DISCUSSION PAPER SERIES

DP12400
(v. 2)

ETHNIC DIVERSITY AND GROWTH: REVISITING THE EVIDENCE

Marta Reynal-Querol and Jose G Montalvo

DEVELOPMENT ECONOMICS

ISSN 0265-8003

ETHNIC DIVERSITY AND GROWTH: REVISITING THE EVIDENCE

Marta Reynal-Querol and Jose G Montalvo

Discussion Paper DP12400
First Published 28 October 2017
This Revision 06 July 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

ETHNIC DIVERSITY AND GROWTH: REVISITING THE EVIDENCE

Abstract

The relationship between ethnic heterogeneity and economic growth is complex. Empirical research working with cross-country data finds a negative, or statistically insignificant, relationship. However, analysis at city level finds a positive effect of diversity on wages and productivity. Generally, there is a trade-off between the economic benefits of diversity and the costs of heterogeneity. Using cells of fixed size we find that the relationship between diversity and growth is positive for small geographical areas. In the case of Africa, we argue that the explanation is the increase in trade at the boundaries between ethnic groups due to ethnic specialization.

JEL Classification: N/A

Keywords: Scale, Ethnicity, economic growth

Marta Reynal-Querol - marta.reynal@upf.edu
Universitat Pompeu Fabra and CEPR

Jose G Montalvo - jose.garcia-montalvo@upf.edu
Universitat Pompeu Fabra

Acknowledgements
We would like to thank Daron Acemoglu, Alberto Alesina, Tim Besley, Klaus Desmet, Ruben Durante, Oded Galor, Saumitra Jha, Massimo Morelli, Gerard Padro, Elias Papaioannou, Luigi Pascali, Felipe Valencia, Hans-Joachim Voth, and especially, Antonio Ciccone, Stelios Michalopoulos and Roman Wacziarg for their helpful comments. We also thank the participants in the Bocconi Workshop on Conflict, the Meeting of the European Public Choice Society, the EPCS of Groningen, the Berlin Workshop on Advances on the Political Economy of Conflict and Redistribution III, the Invited Session on Political Economy of Development at the EEA Meetings of 2015 in Mannheim, the Yale University Workshop on The Political Economy of Social Conflict, and the NBER Summer Institute. We have also bene ted from the comments of participants in seminars at Stanford GSB, Oxford, LSE, Warwick, Bonn, Stockholm, UPF and CREI. Financial support from the European Research Council (ERC), the Spanish National Science Foundation, and the Government of Catalonia is gratefully acknowledged. The authors also acknowledge the financial support from the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2015-0563).

Ethnic Diversity and Growth: Revisiting the Evidence*

José G. Montalvo†

Marta Reynal-Querol‡

*We would like to thank Daron Acemoglu, Alberto Alesina, Tim Besley, Klaus Desmet, Ruben Durante, Oded Galor, Saumitra Jha, Massimo Morelli, Gerard Padro, Elias Papaioannou, Luigi Pascali, Felipe Valencia, Hans-Joachim Voth, and especially, Antonio Ciccone, Stelios Michalopoulos and Roman Wacziarg for their helpful comments. We also thank the participants in the Bocconi Workshop on Conflict, the Meeting of the European Public Choice Society, the EPCS of Groningen, the Berlin Workshop on Advances on the Political Economy of Conflict and Redistribution III, the Invited Session on Political Economy of Development at the EEA Meetings of 2015 in Mannheim, the Yale University Workshop on The Political Economy of Social Conflict, and the NBER Summer Institute. We have also benefited from the comments of participants in seminars at Stanford GSB, Oxford, LSE, Warwick, Bonn, Stockholm, UPF and CREI. Financial support from the European Research Council (ERC), the Spanish National Science Foundation, the Barcelona GSE Research Network, and the Government of Catalonia is gratefully acknowledged.

†Corresponding author: José García Montalvo, Professor of Economics, Univ. Pompeu Fabra (UPF), BarcelonaGSE and IPEG, Ramón Tras Fargas 25-27, Barcelona 08005. Telephone: (+34) 93 542 2509. E-mail jose.garcia-montalvo@upf.edu

‡Marta Reynal-Querol, ICREA Research Professor, Univ. Pompeu Fabra (UPF), IPEG and BarcelonaGSE Ramón Tras Fargas 25-27, Barcelona 08005, Spain. E-mail: marta.reynal@upf.edu

Abstract

The relationship between ethnic heterogeneity and economic growth is complex. Empirical research working with cross-country data finds a negative, or statistically insignificant, relationship. However, analysis at city level finds a positive effect of diversity on wages and productivity. Generally, there is a trade-off between the economic benefits of diversity and the costs of heterogeneity. Using cells of fixed size we find that the relationship between diversity and growth is positive for small geographical areas. In the case of Africa, we argue that the explanation is the increase in trade at the boundaries between ethnic groups due to ethnic specialization.

Keywords: ethnic diversity, economic growth, spatial analysis

JEL Codes: O10, O40, N17, R12

1 Introduction

The issue of the effect of ethnic diversity on economic development has generated a large body of literature. Initially the literature analyzed the issue by running cross-country regressions. For example, using cross-country differences in ethnic diversity, Easterly and Levine (1997) show that Africa's low level of economic development is associated with its high degree of ethnic heterogeneity. Alesina et al. (2003) and Alesina and La Ferrara (2005), also using cross- country data, similarly show a consistent negative effect of ethnic fractionalization on growth. Further research has qualified the conditions for this negative relationship.[1] By contrast, research based on data from small geographical areas, such as cities, frequently finds a positive effect of diversity on wages and productivity.[2]

Is ethnic diversity good or bad for economic growth? The literature has often emphasized the trade-off between the benefits of diversity and the costs of heterogeneity. On the one hand, ethnic diversity can be beneficial by enhancing productivity through innovation, skill complementarities, increased creativity, trade, and product variety. On the other hand, diversity can generate an inefficient provision of public goods, ethnically biased policies, and conflict or disagreement over common public goods and policies. All of these theories generally imply that there is a size at which benefits and costs are equalized, implying that on a smaller scale we should find a positive effect of ethnic diversity and on a larger scale we should find a negative effect. However, theoretical models on the effects of ethnic diversity on economic development are mostly agnostic about the scale of analysis. The literature has thus generally found that diversity seems to be negative, or irrelevant, for development at high levels of aggregation, but positive at low levels of

[1] Collier (2001), Easterly (2001), Alesina, Spolaore, and Wacziarg (2000), Montalvo and Reynal-Querol (2005), or Gören (2014).

[2] Ottaviano and Peri (2005), Ottaviano and Peri (2006), Alesina, Harnoss, and Rapoport (2016) or Sparber (2010) study the case of the US. Nathan (2011) analyses the case of the UK, while Suedekum, Wolf and Blien (2014) discuss the results of Germany, and Bakens, Mulder and Nijkamp (2013) the example of the Netherlands.

geographical aggregation, such as the city level. Detailed geographical data is, in fact, becoming increasingly popular in the analysis of ethnicity.[3]

In this paper, we argue that the answer to the question of the nature of the relationship between ethnic heterogeneity and growth is, in fact, different depending on the size of the unit of analysis. From a theoretical perspective, if we assume that ethnic specialization increases the variety of goods in the presence of ethnic diversity, it is simple to write down a growth model that depicts a positive relationship between growth and ethnic diversity. In fact, the growth rate of most product-variety models increases with the size of the economy, as measured by total labor supply[4], which is positively correlated with diversity. However, as we increase the size of the unit and, therefore, the degree of diversity, the cost of heterogeneity increases. There is an optimal level of diversity determined by the trade-off between the benefits and the costs of ethnic heterogeneity. The larger the benefits from skill variety in production, the larger the size, while the higher the effect of heterogeneity on the unwillingness to share public goods, the smaller the size. Obviously the specific mechanism that explains the benefits and costs when ethnic heterogeneity increases may be different depending on the level of development and the sectoral structure of the economy.

One mechanism, which we explore in this paper, is based on the increase in trade due to the production specialization or service complementarities of different ethnic groups. The issue of the impact of spatial ethnic heterogeneity on intra-national trade is an underdeveloped topic of research. Aker et al. (2014) argue, using information from Niger, that transaction costs are higher for trade between regions with different ethnicities versus trade in homogeneous areas. It is well known that trust is higher among people of the same ethnic group, and that this consequently reduces transaction costs. However,

[3]Michalopoulos and Papaioannou (2013, 2014), Alesina, Michalopoulos, and Papaioannou (2016) or Michalopoulos (2012). Several papers have recently used the boundaries between ethnic groups to generate quasi-experiments like Michalopoulos and Papaioannou (2014, 2016).

