,545

In this table, I show that ethnic connections predict bilateral exports between countries, even at the ethnicity level. Instead of estimating the gravity equation at the bilateral level (cf. Table 2), I estimate trade flows exploiting variation at the ethnicity × country-pair level. *Ethnic connections* are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. The main dependent variable are the logarithm of bilateral exports per current capita in the years 1992–2018. Country-pair controls defined in Table 2 are added throughout all columns. Importer×year and exporter×year fixed effects are included throughout all columns to capture all time-varying variables at the country level. In columns (2), I add ethnicity fixed effects and column (3) and (4) control for ethnicity × exporter fixed effects, capturing everything that is constant for the ethnicity in the exporting country. This specification thus only exploits variation in the intensity of connections across multiple trading partners. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. OLS and PPML denote the estimation method. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The effect of bilateral exports on economic development:
Satellite imagery

	Nighttime lights		Fraction pixels lid	
	(1)	(2)	(3)	(4)
	OLS	PPML	OLS	PPML
Panel A: Exports per capita				
Trade Exposure	-0.009**	-0.008**	-0.004**	-0.004**
	(0.004)	(0.003)	(0.002)	(0.002)
Country × year fixed effect	Yes	Yes	Yes	Yes
Country × ethnicity fixed effects	Yes	Yes	Yes	Yes
Observations	10,032	10,032	10,032	10,032
Mean dependent variable	0.068	0.068	0.033	0.033

In this table, I show how predicted exports impact economic development as measured by nighttime luminosity. *Trade Exposure* is defined as the predicted trade flows per capita (cf. Table 3, columns (3) & (4)), aggregated to the exporter level and interacted with the population share of this ethnicity. Country × year fixed effects then account for total trade flows of this exporting country and Country × year fixed effects for the size and impact of each ethnicity in this country. *Trade Exposure* is identified only from the interaction of the two. *Nighttime lights* is defined as the log of the continuous measure from the VIIRMS data plus one. It calculates the average luminosity of each ethnicity-country-year observation. *Fraction pixels lid* is defined from the same source at the same level but calculated as the fraction of pixels not zero. OLS and PPML denote the estimation method to predict exports. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The effect of bilateral exports on economic development:
Individual data from the Afrobarometer

	Household wealth		Household trust		Democratic values	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	PPML	OLS	PPML	OLS	PPML
Panel A: Exports per capita						
Trade Exposure	-0.052***	-0.047***	-0.041**	-0.050**	-0.132**	-0.117**
	(0.020)	(0.017)	(0.020)	(0.024)	(0.056)	(0.050)
Country × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country × ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	217,809	217,809	217,809	217,809	217,809	217,809

In this table, I assess the gains from trade for each individual. *Household wealth*, *Household Trust*, and *Democratic values* are composite scores from variables in the Afrobarometer Rounds 1-7. Geolocated data is used to place individuals in ethnic regions and countries and assign trade exposure. *Trade Exposure* is defined as the predicted trade flows per capita (cf. 3, columns (3) & (4)), aggregated to the exporter level and interacted with the population share of this ethnicity. Country × year fixed effects then account for total trade flows of this exporting country and Country × year fixed effects for the size and impact of each ethnicity in this country. *Trade Exposure* is identified only from the interaction of the two. Individual controls are age, gender, education, and urbanity. OLS and PPML denote the estimation method to predict exports. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: The effect of bilateral exports on economic development:
The effect of political power

	Household wealth		Household trust		Democratic values	
	(1) OLS	(2) PPML	(3) OLS	(4) PPML	(5) OLS	(6) PPML
Panel A: Exports per GDP per capita						
Trade Exposure	-0.046	-0.040	-0.068**	-0.098***	-0.214***	-0.217***
	(0.032)	(0.033)	(0.027)	(0.029)	(0.053)	(0.053)
× in Power	0.304***	0.304***	0.159**	0.155**	0.311*	0.316*
	(0.083)	(0.096)	(0.073)	(0.077)	(0.178)	(0.173)
× Discriminated against	-0.059	-0.057	0.130	0.116	0.016	0.012
	(0.102)	(0.120)	(0.130)	(0.151)	(0.137)	(0.152)
Panel B: Manufacturing Exports per GDP per capita						
Trade Exposure	-0.067**	-0.046*	-0.101***	-0.078***	-0.246***	-0.194***
	(0.033)	(0.024)	(0.028)	(0.023)	(0.051)	(0.040)
× in Power	0.334***	0.360***	0.218**	0.193**	0.432**	0.513***
	(0.089)	(0.116)	(0.086)	(0.087)	(0.188)	(0.172)
× Discriminated against	-0.090	-0.074	0.069	0.107	-0.026	0.004
	(0.080)	(0.084)	(0.097)	(0.110)	(0.125)	(0.149)
Country × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country × ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	167,617	167,617	167,617	167,617	167,617	167,617

In this table, I show how individuals belonging to ethnic groups in power benefit from trade, while marginalized ethnic groups loose. *In Power* and *Discriminated* against are taken from the Ethnics Power Relations Core Data 2021. Individuals from the Afrobarometer are assigned their political status by georeferencing their location with the provided shapefiles. *Household wealth*, *Household Trust*, and *Democratic values* are composite scores from variables in the Afrobarometer Rounds 1-7. Geolocated data is used to place individuals in ethnic regions and countries and assign trade exposure. *Trade Exposure* is defined as the predicted trade flows per capita (cf. 3, columns (3) & (4)), aggregated to the exporter level and interacted with the population share of this ethnicity. Country × year fixed effects then account for total trade flows of this exporting country and Country × year fixed effects for the size and impact of each ethnicity in this country. *Trade Exposure* is identified only from the interaction of the two. Individual controls are age, gender, education, and urbanity. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: The effect of Migration on bilateral exports

	Exports per capita					
	(1)	(2)	(3)	(4)	(5)	(6)
	FS	2SLS	FS	2SLS	FS	2SLS
Ethnic connections	0.353***		0.274***		0.263***	
	(0.059)		(0.059)		(0.081)	
	[0.079]		[0.077]		[0.075]	
log Migration in 1960		0.498***		0.671***		0.655*
		(0.170)		(0.241)		(0.351)
		[0.156]		[0.297]		[0.296]
Country-pair controls			Yes	Yes	Yes	Yes
Exporter and importer × year fixed effects					Yes	Yes
Observations	4,195	4,195	4,195	4,195	3,658	3,658
F-test first stage		35.539		21.822		10.624

In this table, I show that ethnic connections predict bilateral migration in 1960 (odd columns) and demonstrate a positive elasticity between migration and trade flows (even columns). *Ethnic connections* are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. *log Migration in 1960* is defined as the total bilateral migration flows in 1960. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. Odd columns denote the first stage: *log Migration in 1960* $= \beta$*Ethnic connections*. The F-test is denoted in even columns. Even columns instrument *log Migration in 1960* and estimate its elasticity with *log Exports per capita*. Country-pair controls defined in Table 2 are added in columns (3)-(6). The effect for all years is shown in Table C.7. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Gains from trade:
Heterogeneity between sectors

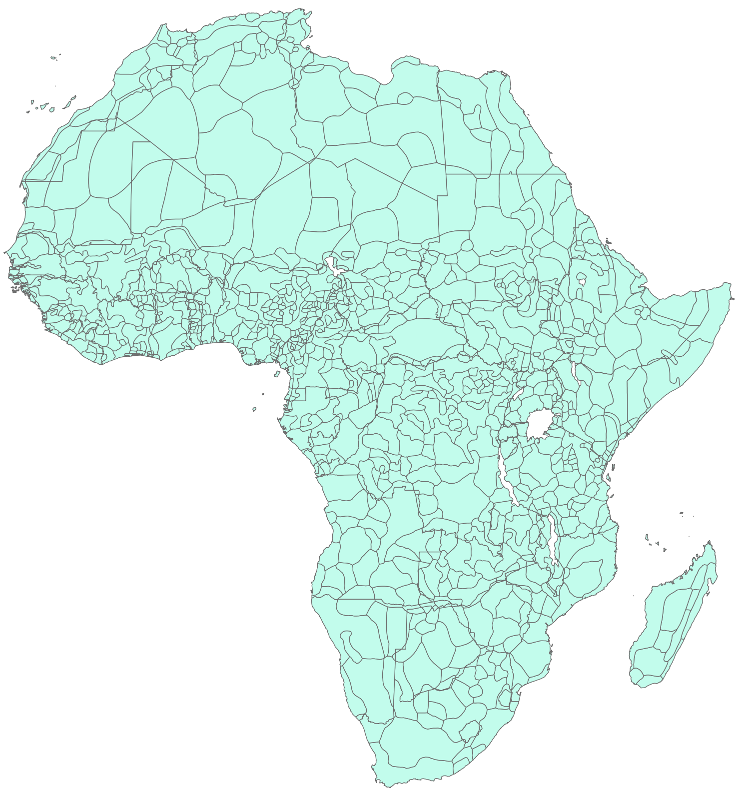
	Agricultural Exports		Manufacturing Exports	
	(1) OLS	(2) PPML	(3) OLS	(4) PPML
Panel A: Fraction pixels lid				
Trade Exposure	-0.002	-0.001	-0.004***	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Mean dependent variable	0.033	0.033	0.033	0.033
Observations	9,918	9,920	10,024	10,026
Panel B: Household welfare				
Trade Exposure	0.002	-0.007	-0.037**	-0.034***
	(0.032)	(0.025)	(0.017)	(0.013)
Observations	219,585	219,585	220,722	220,722
Panel C: Household trust				
Trade Exposure	-0.056***	-0.075***	-0.044**	-0.041**
	(0.019)	(0.021)	(0.022)	(0.017)
Observations	219,585	219,585	220,722	220,722
Panel D: Democratic values				
Trade Exposure	-0.113*	-0.110**	-0.133**	-0.111***
	(0.060)	(0.054)	(0.053)	(0.038)
Observations	219,585	219,585	220,722	220,722

In this table, I assess the importance of exports in different sectors for regional development. *Fraction pixels lid* is defined as the continuous measure from the VIIRMS data. It is calculated as the fraction of pixels not zero in each ethnicity-country-year observation. *Household wealth*, *Household Trust*, and *Democratic values* are composite scores from variables in the Afrobarometer Rounds 1-7. Geolocated data is used to place individuals in ethnic regions and countries and assign trade exposure. *Trade Exposure* is defined as the predicted trade flows per capita (cf. 3, columns (3) & (4)) in the agricultural sector (columns 1-2) and manufacturing sector (3-4), aggregated to the exporter level and interacted with the population share of this ethnicity. Country $\times$ year fixed effects then account for total trade flows of this exporting country and Country $\times$ year fixed effects for the size and impact of each ethnicity in this country. *Trade Exposure* is identified only from the interaction of the two. OLS and PPML denote the estimation method to predict exports. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For Online Publication: Appendix
2nd February 2023

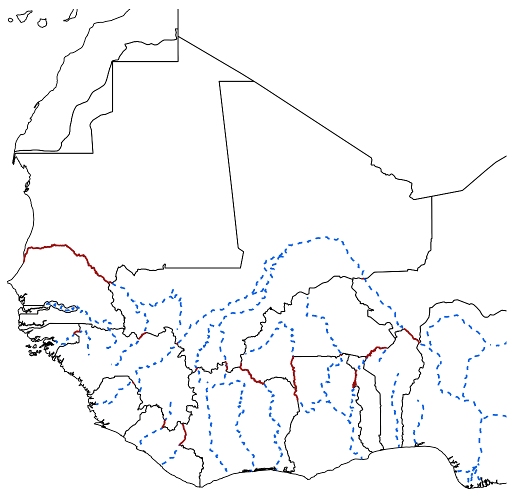
A Supplemental material that support the identification strategy

Figure A.1: The precolonial distribution of ethnic groups in Africa



Notes: Identification assumption i): Variation in the main explanatory variable. This figure shows the distribution of ethnic groups before colonization as recorded by (Murdock, 1959). Every country features at least one split ethnicity, and every border splits at least one ethnicity.

Figure A.2: Rivers as confounders



Notes: Identification assumption ii): Borders were set without sufficient knowledge of local geography. As an example, while rivers are easily observable, most country borders have no rivers (black) and only few contain parts of a river (red).

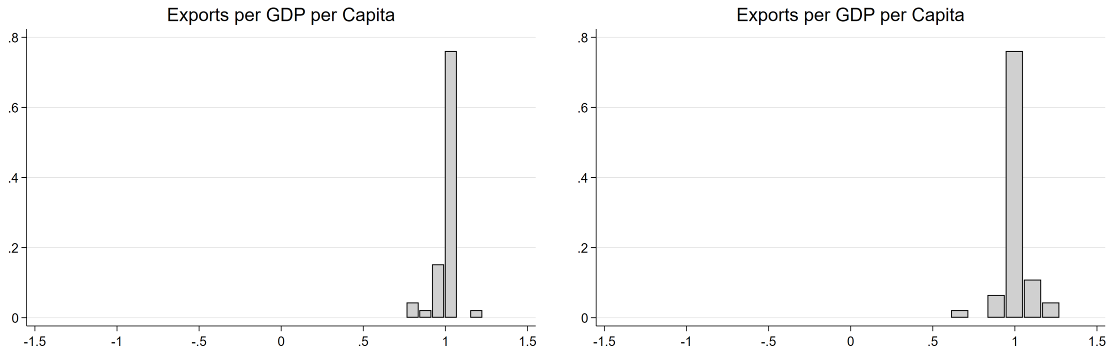
Table A.1: Impact of being divided on ethnic power status as recorded in Wimmer et al. (2009).