[4]Aghion and Howitt (2009)

ethnic specialization in production generates a motive for trade that is absent within homogeneous groups. There are, in fact, many mitigating mechanisms that can reduce the cost of transactions with other ethnic groups. Several papers have analyzed mechanisms that can support trade among agents that belong to different groups like, for instance, Glaeser (2005), Greif (1993, 2000, 2006) or Jha (2013).

At a very high level of resolution of the grid there is obviously no possibility of finding measures of output, value added, or even wages for most countries. We consequently take advantage of luminosity data to proxy for local economic activity. Recent research has shown that light density at night is a good proxy of economic activity.[5] We find that, at the highest degree of resolution, there is a positive association between ethnic heterogeneity and economic growth. Finding this correlation at the country level does not, however, resolve the issue of endogeneity caused by the possibility that other unobserved characteristics can drive the association via, for instance, institutional differences. Using artificially constructed cells mitigates this concern. Taking advantage of arbitrarily drawn borders for arbitrary levels of aggregation allows for the control using fixed effects, and mitigates the concern of endogeneity of the contemporaneous boundaries of countries.

In addition, we show that the results are robust to a large number of potential issues. First, we show that the results are unaffected by the use of a large number of controls to account for within-country variation, such as geography, climate, soil quality, proximity to lakes or political capitals, etc. Second, we run one hundred different regressions randomly changing the initial location of the point that defines, together with the level of resolution, the exact location of the area covered by each cell. The results are robust to the location of the origin of the grid that defines the cells. We also show that the positive effect of heterogeneity on economic growth is not only due to the contribution of urban areas, nor is it simply capturing an agglomeration effect. Finally, we find that reducing the degree of resolution of the grid decreases the association between ethnic diversity and development, to the point of finding no association between heterogeneity and development.[6]

[5] See, for example, Chen and Northaus,(2011) or Pinkowskiy and Sala-i-Martin, (2016)

[6] Montalvo and Reynal-Querol (2017a)

These findings are likely to derive from different mechanisms, depending on the sectoral structure of the economy or its level of development. In particular, we consider the case of Africa. In order to understand why regional development is faster along ethnic borders, we propose a mechanism related to trade. We find that those areas which have more ethnic diversity, also have a higher proportion of markets. Previous research (Michalopoulos, 2012) has shown that ethnic groups in Africa typically specialize in different agricultural products, and therefore have incentives to exchange goods. Ethnic groups that are geographically close to one another, and can potentially monitor each other, may therefore have a large volume of trade despite their different ethnic origins. In order to provide evidence of this mechanism, we show that local markets in Africa are located close to ethnic borders, supporting this interpretation.

The structure of the paper is as follows. Section 2 describes the data. Section 3 presents the basic results and discusses some exercises that show the robustness of the results. Section 4 proposes a mechanism for the case of Africa, and presents supporting evidence. Section 5 concludes.

2 Data

The previous section argues that the effect of ethnic heterogeneity on growth depends on the balance between the positive effects of diversity (skill complementarities, propensity to innovate, specialization and trading, etc.) and its the negative effects (problems in the provision of public goods, potential conflict, etc.). The empirical literature generally finds that the negative effect of diversity prevails over the positive effect at the country level. From a theoretical perspective, it is clear that for small areas the positive effect of diversity should generally dominate the negative effect, while the opposite should be true for large geographical areas. In this paper we therefore investigate the effect of ethnic diversity on development using small geographical units to investigate the sign of that relationship.

Our units of observation are grid-country cells that generate country level data at very high resolution. We construct grids, and calculate the value of the explanatory variables and the outcome for each of these cells. We also check the robustness of the results by changing the origin of the grid that generates the country cells. Table 1 presents the descriptive statistics of the variables used in the empirical analysis. The basic variables for the specification are measures of local growth and ethnic diversity. We also describe the control variables included in the regressions.

Table 1: **Summary statistics**

	Observations	Mean	Std. deviation	Q50	Q90
Growth	21514	1.509	2.800	0.623	6.266
Log night light 1992 pc	21514	-5.312	4.159	-4.332	-0.645
Ethnic Fractionalization	21514	0.113	0.191	0.000	0.467
Distance to Coastline (km)	21514	144.219	181.713	72.026	387.069
Terrain Ruggedness Index, 100 m	21514	9869.915	3338.446	10704.371	11778.844
Coastline Border	21514	0.450	0.498	0.000	1.000
Average temperature from 1961-1980	21514	13.380	11.836	15.190	26.867
Average precipitation from 1961-1980	21514	79.630	73.480	53.872	187.556
Log. Population Density	21514	1.204	2.871	1.507	4.665
Share Mining	21514	0.000	0.011	0.000	0.000
% Fertile Soil	21514	27.683	36.213	2.778	96.528
Distance to River (km)	21514	13.652	14.940	13.334	29.817
Distance to Lake (km)	21514	8.429	54.004	2.778	12.423
Distance to Equatorial Line	21514	24.844	28.699	28.000	60.161
Log. OBS area	21514	7.324	3.585	9.351	9.968
Border (yes=1)	21514	0.271	0.444	0.000	1.000
Ecological Fractionalization	21514	0.220	0.292	0.020	0.610
Pathogen Stress Index	21514	11.48	3.63	11	16

Own elaboration from several sources

2.1 Local growth

To measure growth in each cell we need information on economic development. At high levels of resolution it is difficult to find estimations of GDP and, certainly, many areas of the world do not have information on geocoded high-resolution measures of economic development. It has, however, become increasingly common to use satellite night light density as a proxy for local economic activity when working with small geographical areas. Satellite night light density data are available from the National Oceanic and Atmospheric Administration and have been used recently by scholars such as Henderson, Storeygard, and Weil (2012), Michalopoulos (2012), Michalopoulos and Papaioannou (2013, 2014), and Alesina, Michalopoulos, and Papaioannou (2016). There is also a series of papers that specifically corroborate a high within-country correlation between GDP and light density at night. Chen and Nordhaus (2011) find that luminosity has informational value for countries, regions, and areas with poor quality or missing data. They also argue that night light has a large estimated optimal weight in the estimation of growth rates in countries with low quality statistical systems, following the A to D classification of the Penn World Tables (PWT). In particular, the authors show that the weight is, in these cases, larger than in the estimation of the level of GDP per capita. The importance of night light, as measured by its weight, in the estimation of growth is always higher in low-GDP density countries than those of high-GDP density, for any level of quality of the statistical system.[7] More recently, Pinkovskiy and Sala-i-Martin (2016) have used nighttime lights to show that national accounts are an excellent proxy for actual income, while survey means have very little, if any, informative content to estimate true income. They show that growth rates of GDP per capita are very highly correlated with the growth of night light per capita while the growth rate of survey means is very weakly correlated with the growth of night light per capita. Along similar lines, Jean et al. (2016) use satellite images and machine learning techniques to predict poverty at small scales.

[7]The cross-validation analysis in Michalopoulos and Papaioannou (2013) shows that light density at night is highly correlated with a wealth index across households in four large African countries.

In their application, they use daytime satellite photos to capture details of the landscape (metal roof, water, etc.) that they correlate, using neural networks, with satellite night lights as a proxy for economic activity[8].

All data were obtained from the National Geophysical Data Center, specifically the Earth Observation Group (EOG) reference to version 4 of the Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS) Night-time Lights Time Series. The data is divided by year, from 1992 to 2013 and by six satellites from F10 to F18. From the three available image types, we use the stable light version, which is quantitized into 63 levels of light intensity. We have information on the Night Time Light and the total Night Time Light density by pixel from 1992 to 2010. Population data comes from the Gridded Population of the World. For each cell, we constructed measures of luminosity per capita. Our basic dependent variable is the per capita growth between 1992 and 2010.

2.2 Spatial Ethnic Diversity

We use data from GREG (Geo-referencing of ethnic groups) for the geospatial location of ethnic groups (Weidman, Rod, and Cederman, 2010). Relying on maps and data drawn from the classical Soviet Atlas Narodov Mira (AnM), the GREG dataset employs geographic information systems (GIS) to represent group territories as polygons. The full GREG dataset has global coverage and consists of 929 groups.[9]

---

[8]Night light intensity has also being used to measure inequality at low levels of geographical aggregation. See, for example, Alesina Michalopoulos, and Papaioannou (2016) and Montalvo and Reynal-Querol (2017b)

[9]Desmet, Ortuo-Ortin, and Wacziarg (2012) use a linguistic tree to calculate measures of diversity at different levels of aggregation. They argue that, while deep cleavages are relevant for conflict, more superficial cleavages are relevant for economic growth. We tend to agree with this conclusion, although given the computational challenges of the exercises in this paper, an analysis of that incorporates the degree of ethnic cleavages must necessarily be left for a future project.

For each country cell, we construct two types of measures of diversity. For the first measure, we use the percentage of territory that the homeland of the ethnic group covers in a particular cell. The second measure uses the percentage of the population living in the homeland of the ethnic group in a particular cell. We use the traditional fractionalization measure (Herfindhal index). Since data on population living in the specific homeland of a cell–country unit can only be computed from 1990 on, we use this second measure as a robustness check.