	Ethnicity power status					
	(1) OLS	(2) IV	(3) IV	(4) OLS	(5) IV	(6) IV
Tribe split into #countries	-0.062***	-0.257***	-0.276***			
	(0.020)	(0.076)	(0.086)			
Tribe divided				-0.092***	-0.467***	-0.498***
				(0.034)	(0.140)	(0.145)
F-test first stage		80.027	14.980		71.038	17.347
Observations	785	785	421	785	785	421

In this table, I provide suggestive evidence that ethnic groups being split are less likely to be in power. Columns (1) and (4) denote the ordinary least squares controlling for latitude, longitude and their product. In columns (2) and (5), I use log Ethnic Area as in Table 1, column (2) as an instrument. In columns (3) and (6), I use the specification in column (8) Table 1 as instruments. Significance denoted by standard errors clustered by the country: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

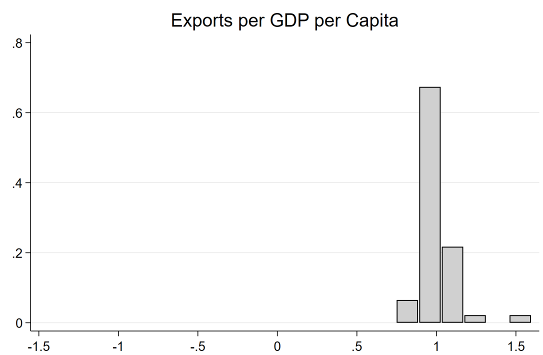
B Sensitivity to dropping outliers

Figure B.1: Sensitivity in the trade prediction



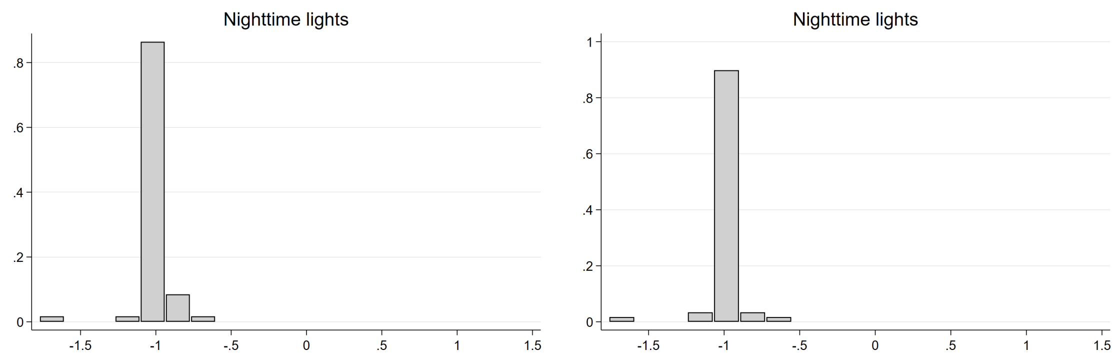
Notes: The left panel uses the specification in Table 2, column (3) and plots the distribution of point estimates when dropping one country at a time relative to the average effect. 1 implies that the point estimate is the same, 0.5 implies its 50% smaller. The right panel uses the specification in Table 2, column (4).

Figure B.2: Sensitivity in the impact of migration on trade



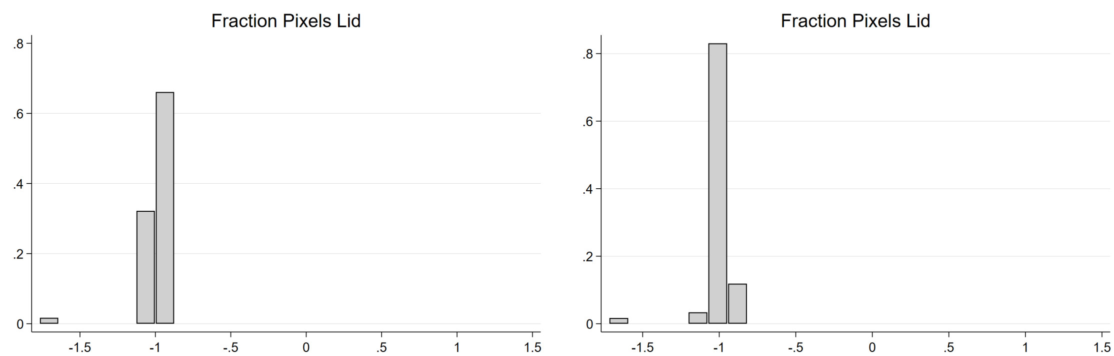
Notes: This panel uses the specification in Table 7, column (6) and plots the distribution of point estimates when dropping one country at a time relative to the average effect. 1 implies that the point estimate is the same, 0.5 implies its 50% smaller.

Figure B.3: Sensitivity in nighttime luminosity specification



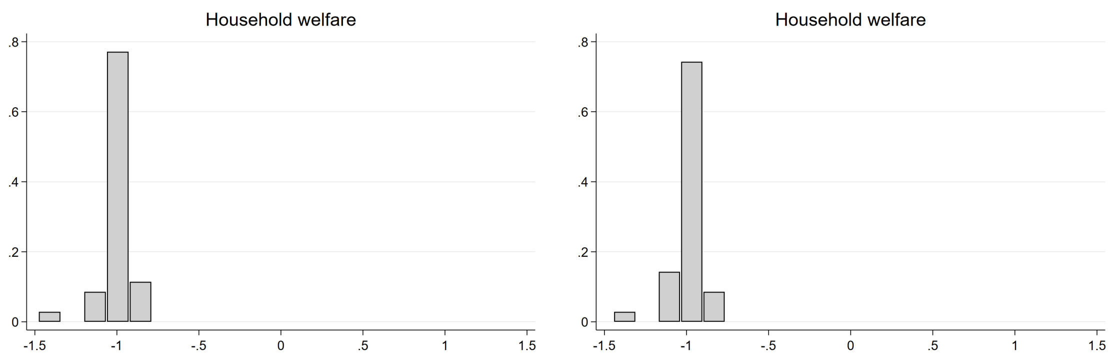
Notes: The left panel uses the specification in Table 4, column (1) and plots the distribution of point estimates when dropping one country at a time relative to the average effect. 1 implies that the point estimate is the same, 0.5 implies its 50% smaller. The right panel uses the specification in Table 4, column (2).

Figure B.4: Sensitivity in the nighttime luminosity specification



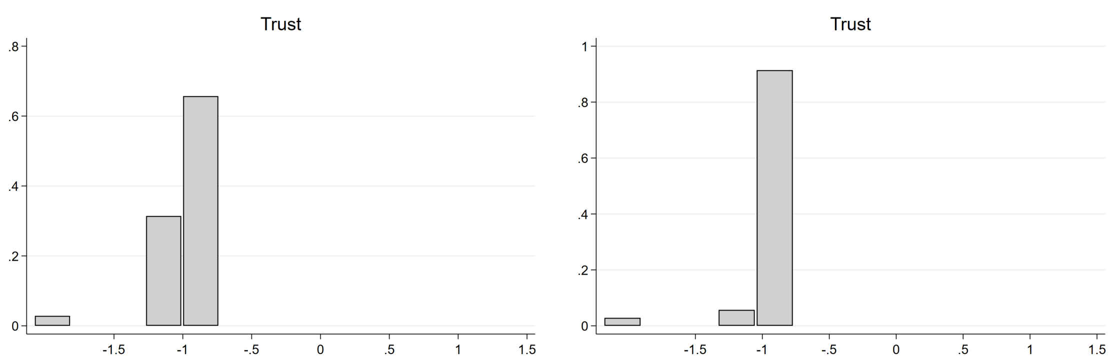
Notes: The left panel uses the specification in Table 4, column (3) and plots the distribution of point estimates when dropping one country at a time relative to the average effect. 1 implies that the point estimate is the same, 0.5 implies its 50% smaller. The right panel uses the specification in Table 4, column (4).

Figure B.5: Sensitivity in the Afrobarometer: Household welfare



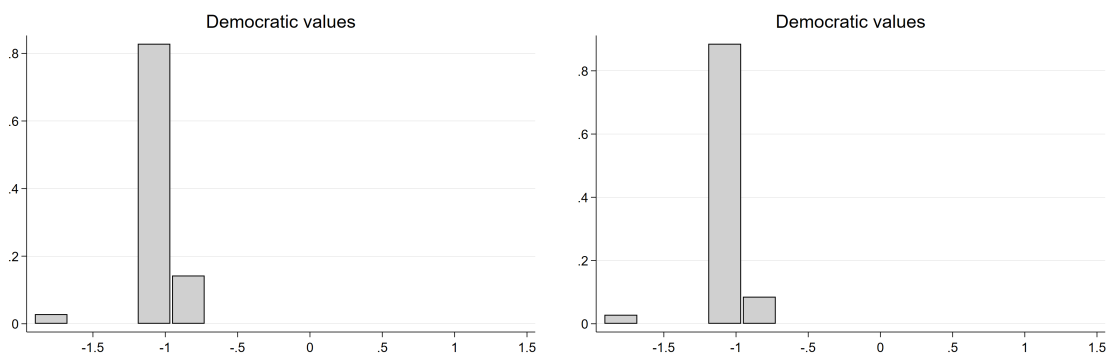
Notes: The left panel uses the specification in Table 5, column (1) and plots the distribution of point estimates when dropping one country at a time relative to the average effect. 1 implies that the point estimate is the same, 0.5 implies its 50% smaller. The right panel uses the specification in Table 4, column (2).

Figure B.6: Sensitivity in the Afrobarometer: Trust



Notes: The left panel uses the specification in Table 5, column (3) and plots the distribution of point estimates when dropping one country at a time relative to the average effect. 1 implies that the point estimate is the same, 0.5 implies its 50% smaller. The right panel uses the specification in Table 4, column (4).

Figure B.7: Sensitivity in the Afrobarometer: Democratic values



Notes: The left panel uses the specification in Table 5, column (5) and plots the distribution of point estimates when dropping one country at a time relative to the average effect. 1 implies that the point estimate is the same, 0.5 implies its 50% smaller. The right panel uses the specification in Table 4, column (6).

C Appendix Tables

Table C.1: Ethnic connections and trade flows

	OLS			PPML
	(1)	(2)	(3)	(4)
Panel A: Exports				
Ethnic connections	0.170***	0.180***	0.164**	0.217***
	(0.050)	(0.045)	(0.071)	(0.065)
	[0.036]	[0.052]	[0.054]	[0.076]
Panel B: Exports per capita				
Ethnic connections	0.176***	0.184***	0.172**	0.170***
	(0.050)	(0.046)	(0.070)	(0.048)
	[0.037]	[0.054]	[0.055]	[0.048]
Panel C: Exports per GDP per capita				
Ethnic connections	0.176***	0.186***	0.178**	0.181***
	(0.052)	(0.047)	(0.071)	(0.061)
	[0.039]	[0.056]	[0.056]	[0.063]
Country-pair controls		Yes	Yes	Yes
Exporter and importer $\times$ year fixed effects			Yes	Yes
Observations	4,128	4,128	4,128	4,131

In this table, I show that ethnic connections predict bilateral exports between countries. *Ethnic connections* are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. The main dependent variable are the logarithm of bilateral exports per current capita in the years 1992–2018. I add the following country-pair controls in columns (2)-(4): log length of the border, log distance between capitals, a binary variable indicating whether parts of the border is a river, a mountain above 1,000 or 2,000 meter. The number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, log border fractionalization (Alesina et al., 2011) and Linguistic and genetic distance $\in [0,1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). In column (3) I add importer$\times$year and exporter$\times$year fixed effects to capture all time-varying variables at the country level. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.2: Ethnic connections and trade flows:
Controlling for regional trade agreements

	OLS			PPML
	(1)	(2)	(3)	(4)
Panel A: Regional Trade Agreements				
Ethnic connections	0.050***	0.054***	0.055***	0.151***
	(0.013)	(0.009)	(0.013)	(0.036)
	[0.015]	[0.011]	[0.011]	[0.049]
Panel B: Exports per capita, controlling for Regional Trade Agreements				
Ethnic connections	0.176***	0.191***	0.184**	0.178***
	(0.050)	(0.052)	(0.079)	(0.052)
	[0.037]	[0.058]	[0.063]	[0.058]
Country-pair controls		Yes	Yes	Yes
Exporter and importer $\times$ year fixed effects			Yes	Yes
Observations	4,195	4,195	4,195	4,198

In this table, I show that while ethnic connections predict regional trade agreements (Panel A), the impact of ethnic connections on bilateral exports between countries is unaffected (Panel B). Regional trade agreements for the years 1989-2020 are obtained from Mario Larch's Regional Trade Agreements Database from Egger and Larch (2008). *Ethnic connections* are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. The main dependent variable are the logarithm of bilateral exports per current capita in the years 1992–2018. I add the following country-pair controls in columns (2)-(4): log length of the border, log distance between capitals, a binary variable indicating whether parts of the border is a river, a mountain above 1,000 or 2,000 meter. The number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, log border fractionalization (Alesina et al., 2011) and Linguistic and genetic distance $\in [0, 1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). In column (3) I add importer$\times$year and exporter$\times$year fixed effects to capture all time-varying variables at the country level. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.3: Ethnic connections and trade flows

	OLS			PPML
	(1)	(2)	(3)	(4)
Panel A: Exports				
Ethnic connections	0.119	0.154*	0.136***	0.156**
	(0.093)	(0.086)	(0.050)	(0.064)
	[0.070]	[0.076]	[0.062]	[0.085]
Panel B: Exports per capita				
Ethnic connections	0.123	0.156*	0.142***	0.157***
	(0.094)	(0.089)	(0.051)	(0.052)
	[0.069]	[0.076]	[0.062]	[0.056]
Panel C: Exports per GDP per capita				
Ethnic connections	0.129	0.164*	0.147***	0.108
	(0.092)	(0.088)	(0.049)	(0.070)
	[0.060]	[0.075]	[0.061]	[0.086]
Country-pair controls		Yes	Yes	Yes
Weighted by observed trade			Yes	Yes
Observations	167	167	4,191	4,191