To capture ethnic diversity we also use the Ethnolingustic Fractionalization Index (ELF). In particular the index takes the form,

$$FRAC\text{=}1\text{-}\sum_{i=1}^{N} \pi_i^2 = \sum_{i=1}^{N} \pi_i(1 - \pi_i) \tag{1}$$

where π is the proportion of people who belong to ethnic group i. The broad popularity of the ELF index is based on its intuitive appeal: the index captures the probability that two randomly selected individuals from a given area will not belong to the same ethnolinguistic group.

2.3 Control variables

The regressions include a long list of control variables. Among the geographic controls we consider the distance to the coastline, the distance to closest river or lake, and the ruggedness index. The climate controls include average temperature and average precipitation from 1961-1980. Additionally we include population density, the log of the area, the share of mining and fertile soil, the distance to the Equator, ecological diversity and the degree of pathogen stress. Ecological diversity captures bio-geographic diversity based on eco-regions, which are defined as relatively large units of land or water containing a distinct assemblage of natural communities sharing a large majority of species, dynamics, and environmental conditions. We construct an index of ecological fractionalization as in Fenske (2014). Pathogen stress measures the extent of disease prevalence during precolonial times using a general measure of pathogen that includes the presence and intensity of seven pathogens[10].

3 Basic Results

The basic specification is

$$\ln y_{ijt} - \ln y_{ij0} = \alpha_j + \beta \ln y_{ij0} + \gamma FRAC_{ij} + \sum \gamma_k z_{kij} + \epsilon_{ij}$$

where i and j refer to a cell and a country respectively, and y_{ijt} and y_{ij0} are night light per capita in 2010 and 1992 respectively[11]. FRAC is the level of ethnic fractionalization at each country-cell. Using this arbitrary geographical area, we minimize concerns over the possible endogeneity of the political boundaries highlighted in the cross-sectional empirical literature. We also include controls for geographic and climate variables. To control for other factors that are country-specific, we include country fixed effects. This is another advantage with respect to cross-country regressions. In fact, if we increase the size of the country-cells we ultimately reach the size of each country[12]. We should notice that cells that include areas of two or more countries are divided in as many cells as countries. In this way we make sure that the limiting case of the expansion of the size of the cells is a particular country. This is also the reason why we describe the elements of the grid as country cells: each cell, or part of a standard size cell, belongs only to one country. Table

[10]For a full description of the variables included in the empirical analysis see the Online Appendix.

[11]The measure of night light per capita is generally adopted as the good proxy for GDP per capita at high resolution. Pinkovskiy and Sala-i-Martin (2016) similarly use night light per capita in all of their baseline regressions.

[12]Montalvo and Reynal-Querol (2017a) perform a systematic analysis of the effect of the size of geographical units on the relationship between ethnic diversity and growth. They find a positive relationship for small geographical areas and no effect for large areas and countries.

2 shows the results. Columns 1 presents the estimators for the specification that only includes country fixed effects and robust standard error. The estimation shows a positive relationship between ethnic fractionalization and growth controlling for the initial level of development.[13] Column 2 presents the same basic regression using clustered standard errors at the country level. The standard errors almost double, but the estimators are still statistically significant and, in the case of ethnic diversity, positive.

[13] In the regressions of Table 2 the speed of convergence implied by the coefficient of the log of the initial level of development ranges from 1.6% to 2.4%, values similar to those typically found for the speed of convergence across regions or countries. This result indicates that night light density generates similar results to the those found with other indicators of economic development.

Table 2: **Ethnic diversity and growth**

	Dependent Variable: Growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log night light 1992	-0.370***	-0.370***	-0.369***	-0.367***	-0.369***	-0.362***	-0.364***	-0.364***
	[0.007]	[0.022]	[0.021]	[0.021]	[0.021]	[0.021]	[0.021]	[0.021]
								(0.013)
Ethnic Fractionalization	0.617***	0.617**	0.695***	0.708***	0.699***	0.568***	0.610**	0.621**
	[0.100]	[0.287]	[0.264]	[0.250]	[0.246]	[0.266]	[0.258]	[0.240]
								(0.146)
Geographic Variables	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Climate Variables	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	No	No	No	Yes	Yes	Yes	Yes	Yes
Share Mining and Fertile Soil	No	No	No	No	Yes	Yes	Yes	Yes
Log. Area Obs	No	No	No	No	No	Yes	Yes	Yes
Country and Coastline Border	No	No	No	No	No	No	Yes	Yes
Dist. Equatorial Line	No	No	No	No	No	No	Yes	Yes
Ecological Diversity	No	No	No	No	No	No	No	Yes
Pathogen stress	No	No	No	No	No	No	No	Yes
Observations	21514	21514	21514	21514	21514	21514	21514	21514
R-squared	0.294	0.294	0.298	0.298	0.299	0.299	0.306	0.306

Notes - In Column 1 robust standard error in brackets. In Columns 2 to 7, Robust standard error clustered at country level are reported in brackets. In Column 8, Conley standard errors in parenthesis (Spatial correlation kernel cutoff = 200km). * Significant at 10%, ** significant at 5%, and *** significant at 1%. Country fixed effects are included. Geographic Variables include: distance to coastline, distance to closest river or lake, and the ruggedness index. Climate Controls: Average temperature from 1961-1980 and average precipitation from 1961-1980.

3.1 Controlling for observable variables

The basic results are robust to adding an large number of geographical and climate controls. Column 3 of Table 2 adds the basic geographic (distance to the coastline, distance to the closest river or lake, and ruggedness) and climate controls (average temperature and precipitation). The effect of ethnic diversity is positive and a bit higher than in the previous specification. Adding population density (column 4) and the share of mining/fertile soil (column 5) does not affect the estimate. It is interesting to notice that including population density does not affect the size of the coefficient on the effect of ethnic diversity on growth[14]. When using cross-country data, the high correlation between ethnic diversity and population density has an important effect on the estimation of the parameter associated with ethnic diversity. However, at the cell level, we do not observe that effect. The borders of ethnic homelands do not concentrate areas with higher/lower proportion of cities, or population density, than other areas. Many big cities, with high population density, are not located at these borders.

The results are also robust to including many other observable variables, like the inclusion of the area of the cells (column 6); the results are also robust to the inclusion of a dummy for country or coastal borders, and distance to the equator (column 7). Finally, Column 8 includes also ecological diversity and pathogen stress as controls[15]. The results are basically unaffected. Inferential conclusions are not altered by using Conley's approach to correct the standard errors for spatial correlation as reflected in the curly brackets of column 8. These results indicate that there is a positive relationship between local ethnic diversity and local growth at a very high degree of geographical resolution. Using the results of the last column, an increase in the degree of ethnic heterogeneity of two standard deviations implies an annual increase of output per capita of 1.1 percentage

[14] We are indebted to Roman Wacziarg for suggesting this exercise.

[15] For the description of these variables see the Online Data Description

points[16].

3.2 Urban agglomeration and migration

The positive relationship between ethnic diversity and local growth could be capturing an agglomeration effect related with the presence of large cities in the cell and, therefore, have a level of relationship different from the country cells that we use as the basic unit of observation. The empirical literature showing a positive impact of ethnic diversity on growth refers mostly to cities. Moreover, related to this "urban premium," cities have higher productivity and higher wages than other areas. Is it only the urban premium what drives our results? The results of Table 3 indicate that it is not only agglomeration effects what support the previous findings. In Column 1 we include a dummy for the national capital while in Column 2 we also add dummies for provincial capitals. The effect of diversity is still present after controlling for capital cities. Column 3 adds dummies for urban agglomerations, considering as such urban areas with more than 500,000 inhabitants. In all the cases, the basic results of Table 2 are maintained: ethnic diversity has a positive and significant effect on growth.

[16] To relate this to GDP per capita, we note that Pinkovskiy and Sala-i-Martin (2016) cannot reject the null hypothesis that the weight of log GDP per capita is 1 in the optimal light night-based proxy of true income. The same result holds for the subsample of Africa.