In this table, I show that ethnic connections significantly and robustly predict bilateral exports between countries. Ethnic connections are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. The main dependent variable are bilateral exports in the years 1992–2018. Panel A studies the impact on the logarithm of bilateral exports. Panel B studies the impact on the logarithm of bilateral exports per 1960s population. Panel C studies the impact on the logarithm of bilateral exports per current GDP per capita. I add the following country-pair controls in columns (2)-(4): log length of the border, log distance between capitals, a binary variable indicating whether parts of the border is a river, a mountain above 1,000 or 2,000 meter. The number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, log border fractionalization (Alesina et al., 2011) and Linguistic and genetic distance $\in [0,1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). In column (3) I add importer$\times$year and exporter$\times$year fixed effects to capture all timevarying variables at the country level. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: The effect of bilateral exports on economic development:
Satellite imagery, alternative export measures

	Nighttime lights		Fraction Pixels Lid	
	(1) OLS	(2) PPML	(3) OLS	(4) PPML
Panel A: Exports				
Trade Exposure	-0.009**	-0.004**	-0.008**	-0.004**
	(0.004)	(0.002)	(0.003)	(0.002)
Panel B: Exports per capita				
Trade Exposure	-0.009**	-0.004**	-0.008**	-0.004**
	(0.004)	(0.002)	(0.003)	(0.002)
Panel C: Exports per GDP per capita				
Trade Exposure	-0.010**	-0.005**	-0.009**	-0.004**
	(0.004)	(0.002)	(0.004)	(0.002)
Country $\times$ year fixed effecst	Yes	Yes	Yes	Yes
Country $\times$ ethnicity fixed effects	Yes	Yes	Yes	Yes
Observations	10,032	10,032	10,032	10,032

In this table, I show how predicted exports impact economic development as measured by nighttime luminosity. *Trade Exposure* is defined as the predicted trade flows (cf. Table 3, columns (3) & (4)), aggregated to the exporter level and interacted with the population share of this ethnicity. Country $\times$ year fixed effects then account for total trade flows of this exporting country and Country $\times$ year fixed effects for the size and impact of each ethnicity in this country. *Trade Exposure* is identified only from the interaction of the two. *Nighttime lights* is defined as the continuous measure from the VIIRMS data. It calculates the average luminosity of each ethnicity-country-year observation. Since 97% of values are smaller than one, a level specification is used. *Fraction pixels lid* is defined from the same source at the same level but calculated as the fraction of pixels not zero. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: All answers, correcting for multiple hypothesis testing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS				PPML			
	beta	s.e.	*p*-value	FDR adj. *p*-value	beta	s.e.	*p*-value	FDR adj. *p*-value
Household wealth								
Country's condition	-0.106	0.085	0.215	0.170	-0.098	0.083	0.234	0.172
Household condition, today	-0.054	0.041	0.187	0.170	-0.053	0.038	0.166	0.158
Household condition, past	-0.098	0.056	0.081	0.128	-0.089	0.054	0.096	0.156
Household condition, compared	-0.082	0.069	0.234	0.170	-0.093	0.071	0.190	0.158
Enough access to food	-0.087	0.044	0.051	0.097	-0.098	0.052	0.060	0.117
Enough access to water	-0.149	0.033	0.000	0.001	-0.137	0.038	0.000	0.003
Enough access to cash	-0.037	0.040	0.356	0.188	-0.014	0.034	0.668	0.325
Enough access to fuel	-0.125	0.046	0.006	0.026	-0.101	0.037	0.006	0.025
Enough access to health	-0.179	0.091	0.050	0.097	-0.139	0.073	0.056	0.117
Trust								
Trust in President	-0.197	0.152	0.196	0.563	-0.153	0.133	0.247	0.418
Trust in Police	0.010	0.032	0.754	0.908	0.029	0.031	0.352	0.507
Trust in Law	-0.075	0.046	0.098	0.487	-0.077	0.044	0.079	0.367
Trust in Army	-0.087	0.033	0.008	0.101	-0.089	0.030	0.004	0.041
Trust in Electoral Commission	-0.141	0.082	0.086	0.487	-0.124	0.077	0.107	0.367
Trust in Parliament	-0.015	0.056	0.793	0.908	-0.001	0.052	0.985	0.830
Trust in local council	-0.047	0.077	0.546	0.756	-0.031	0.071	0.667	0.802
Trust in ruling party	-0.126	0.132	0.339	0.634	-0.106	0.122	0.384	0.507
Trust in opposition party	-0.004	0.046	0.933	1.000	-0.004	0.046	0.927	0.830
Trust in traditional leaders	-0.109	0.075	0.146	0.563	-0.088	0.066	0.181	0.418
Trust in religious leaders	-0.091	0.086	0.292	0.634	-0.114	0.067	0.090	0.367
Democratic values								
How much of a democracy is your country	-0.105	0.051	0.042	0.068	-0.116	0.057	0.044	0.055
Satisfied with democracy	-0.070	0.060	0.244	0.089	-0.077	0.063	0.219	0.079
Freedom of Speech	-0.239	0.106	0.024	0.068	-0.171	0.072	0.017	0.055

Notes: The unit of observation is an individual throughout. All regression are estimated using the individual-level equation (5). All data obtained from Afrobarometer rounds 1-7.

Table C.6: Individual exposure to bilateral trade:
Individual data from the Afrobarometer, using alternative export measures

	log(Exports)		log(Exports, p.c.)		log(Exports, per GDP p.c.)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Household wealth						
Trade Exposure	-0.052***	-0.047***	-0.052***	-0.047***	-0.058**	-0.052**
	(0.020)	(0.017)	(0.020)	(0.017)	(0.024)	(0.021)
Panel B: Household Trust						
Trade Exposure	-0.041**	-0.050**	-0.041**	-0.050**	-0.035*	-0.045*
	(0.020)	(0.024)	(0.020)	(0.024)	(0.018)	(0.024)
Panel C: Democratic Values						
Trade Exposure	-0.132**	-0.117**	-0.132**	-0.117**	-0.132**	-0.113**
	(0.056)	(0.050)	(0.056)	(0.050)	(0.060)	(0.051)
Country $\times$ year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country $\times$ ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	217,809	217,809	217,809	217,809	217,809	217,809

In this table, I show how individuals respond to increased trade exposure. In this table, I assess the gains from trade for each individual. *Household wealth*, *Household Trust*, and *Democratic values* are composite scores from variables in the Afrobarometer Rounds 1-7. Geolocated data is used to place individuals in ethnic regions and countries and assign trade exposure. *Trade Exposure* is defined as the predicted trade flows (cf. 3, columns (3) & (4)), aggregated to the exporter level and interacted with the population share of this ethnicity. Country $\times$ year fixed effects then account for total trade flows of this exporting country and Country $\times$ year fixed effects for the size and impact of each ethnicity in this country. *Trade Exposure* is identified only from the interaction of the two. Individual controls are age, gender, education, and urbanity. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure C.1: DHS: Effect by income group:

Notes: In this panel, I plot the average marginal effects of trade exposure by initial income group. The richest group in every country gains, every other group does not benefit.

Table C.7: The effect of Migration on bilateral exports:
Effects per migration year

	(1) 1960	(2) 1970	(3) 1980	(4) 1990	(5) Pooled
Panel A: 2SLS, Migration flows on exports (1990-2020)					
log Migration flows	0.655*	0.639*	0.719*	0.796*	0.686*
	(0.351)	(0.362)	(0.430)	(0.466)	(0.390)
	[0.296]	[0.321]	[0.371]	[0.417]	[0.339]
F-test first stage	10.624	13.296	9.198	10.456	11.706
Country-pair controls	Yes	Yes	Yes	Yes	Yes
Exporter and importer $\times$ year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	3,658	3,658	3,658	3,658	3,658

In this table, I show that ethnic connections predict bilateral migration and demonstrate a positive elasticity between migration and trade flows across all years. *Ethnic connections* are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. *log Migration* is defined as the total bilateral migration flows as defined by the column title. 1960 refers to migration in 1960, 1970 to migration in 1970, 1980 to migration in 1980, 1990 to migration in 1990. Pooled refers to the sum migration between 1960–1990, ignoring return migration. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. The first stage $log\ Migration = \beta Ethnic\ connections$ is denoted by the F-test . Country-pair controls defined in Table 2 are added in columns (3)-(6). Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Comparing results to the dyadic-level estimation of exports

I estimate the effect of trade flows on ethnic groups by interacting country-level export flows with ethnic level population shares. This '*Trade Exposure*' can be calculated in two —a priori— equally reasonable procedures following the shift-share literature. The key difference lies in the construction of the *shift*; it is either constructed as an ethnicity-level aggregate or a national-level aggregate. In this appendix, I show that both ways lead to the same conclusion.

Prediction of trade flows: In the main part of the paper, I use the following dyadic equation to estimate the impacts of ethnic networks on trade:

$$\log(X_{cd,t}) = \beta \ \log\left(\sum_{e \in E_c \cap E_d}^{E} PS_{c,e} \times PS_{d,e}\right) + \Gamma_{cd} + \delta_c + \delta_d + \delta_t + \varepsilon_{cd,t} \tag{D.1}$$

Which I then break down into the effects at the country-pair by ethnicity level, keeping the variation on the left-hand side constant:

$$\log(X_{cd,t}) = \beta \log\left(PS_{c,e} \times PS_{d,e}\right) + \Gamma_{cd} + \delta_{c,t} + \delta_{d,t} + \delta_{e,c} + \varepsilon_{cde,t} \tag{D.2}$$

Both equations utilize the same data and identification, but exploit the variation at a different level. Hence, the estimated effects in Table D.1 are a magnitude smaller due to the missing aggregation.

The idea behind this exercise is to predict each ethnic groups' contribution to shaping trade flows. Since the value of exports does not vary between ethnic groups of the same country pair, $\hat{\beta}$ will only estimate the average elasticity between export flows and all ethnic groups. When predicting trade flows $\widehat{X_{cde,t}}$ this elasticity $\hat{\beta}$ will be multiplied with each ethnicity's network $(PS_{c,e} \times PS_{d,e})$ to generate variation in each ethnicity's contribution to exports. Larger ethnic groups will have larger export flows and smaller ethnic groups smaller, governed by their network size multiplied with their average elasticity $\hat{\beta}$.

In a final step, both, $\widehat{X_{cde,t}}$ and $\widehat{X_{cd,t}}$ are aggregated. Yet, while $\widehat{X_{cde,t}}$ is aggregated to the country $\times$ ethnicity level $X_{ce,t} = \sum_d \widehat{X_{cde,t}}$, $\widehat{X_{cd,t}}$ is aggregated to the country level $X_{c,t} = \sum_d \widehat{X_{cd,t}}$. The difference is thus, that $X_{ce,t}$ allows for a more precise identification of each ethnic group's contribution to exports than $X_{c,t}$.

Assignment of trade flows: In the main paper, I then interact the aggregate with population shares and control for country $\times$ year and ethnicity $\times$ country fixed effects (substituting in variables for $Trade\ Exposure_{ce,t}$):

$$Y_{cei,t} = \beta X_{ce,t} \times PS_{c,e} + \delta_{c,t} + \delta_{e,c} + X_i + \varepsilon_{ce,t} \tag{D.3}$$

In this appendix, I compare this equation (D.3) to the following equation run at the dyadic level (D.4):

$$Y_{cei,t} = \beta X_{c,t} \times PS_{c,e} + \delta_{c,t} + \delta_{e,c} + X_i + \varepsilon_{ce,t} \tag{D.4}$$

Since the aggregate trade and population shares are dropped due to the fixed effects, only the variation between them is exploited for identification. As the underlying variation is the same, the results in Table D.2-D.3 are very similar across these two methods.

Table D.1: Ethnic connections and trade flows:
Comparing dyadic and ethnicity level estimations

	Exports per capita			
	(1) OLS	(2) OLS	(3) OLS	(4) PPML
Panel A: Dyadic level equation (D.1)*:*				
Ethnic connections	0.176***	0.184***	0.172**	0.170***
	(0.050)	(0.046)	(0.070)	(0.048)
	[0.037]	[0.054]	[0.055]	[0.048]
Country-pair controls		Yes	Yes	Yes
Exporter and importer × year fixed effects			Yes	Yes
Observations	4,195	4,195	4,195	4,198
Panel B: Ethnic level equation (D.2)*:*				
Ethnic connections	0.013***	0.032***	0.055***	0.014***
	(0.004)	(0.008)	(0.015)	(0.005)
	[0.008]	[0.015]	[0.024]	[0.007]
Country-pair controls	Yes	Yes	Yes	Yes
Exporter and importer × year fixed effects	Yes	Yes	Yes	Yes
Ethnic fixed effects		Yes	Yes	Yes
Ethnic × exporter fixed effects			Yes	Yes
N	135,314	135,314	135,314	135,545

In this table, I show that ethnic connections predict bilateral exports between countries. *Ethnic connections* are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. The main dependent variable are the logarithm of bilateral exports per current capita in the years 1992–2018. Panel A estimates the dyadic equation (3), Panel B the ethnicity-level equation (5). I add the following country-pair controls: log length of the border, log distance between capitals, a binary variable indicating whether parts of the border is a river, a mountain above 1,000 or 2,000 meter. The number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, log border fractionalization (Alesina et al., 2011) and Linguistic and genetic distance $\in [0,1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. OLS and PPML denote the estimation method. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.2: The effect of bilateral exports on economic development:
Satellite imagery, comparing dyadic and ethnicity level estimations