Table 3: **Ethnic diversity and growth: the role of agglomeration**

Dependent Variable: Growth

				Drop all-urban center	Drop 10% richest	Drop 20% richest	Drop 10% most densely	Drop 20% most densely
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log night light 1992	-0.363***	-0.368***	-0.367***	-0.358***	-0.364***	-0.372***	-0.369***	-0.358***
	[0.021]	[0.021]	[0.021]	[0.021]	[0.015]	[0.016]	[0.022]	[0.022]
Ethnic Fractionalization	0.619**	0.611**	0.609**	0.620**	0.575**	0.746***	0.628**	0.523*
	[0.240]	[0.238]	[0.237]	[0.282]	[0.251]	[0.163]	[0.265]	[0.275]
Nat. Capital (yes=1)	-0.344***	-0.575***	-0.536***		-0.571***	-0.605***	-0.681***	-1.312***
	[0.091]	[0.090]	[0.078]		[0.074]	[0.076]	[0.190]	[0.274]
Prov. Capital (yes=1)		0.416***	0.451***		0.465***	0.472***	0.845***	1.246***
		[0.128]	[0.132]		[0.128]	[0.124]	[0.169]	[0.213]
Urb. Agglom (yes=1)			-0.150		-0.337**	-0.620***		
			[0.144]		[0.131]	[0.117]		
Controls from Table 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,514	21,514	21,514	18,174	19,151	16,997	19,148	16,997
R-squared	0.306	0.307	0.307	0.298	0.295	0.314	0.303	0.294

Notes - Robust standard error clustered at country level are reported in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%. Country fixed effects are included. Controls from Table 2 include: distance to coastline, Ruggedness Index, Average temperature from 1961-1980, average precipitation from 1961-1980, Population Density, Area, Share Mining, % Fertile Soil, Distance to River, Distance to Lake and Border (yes=1), Country and Coastline Border, Dist. Equatorial Line, Ecological Diversity, Pathogen stress.

To continue investigating the issue of the influence of urban agglomeration, Columns 4 to 8 of Table 3 restrict the sample to different subsets of country cells without urban centers. Column 4 excludes the cells that contain an urban center.[17] Columns 5 and 6 exclude the richest areas (upper 10% and 20% respectively) mostly associated with the presence of urban areas. The most densely populated cells are also associated with the presence of urban metropolitan areas. For this reason, Columns 7 and 8 present the estimation dropping from the sample the cells with the highest population density (upper 10% and 20% respectively). Overall, the results of Table 3 indicate that the relationship between ethnic diversity and local growth are not driven only by the agglomeration effect associated with the presence of urban areas.

This exercise also addresses a potential measurement problem associated with the ethnic distribution of the population living in urban centers and, in particular, in capitals. Papaionnaou and Michalopoulos (2014) claim that, under the assumption that in a given urban center the respective indigenous group is relatively more populous than recent migrant groups, this should not be an important concern. We have shown in this section that our basic results are robust to the elimination of urban centers from the sample.

Finally, there is the issue of potential simultaneity of ethnic heterogeneity and growth: high growth areas may attract more diverse populations than stagnant ones[18]. In order to address the potential impact of postcolonial migration to prosperous countries, and subsequent increases in ethnic diversity in those areas, we follow the strategy of Ashraf and Galor (2013)[19]. We perform the analysis restricting the sample to specific sets of countries depending on their attractiveness to migrants or the migratory distance from East Africa: countries that do not belong to the OECD and, therefore, are less attractive

[17]See Online Appendix.

[18]Ashraf and Galor (2013) argue that the direction of the potential endogeneity bias is ambiguous a priori since wealthier societies can have advance military technology to minimize invasions of foreigners.

[19]Previous versions of the paper addressed this issue using two instrumental variables. The results confirmed the findings discussed above.

to migrants; non-Neo-European countries (i.e. excluding US, Canada, Australia and New Zeeland); non-Latin-American countries; non-Sub-Saharan African countries; and the complementary of all the previous samples. Montalvo and Reynal-Querol (2017a) show that the effect of diversity on growth remains statistically significant in all these restricted samples, with the parameter estimate moving between 0.7 and 0.8.

3.3 Some additional analyses of robustness

We have run many other robustness analyses. In Table 4 we check the robustness of the results to the use of alternative measures of ethnic diversity. In Column 1 we find that the results are robust to the use of ethnic fractionalization calculated as the percentage of population living in the ethnic homeland. The empirical findings are basically unaffected if we use other sources of ethnic diversity such as Ethnologue (Column 2). We also calculate fractionalization using ancestral ethnic homelands. To construct this variable we use Murdock's data in the analysis of the African case (Columns 3)[20]. This is reassuring because calculating ELF using ancestral ethnic homelands mitigates the concerns about the endogeneity of migration to local growth given that ancestral ethnic homelands do not reflect migrations.

In Column 4 we show that results are robust to dropping the outliers.[21]

[20]The results are also unaffected if we measure diversity as the number of ethnic groups.

[21]The results are also robust to the use of the level of night light per capita instead of the growth rate. Including regressions by continents do not alter the results.

Table 4: **Ethnic diversity and growth: additional robustness checks**

Dependent Variable: Growth

				Without Outliers FRAC
	(1)	(2)	(3)	(4)
Log night light 1992	-0.363***	-0.378***	-0.384***	-0.355***
	[0.021]	[0.022]	[0.035]	[0.022]
Ethnic Fractionalization POP	0.647***			
	[0.224]			
Ethnic Fractionalization (Ethnologue)		0.553***		
		[0.161]		
Ethnic Fractionalization 1800 (Murdock)			0.883***	
			[0.258]	
Ethnic Fractionalization				0.682**
				[0.341]
Controls from Table 2	Yes	Yes	Yes	Yes
Observations	21481	19822	3654	19148
R-squared	0.306	0.314	0.270	0.307

Notes - Robust standard error clustered at country level are reported in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%. Country fixed effects are included. Controls from Table 2 include: distance to coastline, Ruggedness Index, Average temperature from 1961-1980 , average precipitation from 1961-1980, Population Density, Area, Share Mining, % Fertile Soil, Distance to River, Distance to Lake and Border (yes=1), Country and Coastline Border,Dist. Equatorial Line,Ecological Diversity, Pathogen stress

We have also analyzed the sensitivity of the results to the grid generating coordinates. Therefore, we consider the possibility that the results shown in previous sections are the outcome of a specific initial point generating the grid. For this purpose we produce 100 grids with random initial coordinates. More specifically, we take our initial coordinates (longitude -180; latitude -89) and add to both a random number generated by two uniform distributions[22]. The results show that all the parameter estimates are statistically significant no matter the initial coordinates of the grid. In addition, the estimates move mostly in a close range between 0.7 and 0.9[23].

4 The Case of Africa

Why is diversity good for units of small size? In the first section we discussed several mechanisms that could explain the reduction of the influence of diversity on development as the size of the relevant unit of observation increases. The trade off between the positive effect of diversity on the intensity of innovation and creation of knowledge, and the negative effect of reduction of social capital depend on the size of the area. We argued that the specific mechanisms that explain the basic findings of this paper may depend on the characteristics of the group of countries being analyzed. In this section we suggest a mechanism to explain the positive relationship between ethnic fractionalization and growth in Africa based on the possibility that ethnic diversity can increase trade when observed at high resolution. Assuming that members of different ethnic groups have less trust in one another than in members of the same group, trade across ethnic groups im-

[22]See Montalvo and Reynal-Querol (2017a) for a detailed description of the process of generating these one hundred grids by using random initial coordinates.

[23]We have also investigated if the results are heterogeneous depending on country characteristics (political institutions, level of development of the country and country-wide ethnic diversity) showing that the effect of ethnic heterogeneity, at this level of local analysis, does not depend robustly on the quality of country's political institutions. See Montalvo and Reynal-Querol (2017a).

plies the need to monitor, and be able to retaliate in cases of non-fulfillment of contract conditions. Trade across ethnic groups therefore requires proximity. As it is not possible to find data on trade across ethnic groups and thus provide some evidence of the likelihood of this mechanism, we rely on an indirect argument. As shown by Michalopoulous (2012), ethnic groups tend to specialize. Using our data we similarly find evidence of ethnic specialization. Moreover, high variability in the proportion of crops in an area is associated with high growth. Therefore, one mechanism that can explain the positive effect of ethnic diversity on development is the fact that the specialization in production of different ethnic groups provides larger opportunities for welfare improvement through trading with other ethnic groups. This implies that if this effect is larger than the transaction cost associated with lower levels of trust or communication issues, we should find local markets at the ethnic borders.

A theoretical explanation for these effects can be derived from a variation of the trade game with social norms presented in Rohner, Thoening, and Zilibotti (2013). The salience of social norms is heterogeneous across individuals and groups, and determines the psychological benefit derived from agents by playing cooperatively. This benefit is assumed to be group specific but exogenous. However, it seems reasonable to assume that this psychological benefit depends on the proximity of ethnic groups: it is more likely that nearby groups share some social norms, and have less prejudice, than those far away from one another. The contact theory proposed by Allport (1954), a well established idea in social psychology, suggest that contact between members of different groups can work to reduce prejudice and intergroup conflict. Desmet, Gomes and Ortuño (2019) find, in agreement with contact theory, that local learning reduces the antagonism felt towards other ethnic groups. The findings of Robinson (2017) are also consistent with contact theory: local diversity increases interethnic trust. It is important to get along with your neighbors but less important to follow the social norms of individuals with whom little interaction is foreseen. Therefore, the boundaries between ethnic groups should attract trade.

To investigate this mechanism we have chosen the case of Africa, as this is the region of the world where most of the research on the issue of ethnic diversity is concentrated. Africa is a particularly interesting case for analysis when dealing with the relationship between ethnic diversity and growth as the latter is the most ethnically diverse region of the world. Before analyzing the location of markets in Africa, we first show that the general results on the relationship between ethnicity and growth also hold for Africa.

Table 5: **Ethnic diversity and growth: the case of Africa**

Dependent Variable: Growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log night light 1992	-0.382***	-0.378***	-0.379***	-0.380***	-0.380***	-0.381***	-0.380***	-0.380***
	[0.034]	[0.033]	[0.032]	[0.033]	[0.035]	[0.035]	[0.035]	[0.035]
Ethnic Fractionalization	0.912**	0.801**	0.773**	0.743**	0.751**	0.728**	0.920**	0.788**
	[0.374]	[0.369]	[0.372]	[0.377]	[0.380]	[0.378]	[0.464]	[0.403]
"+/- 10 degree from Equatorial Line"							-0.564	
							[0.393]	
"+/- 10 degree from Equatorial Line x Ethnic Frac."							-0.528	
							[0.869]	
"+/- 5 degree from Equatorial Line"								0.089
								[0.346]
"+/- 5 degree from Equatorial Line x Ethnic Frac."								-0.302
								[0.969]
Geographic Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climate Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share Mining and Fertile Soil	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Distance to River and Lake	No	No	No	Yes	Yes	Yes	Yes	Yes
Log. Area Obs	No	No	No	No	Yes	Yes	Yes	Yes
Country and Coastline Border	No	No	No	No	Yes	Yes	Yes	Yes
Dist. Equatorial Line	No	No	No	No	Yes	Yes	Yes	Yes
Ecological Diversity	No	No	No	No	No	Yes	Yes	Yes
Pathogen stress	No	No	No	No	No	Yes	Yes	Yes
Observations	3713	3713	3713	3713	3713	3713	3713	3713
R-squared	0.259	0.264	0.265	0.267	0.268	0.268	0.271	0.268

Notes - Robust standard error clustered at country level are reported in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%. Country fixed effects are included. Controls from Table 2 include: distance to coastline, Ruggedness Index, Average temperature from 1961-1980 , average precipitation from 1961-1980, Population Density, Area, Share Mining, % Fertile Soil, Distance to River, Distance to Lake and Border (yes=1), Country and Coastline Border, Dist. Equatorial Line, Ecological Diversity, Pathogen stress

In Table 5 we run the main regression for the African continent and the relationship between diversity and growth remains positive at the same level of resolution as for the original results for the whole world. Columns 1 to 6 show that the basic result is unaffected by the inclusion of an increasing set of observables. Within Africa, areas around the equator seem to be the poorest and most diverse. For this reason we check carefully the role of the equator in the basic relationship. In Column 7 we include a dummy that has value 1 if the cell is on or within an area of +10/-10 degrees from the equatorial line, and zero otherwise. Column 8 reports the results of the regression using a dummy that takes value 1 if the cell is within +5/-5 degrees from the equator. The results are robust to the inclusion of these controls for the equator. Therefore, we also observe in Africa that more ethnically diverse areas are also those that grow faster.

4.1 Interethnic trade and markets in Africa

There are a number of indications that interethnic trade is at the origin of many African local markets. Most of this interethnic trade took place at the boundaries of ethnic homelands. The evidence draws mostly on work by geographers, anthropologists, and historians who have studied the origin of traditional markets in Africa before the arrival of Europeans. For example, Hodder (1965) provides a voluminous body of evidence to support the view that external trading contacts were critical for the genesis of markets in Africa. Hodder (1965) remarks that "the analysis of Yorubaland, for example, indicated that traditional markets were often located at junction zones, areas in which products of each area could be easily exchanged. Also markets were found at the junction of different people: Ketu market, for instance, was regarded as an important link between Yoruba and Dahomey peoples; Iperu market was a contact point between Egba and Ijebu groups of the Yoruba; mamu market was a traditionally frontier market between Ijebu and Ibadan Yoruba" (p. 99). Hodder (1965) adds that "traditional markets are also found among the more southerly groups of the largest tribe, the Kikuyu. These Southern Kikuyu are known to have traded not only among their own tribal groups and sub-groups but also with neighboring tribes, notably the Masai and Kamba, who in turn traded with Arab and Swahili caravans and acted as middlemen between the coastal and interior traders. In Kenya, too, the Teita have traditional markets and have long been noted for their caravan trading to and from the coast. The Buganda, Busoga and Swahili-speaking coastal peoples also have traditional markets; and all are known to have had important trading contacts with peoples and routes outside their own territorial boundaries. Finally in East Africa, the coastal Digo tribe of the north-eastern Bantu are unlike their immediate Bantu neighbors in having traditional market institutions; and these Digo, significantly, have long enjoyed 'an influential position as middlemen in the ivory trade and traded with the Swahili, Arab and Indian merchants..... Even among those peoples where traditional markets do not exist, a few isolated traditional markets may often be found around the periphery of the tribal lands where inter-tribal trade, for instance, seems for long to have existed along the Ubangi River in boundary between the Ngbandi and Banda peoples. Similar peripheral found along the borders of the Ruandi and the Urundi groups" (p. 101).

There is plenty of additional evidence of the interethnic origin of markets in Africa. Meillassoux (1965) analyzes the case of the Guro land in Ivory Coast, where "markets among the Guro of central Ivory Coast tend to be localized at the contact area between complementary zones," supporting the conclusion that "markets are primarily induced by external exchange of complementary products with an alien population. When such a situation occurs, the markets tend to be localized at the contact area between complementary zones. Hence, they can help to indicate the limits of substantive economic areas" (p. 297-298). Vansina (1965) draws an almost identical conclusion with regard to the traditional markets of the Kuba peoples of present-day Zaire.

Roberts (1970) highlights that precolonial commerce in the interior of Tanzania was an activity involving different peoples including Nyamwezi, Sumbwa, Gogo, Taturu, Sukuma, Vinza and Sagara who exchanged complementary products which circulated within and between regional trade networks.

Yet another compelling example is that of the Abyssinian market town covered by Messing (1965): "There is relatively little exchange of any kind outside the extended-family and rural hamlets except for that taking place on daily and weekly markets. On certain seasonal occasions, over 1,000 persons may gather in and about the weekly market at Gondar. Money is used as both medium of exchange and standard of value. The market is closely related to the division of labor which is caste-like in its ethnic specialization of occupations, such as smiting, pottery-making and tanning" (p. 387).

4.2 Empirical results

Assuming that trade at the local level usually takes place in markets, we look for data on the latter. It is difficult to find the location of markets in Africa since there are potentially thousands of very small markets where farmers may sell their products. Therefore, we need some criteria to identify relevant markets which is exogenous to our methodology. Porteous (2019) identifies the location of 223 regionally important hub markets in Africa.[24] Around 60% of markets located in cities of more than 100,000 inhabitants and 40% in smaller villages.

[24]Porteous (2019) considers 230 markets but there are 223 placed in different locations. We should notice that, obviously, there are many small markets not included in the list.