	Nighttime lights		Fraction pixels lid	
	(1) OLS	(2) PPML	(3) OLS	(4) PPML
Panel A: Dyadic level equation (D.4)*:*				
Trade Exposure	-0.008**	-0.008**	-0.004**	-0.004**
	(0.004)	(0.003)	(0.002)	(0.002)
Panel B: Ethnic level equation (D.3)*:*				
Trade Exposure	-0.009**	-0.008**	-0.004**	-0.004**
	(0.004)	(0.003)	(0.002)	(0.002)
Country × year fixed effect	Yes	Yes	Yes	Yes
Country × ethnicity fixed effects	Yes	Yes	Yes	Yes
Observations	10,032	10,032	10,032	10,032
Mean dependent variable	0.068	0.068	0.033	0.033

In this table, I show how predicted exports impact economic development as measured by nighttime luminosity. *Trade Exposure* is defined as the predicted trade flows per capita (cf. Table 3, columns (3) & (4)), aggregated to the exporter level and interacted with the population share of this ethnicity. Country × year fixed effects then account for total trade flows of this exporting country and Country × year fixed effects for the size and impact of each ethnicity in this country. Panel A predicts total trade flows using the dyadic equation (3), Panel B the ethnicity-level equation (5). *Trade Exposure* is identified only from the interaction of the two. *Nighttime lights* is defined as the log of the continuous measure from the VIIRMS data plus one. It calculates the average luminosity of each ethnicity-country-year observation. *Fraction pixels lid* is defined from the same source at the same level but calculated as the fraction of pixels not zero. OLS and PPML denote the estimation method to predict exports. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.3: The effect of bilateral exports on economic development:
Individual data from the Afrobarometer, comparing dyadic and ethnicity level estimations

	Household wealth		Household trust		Democratic values	
	(1) OLS	(2) PPML	(3) OLS	(4) PPML	(5) OLS	(6) PPML
Panel A: Dyadic level equation (D.4)*:*						
Trade Exposure	-0.050***	-0.047***	-0.038*	-0.050**	-0.128**	-0.117**
	(0.019)	(0.017)	(0.019)	(0.024)	(0.056)	(0.050)
Panel B: Ethnic level equation (D.3)*:*						
Trade Exposure	-0.052***	-0.047***	-0.041**	-0.050**	-0.132**	-0.117**
	(0.020)	(0.017)	(0.020)	(0.024)	(0.056)	(0.050)
Country × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country × ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	217,809	217,809	217,809	217,809	217,809	217,809

In this table, I assess the gains from trade for each individual. *Household wealth*, *Household Trust*, and *Democratic values* are composite scores from variables in the Afrobarometer Rounds 1-7. Geolocated data is used to place individuals in ethnic regions and countries and assign trade exposure. *Trade Exposure* is defined as the predicted trade flows per capita (cf. 3, columns (3) & (4)), aggregated to the exporter level and interacted with the population share of this ethnicity. Panel A predicts total trade flows using the dyadic equation (3), Panel B the ethnicity-level equation (5). Country × year fixed effects then account for total trade flows of this exporting country and Country × year fixed effects for the size and impact of each ethnicity in this country. *Trade Exposure* is identified only from the interaction of the two. Individual controls are age, gender, education, and urbanity. OLS and PPML denote the estimation method to predict exports. Significance denoted by standard errors clustered by country and ethnicity: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E 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 model while remaining tractable.

The economy consists of N countries which contain a subset $e \in E$ of predefined ethnic groups. 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:

$$\pi_{ij,ee'}(\omega) = p_{ij}(\omega)q_{ij}(\omega) - c_{ij,ee'}(\omega) \tag{E.5}$$

Where the price of a good $p_{ij}(\omega)$ is country specific, as is the demand for a good $q_{ij}(\omega)$.[26] $\tau_{ij} > 1$ 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:

$$c_{ij,ee'}(\omega) = \frac{\tau_{ij}}{\varphi} q_{ij}(\omega) + \left(\frac{L_{j,e'}}{L_j}\right)^{-\eta} f_{ij} \tag{E.6}$$

Here, φ denotes productivity which every firm draws from a Pareto distribution $G(\varphi) = 1 - \varphi^{-\gamma}$.[27] 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.[28] 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.[29] I call the effect of

[26] Although Aker et al. (2014) show that ethnic groups 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 ethnic groups which would yield a result similar to including different sectors.

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

[28] 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.

[29] 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

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 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:

$$U = q_0^{1-\mu} \left(\int_{\omega \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}\mu} \tag{E.7}$$

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:

$$q_{ij}(\varphi) = p_{ij}(\varphi)^{-\sigma} P_j^{\sigma-1} \mu \left(1 + \frac{\Pi}{L}\right) L_j. \tag{E.8}$$

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.[30] 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 $\pi_{ij,ee'} \geq 0$:

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

The price index P_j can be solved explicitly by summing all prices from all exporting countries together, taking their productivity cutoffs into account.[31] Then, the productivity cutoff can be expressed in terms of primitives:

to pay when dealing with non co-ethnic members. The results are robust.

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

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

$$\varphi^*_{ij,ee'} = \left[\frac{\gamma}{\gamma-(\sigma-1)}\right]^{\frac{1}{\gamma}} \left[\frac{\mu}{\sigma}\left(1+\frac{\Pi}{L}\right)\right]^{-\frac{1}{\gamma}} L_j^{\frac{\eta-1}{\gamma}} \frac{\tau_{ij}}{\theta_j} f_{ij}^{\frac{1}{\sigma-1}} (L_{j,e'})^{\frac{\eta}{1-\sigma}} \tag{E.10}$$

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.[32] θ denotes the multilateral resistance term that approximates how distant a market is in comparison to all other markets.[33] Equation (E.10) 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.[34]

In order to obtain a testable equation, I aggregate individual demand[35] to an network extended gravity equation:

$$X_{ij} = \mu\left(1+\frac{\Pi}{L}\right) L_j f_{ij}^{\frac{\sigma-1-\gamma}{\sigma-1}} \left(\frac{\tau_{ij}}{\theta_j}\right)^{-\gamma} \sum_{e\in E_i\cap E_j} L_{i,e}(L_{j,e'})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}} \tag{E.11}$$

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.[36] If the number of ethnic groups 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

[32] 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.

[33] $\theta_j = \left[\sum_{k=1}^{N} f_{kj}^{\frac{\sigma-1-\gamma}{\sigma-1}} \tau_{kj}^{-\gamma} \sum_{e\in E} L_{k,e}(L_{j,e})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}}\right]^{-\frac{1}{\gamma}}$. 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.

[34] 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.

[35] $X_{ij} = L_i \sum_{e'=e\in E} \frac{L_{i,e}}{L_i} \int_{\varphi^*_{ij,ee'}}^{\infty} dG(\varphi)$, 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).

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

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 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:

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

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:

$$\log(X_{ij,t}) = \beta \log \left(\sum_{e \in i \cap j}^{E} \frac{L_{i,e}}{L_i} \times \frac{(L_{j,e'})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}}}{L_j} \right) + \Gamma_{ij,t} + \delta_i + \delta_j + \varepsilon_{ij,t} \tag{E.13}$$

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.

Table E.1: Ethnic connections and trade flows:
Varying the exponent on the the foreign ethnic population.