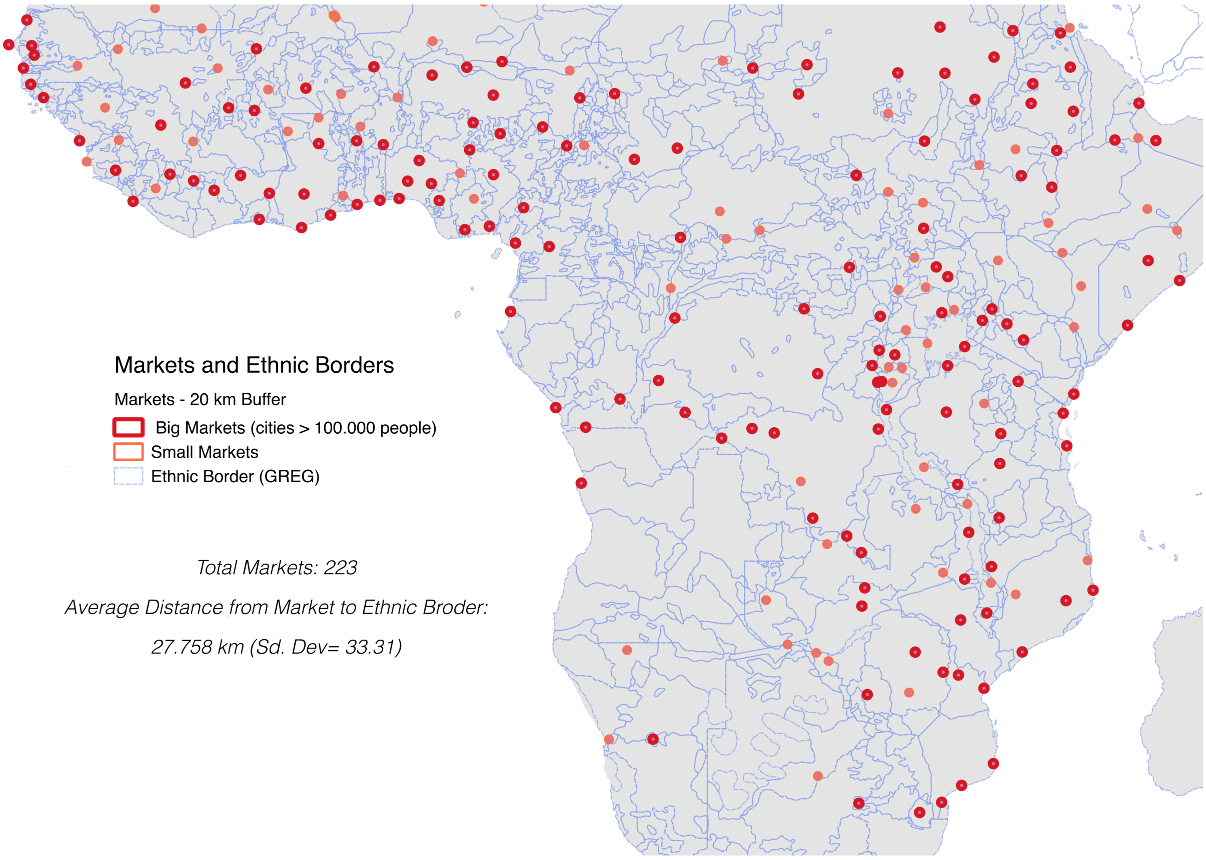

Figure 1: Market location and ethnic borders.

Source: Own elaboration using merged datasets.

We overlap Porteous' map with our own map of the spatial distribution of ethnic groups in Africa in Figure 1. It is easy to see that many of the markets are very close to ethnic borders. In fact, the average distance to the closest ethnic border among all the markets is just 27 km, which seems to indicate that trading markets are located close to ethnic borders[25]. In order to show how far the actual distribution of markets is with respect to a random geographic distribution, we run a simulation with 500 random samples of 223 locations in Africa (equal to the number of markets taken from Porteous, 2019)[26]. We consider continental sub-Saharan Africa, which is the area similarly covered by Porteous (2019), and use the Haversine formula to estimate the distance of each simulated market to the closest ethnic border. Finally, we take the average distance to the closest ethnic border for each of the 500 simulations. The results show that the average distance of 27 km is at the 1% of the distribution[27]. This indicates that markets are much closer to ethnic borders than randomly generated locations.

[25]Montalvo and Reynal-Querol (2017a) zoom in one of Figure 1's typical markets (the Gambela market), to show how ethnic groups and ethnic borders are distributed around it.

[26]We thank Stelios Michalopoulos for this suggestion.

[27]See Montalvo and Reynal-Querol (2017a)

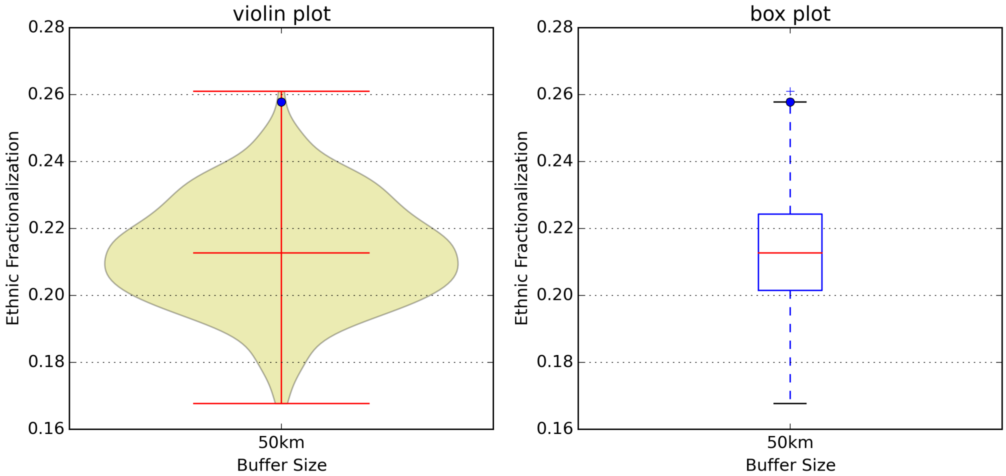

Figure 2: Market location and ethnic borders.

Source: Own elaboration using merged datasets.

Figure 2 provides additional evidence for the concentration of trading along the ethnic borders. It shows the average ethnic diversity index for the actual location of the markets and the 500 simulations of markets' distribution in Sub-Saharan Africa. For the placebo analysis we randomly generate the location of 223 "virtual" markets, or the actual number of markets identified by Porteous (2019). In each simulation we use a buffer of 50km around each point to calculate their index of ethnic diversity. Figure 2 shows that the ethnic diversity of the actual markets is in the tail of the distribution of heteregeneity indices of the "virtual" markets.

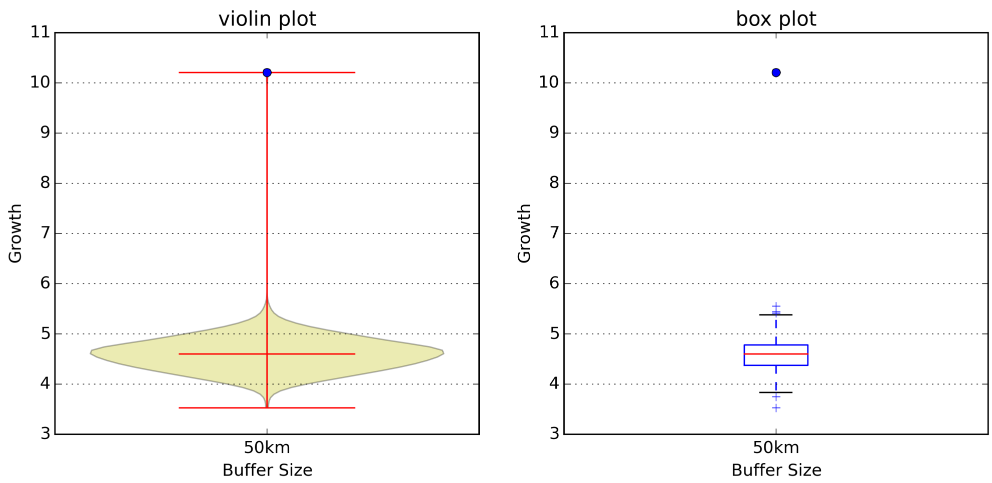

Figure 3: Market location and ethnic borders.

Source: Own elaboration using merged datasets.

Figure 3 runs a similar exercise to Figure 2 but relating market location and growth. It shows that the growth rate around the markets is much higher than the growth rate around the "virtual" markets.

Previous evidence shows that areas that have more ethnic diversity also have more markets. An explanation is based on the specialization of ethnic groups. The geographic proximity of ethnic groups may also increase trade if they are highly specialized in the production of specific agricultural products or services. While the level of trust among different ethnic groups may, in general, be smaller than intragroup trust, the fact that they are geographically close to one another can facilitate monitoring and, therefore, counterbalance the potential lack of trust. Jha (2013) shows that medieval Hindus and Muslims could provide complementary services and a mechanism to share gain from trade which increased tolerance between these two groups. The development of these practices into formal institutions generated inertia in the degree of ethnic tolerance.[28] The location of local markets in Africa seems to support this interpretation.

5 Concluding Remarks

The relationship between ethnic heterogeneity and development is complicated. Empirical research working with cross-country data finds a negative, or null, relationship. However, research at the city level usually finds a positive relationship between diversity and wages and/or productivity. In this paper we find that small areas tend to generate a positive relationship. We argue that an explanation of the positive relationship between diversity and growth in Africa consistent with the data is the increase in trade at the boundaries between ethnic groups due to ethnic specialization.

[28]Jha (2013) also finds that medieval ports, despite being more ethnically diverse, were less prone to conflicts between ethnic groups.

References

Aghion, Philippe, and Peter W. Howitt, *The Economics of Growth*, Cambridge, Massachusetts, MIT Press, (2009).

Aker, Jenny C., Klein Michael, O'Connell Stephen A., and Yang Muzhe, "Borders, ethnicity and trade", *Journal of Development Economics*, 107 (2014), 1-16.

Alesina, Alberto, Johann Harnoss, and Hillel Rapoport, "Birthplace Diversity and Economic Prosperity", *Journal of Economic Growth*, 21 (2) (2016), 101-138.

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou, "Ethnic Inequality", *Journal of Political Economy*, 124 (2) (2016), 428-488.

Alesina, Alberto, Devleeschauwer Arnaud, Easterly William, Kurlat Sergio, and Wacziarg Romain, "Fractionalization", *Journal of Economic Growth*, 8 (2003), 155-194.

Alesina, Alberto, and Eliana La Ferrara, "Ethnic diversity and economic performance", *Journal of Economic Literature*, 43 (3) (2005), 762-800.

Alesina, Alberto, Enrico Spolaore, and Romain Wacziarg, "Economic Integration and Political Disintegration", *American Economic Review*, 90 (5) (2000), 1276-96.

Allport, Gordon Willard, Kenneth Clark, Thomas Pettigrew, *The nature of prejudice*, Cambridge/Reading, MA: Addison-Wesley, (1954).

Ashraf, Quamrul and Galor Oded, "The out of Africa hypothesis, human genetic diversity and comparative development", *American Economic Review*, 103 (1) (2013), 1-46.

Bakens, Jessie, Peter Mulder, and Peter Nijkamp, "Economic impacts of cultural diversity in the Netherlands: productivity, utility, and sorting", *Journal of Regional Science*, 53 (2013), 8-36.

Chen, Xi, and William D. Nordhaus, "Using Luminosity Data as a Proxy for Economic Statistics", *Proceedings of the National Academy of Sciences*, 108 (21) (2011), 8589-8594.

Collier, Paul, "Ethnic diversity: An Economic Analysis of its Implications", *Economic Policy*, 32 (2001), 129-166.

Desmet, Klaus, Joseph Flavian Gomes, and Ignacio Ortuño-Ortín, "The geography of linguitic diversity and the provision of public goods", forthcoming *Journal of Development Economics*, (2019).

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg, "The Political Economy of Linguistic Cleavages", *Journal of Development Economics*, 97 (2012), 322-338.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg, "Culture, ethnicity and diversity", *American Economic Review*, 107 (9) (2017), 2479-2513.

Easterly, William, "Can Institutions Resolve Ethnic Conflict", *Economic Development and Cultural Change*, 49 (4) (2001), 687-706.

Easterly, William, and Ross Levine, "Africa's growth tragedy: Policies and Ethnic divisions", *Quarterly Journal of Economics*, 112 (4) (1997), 1203-1250.

Easterly, William, and Fenske, James, "Ecology, trade and states in pre-colonial Africa", *Journal of the European Economic Association*, 12 (3) (2014), 612-640.

Glaeser, Edward L., "The political economy of hatred", *Quarterly Journal of Economics*, (2005), 45-86.

Gören, Erkan, "How Ethnic Diversity Affects Economic Growth", *World Development*, 59 (2014), 275-297.

Greif, Avner., "The birth of impersonal exchange: the Community responsibility system and impartial justice", *Journal of Economic Perspectives*, 20 (2) (2006), 221-236.

Greif, Avner., "The fundamental problem of exchange: A research agenda in Historical Institutional Analysis", *European Review of Economic History*, 4 (2000), 251-284.

Greif, Avner., "Contract enforceability and economic institutions in Early trade: the Maghribi traders' coalition", *American Economic Review*, 83 (3) (1993), 525-548.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space", *American Economic Review*, 102 (2012), 994-1028.

Hodder, Bramwell William, "Some comments on the origin of traditional Markets in Africa South of the Sahara", *Transactions of the Institute of British Geographers*, 36 (1965), 97-105.

Jha, Saumitra, "Trade, Institutions and Ethnic Tolerance: Evidence from South Asia", *American Political Science Review*, 107 (4) (2013), 806-832.

Jean, Neal, Burke Marshall, Xie Michael, Davis W. Matthew, Lobell David B., and Ermon Stefano, "Combining satellite imagery and machine learning to predict poverty", *Science*, 353.6301 (2016): 790-794.

Meillassoux, Claude, "Social and economic factors affecting markets in Guro land", in Bohannan, P. and Dalton, G. (Eds.), *Markets in Africa*, Evanston: Northwestern University Press, (1965).

Messing, Simon D., "The Abyssinian market town", in Bohannan, P. and Dalton, G. (Eds.), *Markets in Africa*, Evanston: Northwestern University Press, (1965).

Michalopoulos, Stelios, "The Origins of Ethnolinguistic Diversity", *American Economic Review*, 102 (4) (2012), 1508-1539.

Michalopoulos, Stelios, and Elias Papaioannou, "Pre-colonial Ethnic Institutions and Contemporary African Development", *Econometrica*, 81 (1) (2013), 113-152.

Michalopoulos, Stelios, and Elias Papaioannou, “National Institutions and Subnational Development in Africa”, *Quarterly Journal of Economics*, 129 (1) (2014), 151-213.

Michalopoulos, Stelios, and Elias Papaioannou, “The Long Run Effects of the Scramble for Africa”, *American Economic Review*, 106 (7) (2016), 1802-1848.

Montalvo, Jose G., and Reynal-Querol Marta, “Ethnicity and growth: revisiting the evidence”, BarcelonaGSE Working Paper 992, (2017a).

Montalvo, Jose G., and Reynal-Querol Marta, “Earthquakes and Inequality”, mimeo, (2017b).

Montalvo, Jose G., and Reynal-Querol Marta, “Ethnic Diversity and Economic Development”, *Journal of Development Economics*, 76 (2005), 293-323.

Nathan, Max, “The long term impact of migration in British cities: diversity, wages, employment and prices”, SERC Discussion Paper, 67 (2011).

Ottaviano, Gianmarco IP, and Giovanni Peri, “Cities and Cultures”, *Journal of Urban Economics*, 58 (2) (2005), 304-307.

Ottaviano, Gianmarco IP, and Giovanni Peri, “The economic value of cultural diversity: evidence from US cities”, *Journal of Economic Geography*, 6 (2006), 9-44.

Pinkovskiy, Maxim, and Xavier Sala-i-Martin, “Lights, camera... income! Illuminating the national account-household surveys debate”, *Quarterly Journal of Economics,*, (2016) 579-631.

Porteous, Obie, “High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Agricultural Storage and Trade”, forthcoming *American Economic Journal: Applied Economics*, (2019).

Roberts, Andrew, “Nyamwezi Trade,” in Gray, R. and Birmingham, D. (Eds.), *Precolonial African Trade*, London, (1970).

Robinson, Amanda Lea, “Ethnic diversity, segregation and ethnocentric trust in Africa,” *British Journal of Political Science*, (2017).

Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti, “War Signals: a Theory of Trade, Trust and Conflict”, *Review of Economic Studies*, 80 (3) (2013), 1114-1147.

Sparber, Chad, “Racial Diversity and Macroeconomic Productivity across US States and Cities”, *Regional Studies*, 44 (1) (2010), 71-85.

Suedekum, Jens, Katja Wolf, and and Uwe Blien, “Cultural diversity and local labor markets,” *Regional Studies*, 40 (2014), 173-191.

Vansina, Jan, “Trades and markets among the Kuba”, in Bohannan, P. and Dalton, G. (Eds.), *Markets in Africa*, Evanston: Northwestern University Press, (1965).

Weidmann, Nils B., Jan Ketil Rd, and Lars-Erik Cederman, “Representing Ethnic Groups in Space: A New Dataset”, *Journal of Peace Research*, 47 (4) (2010): 491-99.

Ethnic Diversity and Growth: Revisiting the Evidence

Online Appendix

Data Description

José G. Montalvo*

Marta Reynal-Querol†

*Corresponding author: José García Montalvo, Professor of Economics, Pompeu Fabra University, Ramón Tras Fargas 25-27, Barcelona 08005. Telephone: (+34) 93 542 2509. E-mail jose.garcia-montalvo@upf.edu

†Marta Reynal-Querol, ICREA Research Professor, Pompeu Fabra University, Ramón Tras Fargas 25-27, Barcelona 08005, Spain. E-mail: marta.reynal@upf.edu

1 Unit of Analysis

Our baseline units of analysis (OBS) are the result of intersecting 1 degree per 1 degree cell grids and the national political borders [1]. Political borders are provided by The Global Administrative Unit Layers (2010) –GAUL– project from UN Food and Agriculture Organization (FAO). The GAUL compiles a very high quality information on the different administrative units for all the countries in the world [2]. We use EPSG:4326 - WGS 84 as our coordinate system. To estimate distance and areas in meters, we re-project to EPSG:3857 - WGS 84.

2 Variables

Variable	Definition

[1]Based on the extend of the political borders layer, we set the starting point for the grid cells at longitude = -180 and latitude = -89.