	OLS	PPML	OLS	PPML	OLS	PPML
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Exports						
Model Network	0.172**	0.197***				
	(0.070)	(0.062)				
	[0.055]	[0.073]				
Model Network, exponent: 0.5			0.211**	0.173**		
			(0.088)	(0.074)		
			[0.072]	[0.087]		
Model Network, exponent: 0.2					0.220**	0.133*
					(0.097)	(0.081)
					[0.083]	[0.101]
Panel B: Exports per capita						
Model Network	0.172**	0.170***				
	(0.070)	(0.048)				
	[0.055]	[0.048]				
Model Network, exponent: 0.5			0.211**	0.108		
			(0.088)	(0.078)		
			[0.072]	[0.087]		
Model Network, exponent: 0.2					0.220**	0.037
					(0.097)	(0.103)
					[0.083]	[0.122]
Panel C: Exports per GDP per capita						
Model Network	0.172**	0.170***				
	(0.070)	(0.048)				
	[0.055]	[0.048]				
Model Network, exponent: 0.5			0.211**	0.108		
			(0.088)	(0.078)		
			[0.072]	[0.087]		
Model Network, exponent: 0.2					0.220**	0.037
					(0.097)	(0.103)
					[0.083]	[0.122]
Country-pair controls	Yes	Yes	Yes	Yes	Yes	Yes
Exporter and importer × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,128	4,131	4,128	4,131	4,128	4,131

In this table, I show that the relationship between ethnic networks and trade is robust, even when allowing for more realistic assumptions on the foreign ethnic networks importance $\frac{\eta(\sigma-1-\gamma)}{1-\sigma}$. Ethnic connections are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. The main dependent variable are bilateral exports in the years 1992–2018. Panel A studies the impact on the logarithm of bilateral exports. Panel B studies the impact on the logarithm of bilateral exports per 1960s population. Panel C studies the impact on the logarithm of bilateral exports per current GDP per capita. I add the following country-pair controls in columns (2)-(4): log length of the border, log distance between capitals, a binary variable indicating whether parts of the border is a river, a mountain above 1,000 or 2,000 meter. The number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, log border fractionalization (Alesina et al., 2011) and Linguistic and genetic distance $\in [0, 1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). In column (3) I add importer×year and exporter×year fixed effects to capture all timevarying variables at the country level. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E.1 Inter-ethnic Trade

So far I assumed that connections can only exist within ethnic groups 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 ethnic groups 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 ethnic groups. Therefore, I assume there exists a matrix $F_{E\times E}$ that reflects this ordering between every possible combination of ethnic groups. The cost of producing and exporting are then given by:

$$c_{ij,ee'}(\varphi) = \frac{\tau_{ij}}{\varphi} q_{ij}(\varphi) + \left(\frac{L_{j,e'}}{L_j}\right)^{-\eta} f_{ij} f_{ij,ee'} \tag{E.14}$$

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.[37] The gravity equation is then given by:

$$X_{ij} = L_j \mu \left(1 + \frac{\Pi}{L}\right) f_{ij}^{1-\frac{\gamma}{\sigma-1}} \left(\frac{\tau_{ij}}{\theta_j}\right)^{-\gamma} \sum_{e\in E\cap E'} L_{i,e} (L_{j,e'})^{\frac{\eta(\sigma-1-\gamma)}{1-\sigma}} f_{ij,ee'}^{1-\frac{\gamma}{\sigma-1}} \tag{E.15}$$

Now, the effect of ethnic match probabilities is not only measured within ethnic groups, but also between ethnic groups. If the fixed costs of creating ties between ethnic groups are low enough, this specification should fit the data better. Combining the findings on the extensive margin formulation and the ethnic specific fixed costs, ethnic groups 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 ethnic groups.

[37] 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.

Table E.2: Ethnic connections and trade flows:
Allowing for inter-ethnic trade

	OLS		PPML	
	(1)	(2)	(3)	(4)
Panel A: Exports between bordering countries				
Ethnic connections	0.172^{**}		0.197^{***}	
	(0.070)		(0.062)	
	[0.055]		[0.073]	
log Distance weighted match probability		1.549^{**}		0.966^{**}
		(0.724)		(0.376)
		[0.533]		[0.518]
Effect size	0.564	2.437	0.644	1.520
Country-pair controls	Yes	Yes	Yes	Yes
Exporter and importer $\times$ year fixed effects	Yes	Yes	Yes	Yes
Observations	4,195	4,195	4,198	4,198

In this table, I show that ethnic connections significantly and robustly predict bilateral exports between countries, even when allowing for inter ethnic trade Ethnic connections are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. The main dependent variable are bilateral exports in the years 1992–2018. Panel A studies the impact on the logarithm of bilateral exports between bordering countries. I add the following country-pair controls in columns (2)-(4): log length of the border, log distance between capitals, a binary variable indicating whether parts of the border is a river, a mountain above 1,000 or 2,000 meter. The number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, log border fractionalization (Alesina et al., 2011) and Linguistic and genetic distance $\in [0,1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). In column (3) I add importer$\times$year and exporter$\times$year fixed effects to capture all timevarying variables at the country level. Effect size scales the point estimate by a standard deviation change in either Ethnic connections (Columns 1 and 3) or the Distance weighted Match probability (Columns 2 and 4). Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.3: Ethnic connections and trade flows:
Allowing for inter-ethnic trade between all countries

	Reported Trade		All Trade	
	(1) OLS	(2) PPML	(3) OLS	(4) PPML
Panel A: Exports between all countries				
log Distance weighted match probability	1.754***	1.584***	1.494***	1.612***
	(0.129)	(0.171)	(0.175)	(0.173)
	[0.191]	[0.159]	[0.251]	[0.158]
Effect size				
Country-pair controls	Yes	Yes	Yes	Yes
Exporter and importer × year fixed effects	Yes	Yes	Yes	Yes
Observations	38,271	38,271	66,240	54,754

In this table, I show that ethnic connections significantly and robustly predict bilateral exports between countries, even when allowing for inter ethnic trade. Ethnic connections are defined as the log ethnic match probability as defined in equation (2) and capture the likelihood of drawing two individuals from either country with the same ethnicity. Sample consist of bordering countries in Africa and includes importer, exporter, and year fixed effects in all regressions. Inference is made on the basis of standard errors clustered by each country-pair, as shown in parenthesis. Two-way clustered standard errors allowing for separate home- and foreign-clusters shown as robustness in brackets. The main dependent variable are bilateral exports in the years 1992–2018. Columns (1) and (2) use all reported trade, columns (3) and (4) all possible trade, setting missing trade as zero. Panel A studies the impact on the logarithm of bilateral exports between all countries. I add the following country-pair controls in columns (2)-(4): log distance between capitals, whether the countries share a border, the number of ethnic groups shared between the countries, whether the countries share a colonial past or judicial language, and Linguistic and genetic distance $\in [0, 1]$ to capture the similarity between the countries (Spolaore and Wacziarg, 2015). In column (3) I add importer×year and exporter×year fixed effects to capture all timevarying variables at the country level. Significance denoted by standard errors clustered by the country pair: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$