[2]Data is available at: http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691&currTab=simple

Log night light 1992	To build a proxy income per-capita we estimate the nighlight per-capita in 1992 and 2010 for each cell. Firstly, we use the cloud-free night-light data provided by the NOAA'S National Geophysical Data Center, specifically the Earth Observation Group (EOG)[3]. It provides information at 30 arc second grids[4], on the average quantity of light observed at each grid across cloud-free nights for every year[5]. We use information from 1992 and 2010 collected by satellites F10 and F18 respectively. Although information are collected using different satellites,it is comparable. Secondly, to estimate population we use the Gridded Population of the World (GPWFE) [6]. Based on national census and satellite images, it provides information on human population at 2.5 arc-minutes resolution for 1990, 1995, and 2000, 2005 (projected), 2010 (projected) and 2015 (projected). Once we estimate the total nighlight and population per cell, we estimate the proxy of the economy growth as follows: $$Growth_{1992-2010} = ln\left(\frac{0.1 + nighlight_{2010}}{0.1 + population_{2005}}\right) - ln\left(\frac{0.1 + nighlight_{1992}}{0.1 + population_{1990}}\right)$$

[3]For further information and data at http://www.ngdc.noaa.gov/eog/

[4]Spanning from -180 to 180 degrees longitude and -65 to 75 degrees latitude

[5]Values range from 1 to 63.

[6]Center for International Earth Science Information Network (CIESIN), Columbia University; United Nations Food and Agriculture Programme (FAO); and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World: Future Estimates (GPWFE). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw. (downloaded on July 2015).

Ethnic Fractionalization	As measure of ethnic diversity at cell level we use the Herfindhal Index –HI– of the ethnic groups territories. We use Geo-referencing of Ethnic Groups -GREG- data set[7], which provides the geospatial location of ethnic groups as polygons. By intersecting our cells and group territories, we are able to estimate the share of area within each cell by groups. Then, we estimate the HI as follows: $$FRAC = 1 - \sum_{i=1}^{N} \pi_i^2 = \sum_{i=1}^{N} \pi \left(1 - \pi\right) \quad (1)$$ where π is the proportion of area/population who belong to ethnic group i in a given ($\forall i = 1, ..., N$).
Terrain Ruggedness Index, 100m	We use the Terrain Ruggedness Index used by Nunn and Puga (2012)[8]. Each cells on a 30 arc-seconds grid across the surface of the Earth provides the Terrain Ruggedness Index, in millimeters[9]. Then, we estimate the weighted average, using as weights the area of each cell[10], for each of our obs. Finally, as suggested by Nun and Puga (2012), we obtain divide values by 100,000 to obtain the Terrain Ruggedness Index in hundreds of meters.

[7]Further information and data available at http://www.icr.ethz.ch/data/other/greg

[8]Nun, N and Puga, D (2012). Ruggedness: The blessing of bad geography in Afric. Review of Economics and Statistics 94(1), February 2012: 20-36

[9]Rawdata is available at http://diegopuga.org/data/rugged/tri.zip

[10]Getting the weighted average is important to take into account that the sea-level surface that corresponds to a 30 by 30 arcsecond cell varies in proportion to the cosine of its latitude (Nun and Puga, 2012).

Fertile Soil	We use the information of fertile soil used by Nun and Puga (2012), which determines whether a each cell on a 5-minute grid covering almost the entire land area of the Earth is subject to various constraints for growing rain-fed crops[11]. Thereby, we estimate the overall mean of the fertile cell at our cells level.

[11]This information is originally created by Fischer, van Velthuizen, Shah, and Nachtergaele (2002), based on the FAO/UNESCO Digital Soil Map of the World, the soil association composition table and climatic data compiled by the Climate Research Unit of the University of East Anglia. Based on plates 20 (soil moisture storage capacity constraints), 21 (soil depth constraints), 22 (soil fertility constraints), 23 (soil drainage constraints), 24 (soil texture constraints), and 25 (soil chemical constraints) in Fischer et al. (2002) and the country boundaries described above, we calculate the percentage of the land surface area of each country that has fertile soil (defined as soil that is not subject to severe constraints for growing rain-fed crops in terms of either soil fertility, depth, chemical and drainage properties, or moisture storage capacity). In addition, Nun and Puga (2012) include Cape Verde, French Polynesia, Mauritius and Seychelles that were not covered by the Fischer et al. (2002) data, they we use instead the percentage of their land surface area that is classified by the Food and Agriculture Organization (2008) as arable land or permanent crop land.

Share Mining	We use the Seamless Digital Chart of the World (SDCW)[12], which provides a unique information on areas where natural resources are being extracted from the earth[13]. The SDCW is based on the best currently-available global vector base map, Digital Chart of the World (Vector Smart Map 0, Edition 5 from National Geospatial Agency-Intelligence Agency). We are therefore able to determinate the area within each cell that is being used for different type of mining. Then, the share of mining for a given cell i is: $$ShareMining_i = \frac{\mathit{Area\ Natural\ Resources\ Extraction\ (km2)}_i}{\mathit{Total\ Area\ (km2)}_i} \quad (2)$$

[12]www.worldgeodatasets.com/basemaps/

[13]It includes mines/quarries, oil/gas fields, and salt evaporators.

Geographical variables	In order to capture the information for Urban Centers, Rivers and Lakes we use the SDCW DataSet. It provides the exact location of urban centers (either as points or polygons), rivers (polylines) and lakes(polygons). Based on this information, we built for different variables: **Distance to River (km):** Euclidean distance from the centroid of our cell to the nearest river. **Distance to Lakes (km):** Euclidean distance from the centroid of our cell to the nearest Lake. **Distance to Coastline (km):** Euclidean distance from the centroid of our cell to the nearest Coastline. **Distance to Equatorial Line:** It indicates the latitude of the centroid of our cell **Border (yes=1):** It indicates whether a obs is located at the country border **Distance to Equatorial Line:** It indicates the latitude of the centroid of our cell **Coastline Border:** It indicates whether a cell/obs is located at the coastline border **Log. OBS area:** It is the geographical area of each obs

Average precipitation and temperature, 1961-1990	We use 10-minute latitude/longitude data set of mean monthly surface climate over global land areas, excluding Antarctica (CRU TS3.10 Dataset) [14]. It provides a detailed information on the monthly average precipitation (mm/month) and the temperature (Degrees Celsius) from 1961 to 1990 [15]. We first take the overall average of the monthly information at 10-minute latitude/longitude, and then average at our cell level.
Log. Population Density	We use the Gridded Population of the World (GPWFE) [16]. Based on national census and satellite images, it provides information on human population at 2.5 arc-minutes resolution for 1990, 1995, and 2000, 2005 (projected), 2010 (projected) and 2015 (projected).

[14] Harris, I; Jones,P.D.; Osborn, T.J. and Lister, D.H.. "Updated high-resolution grids of monthly climatic observations the CRU TS3.10 Dataset". International Journal of Climatology: Volume 34, Issue 3, pages 623642, 15 March 2014

[15] Data and further methodological information are available through the School of Geography Oxford (http://www.geog.ox.ac.uk) , the International Water Management Institute "World Water and Climate Atlas" (http://www.iwmi.org) and the Climatic Research Unit (http://www.cru.uea.ac.uk).

[16] Center for International Earth Science Information Network (CIESIN), Columbia University; United Nations Food and Agriculture Programme (FAO); and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World: Future Estimates (GPWFE). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw. (downloaded on July 2015).

Ecological Diversity	As measure of ecological diversity at cell level we use the Herfindhal Index –HI– of the ecological zones in the world. We use the Terrestrial Ecoregions of the World (TEOW), which is map with a bio-geographic regionalization of the Earth's terrestrial biodiversity. The bio-geographic units are eco-regions, which are defined as relatively large units of land or water containing a distinct assemblage of natural communities sharing a large majority of species, dynamics, and environmental conditions. Information is available at http://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1036295/. In particular, to obtain the ecological diversity worldwide we use the World Wildlife Fund and Nature Conservancy Terrestrial Ecoregion layer (2011) (available at http://maps.tnc.org/gis_data.html). This data set provides information on 827 eco-region for entire world. Based on this information, we calculate following Fenske (2014), the ecological diversity as a Herfindahl index constructed using the share s_i^t of each society is area that is occupied by each ecological type t. $$\text{Ecological Diversity}_i = 1 - \sum (s_i^t)^2$$

Pathogen Stress Index	In order to estimate the pathogen stress index we follow Low (1991)[17], who measures the extent of disease prevalence during precolonial times using a general measure of pathogen that includes seven pathogens (leishmanias, trypanosomes, malaria, schistosomes, filariae, spirochetes, and leprosy) which are rated on a 3-point scale for severity. The individual scores are added to yield a total pathogen stress score that ranges from a score of 3 to a score of 27. A high score represents many types of pathogens and more severe exposure. We set the value for each observation to the closest observation in the Standard Cross-Cultural sample of Murdock and White (1969)[18]

[17]Low, B. (1990), "Marriage system and pathogen stress in human societies," American Zoologist, 30, 325-339

[18]Murdock and White (1969), "Standard cross-cultural sample," Ethnology, 8: 239-369.