# Ethnic Diversity and Growth: Revisisting the Evidence *

José G. Montalvo Marta Reynal-Querol
Universitat Pompeu Fabra-ICREA

**Abstract**

The relationship between ethnic heterogeneity and economic growth is a complex one. Empirical research working with cross section data finds a negative, or statistically insignificant, relationship. However, research at the city level finds usually a positive relationship between diversity and wages/productivity. In this paper we perform a systematic analysis of the effect of the size of geographical units on the relationship between ethnic diversity and growth. We find a positive relationship for small geographical areas and no effect for large areas and countries. We argue that a possible mechanism to explain the positive relationship between diversity and growth is the increase trade in the boundaries across ethnic groups due to ethnic specialization. Therefore, heterogeneity is good for trade and exchange. But homogeneity can promote good institutions as found in the literature. Therefore what we observe indicate that, at the local level, for a given institutional framework, diversity is good for local development.

*We would like to thank the comments of Antonio Ciccone, Stelios Michalopoulos, Joachim Voth, participants in the Workshop on Conflict of Bocconi, the Meeting of the European Public Choice Society, the EPCS of Groningen, the workshop on Advances on the Political Economy of Conflict and Redistribution III of Berlin, the Invited Session on Political Economy of Development at the EEA 2015 in Mannheim and the workshop on The Political Economy of Social Conflict at Yale. We have also benefited from the comments of participants in seminars at Oxford (CSAE), Stockholm, UPF and CREI. We are also greatful to Juan Carlos Muñoz Mora for excellent research assistance. Financial support from the European Research Council (ERC), the Spanish Ministerio de Educacion, the Barcelona GSE Research Network, and the Government of Catalonia is gratefully acknowledged.

# 1 Introduction

The issue of the effect of ethnic diversity on economic development has generated a large body of literature (see Alesina and La Ferrara 2005 for a review). Cross country regressions generally show a negative effect of heterogeneity on development. These findings have supported the view of the tragedy of Africa as the consequence of its high degree of ethnic diversity (Easterly and Levine 1997). By contrast, research based on data from small geographical areas like cities finds frequently a positive effect of diversity on wages or productivity.

This paper develops a systematic strategy to analyze the effect of diversity on growth by analyzing this relationship at different spatial scales. Previous research has usually relied on political and/or administrative frontiers, or cities/SMA boundaries, to define the unit of analysis. Our analysis considers as the basic units of observation grid cells at a resolution of one by one decimal degree. We construct grids of one degree by one degree and our unit of observation is a grid-country cell.

Obviously at the highest level of resolution of the grid there is no possibility of finding measures of output, value added or even wages when we consider most of the countries. For this reason we take advantage of the luminosity data to proxy for local economic activity. Recent research has shown that light density at night is a good proxy of economic activity. 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 would not resolve the issue of endogeneity caused by the possibility that other unobserved characteristics can drive the association via, for instance, institutional differences. Using these artificially constructed cells mitigates this concern. In addition the results are robust to a large number of potential issues. First, the paper shows that the results are unaffected by using a large number of controls to account appropriately for within country variation, geography, climate, soil quality, proximity to lakes or political capitals, etc. Second, we run one hundred different regressions changing randomly the initial location of the point that defined, 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. Third we show that the positive effect of heterogeneity on economic growth is not due to urban areas and it is not simply capturing an agglomeration effect. Fourth, we present results that show that the baseline results are also robust to two instrumental variables. We do not want to stress specially this result since instruments are tricky objects but the fact that our results are backed by these instrumental variable regressions is reassuring. Finally, we show that reducing the degree of resolution of the grid decreases the association between ethnic diversity and development up to the point of finding no association between heterogeneity and development.

We also increase the degree of resolution constructing grids of 0.5 by 0.5 decimal degrees finding similar results. In order to reinforce our results we perform a pixel level analysis working with four millions pixels of approximately 5km by 5km. At this high degree of resolution the usual diversity measures are not appropriate to analyze how diversity affect economic growth because diversity will appear only on the pixels which contain an ethnic border. Therefore we construct measures of the distance of each pixel to the closest ethnic border. We compare the economic growth of pixels close to the ethnic borders with pixels that are far away from the ethnic borders. We show that economic growth is higher among pixels close to the ethnic borders.

In order to understand why regional development is faster close to ethnic borders we propose a mechanism related to trade. We find that these areas, which have more ethnic diversity, have also a higher proportion of markets. We know that ethnic groups usually specialize in different agricultural products, especially in Africa, and therefore they have incentives for the exchange of goods. Therefore, ethnic groups that are geographically close may have more trade. In order to provide evidence of this potential mechanism we show that local markets in Africa are located close to ethnic borders, which supports this interpretation.

The structure of the paper is the following. Section 2 discusses the literature on the economic effects of ethnic heterogeneity on development and the contributions of the paper. Section 3 describes the data. Section 4 presents the basic results. Section 5 discusses some exercises to show the robustness of the basic results. In section 6 we discuss the effect of changing the level of geographical resolution of the grid. Section 7 propose a mechanism and present some evidence that support it. Section 8 includes some concluding remarks.

## 2 Diversity and growth

Is ethnic diversity good or bad for economic development? The literature has frequently emphasized the trade-off between the benefits of diversity and its cost. Ethnic diversity can be beneficial by enhancing productivity through innovation, skill complementarities, or increased creativity, trade and variety in production. On the negative side, diversity can generate an inefficient provision of public goods, ethnically biased policies and, in general, conflict for disagreement on common public goods and public policies. In order to analyze empirically the relative importance of the costs and benefits of ethnic diversity the literature has adopted two alternative approaches. A few papers have analyzed the endogenous formation of jurisdictions (number, size and shape) modeling the optimal trade-off between the benefits of diversity and the costs of heterogeneity to determine its equilibrium size. Alesina and Spolaore (1997) focus on the trade-off between the benefits of large jurisdictions and the costs of population heterogeneity. Alesina and La Ferrara (2005) propose a model with a positive effect of variety in the production function and a decreasing utility of public goods as the heterogeneity of population increases.

However, most of the literature that analyzes the relationship between ethnic diversity and growth, conflict or the quality of institutions take as given the size of the jurisdiction. This empirical research relies on different sizes of jurisdictions: countries, regions, counties, cities and even villages or schools in developing countries. In this paper we analyze the effect of ethnic heterogeneity as the size of the geographical unit of analysis grows. We consider initially a country cell unit of observation, and analyze the effect of increasing its size up to the frontier of the countries. The size of these units is, in general, smaller than regions and larger than cities. This empirical strategy allows us to analyze systematically the impact of ethnic heterogeneity on development for different levels of geographical aggregation.

The empirical analysis of the effect of ethnic diversity on economic development has relied on different sizes of jurisdictions. Initially most of the papers use cross country regressions. The seminal work by Easterly and Levine (1997) shows, using cross-country differences in ethnic diversity, 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) show, using also cross country data, a consistent negative effect of ethnic fractionalization on growth. Further research has qualified the conditions for this negative relationship. For instance, Collier (2000) finds that ethnic diversity has a negative effect on growth only in non-democratic situations. In the same spirit, Easterly (2001) finds that when ethnic diversity is high, poor institutions have an even more adverse effect on growth and policy. In countries with sufficiently good institutions, however, ethnic diversity does not reduce growth or worsen economic policies. Good institutions also reduce the risk of wars and genocides that may otherwise result from ethnic fractionalization. Alesina and La Ferrara (2005) argue that heterogeneity has more negative effect when the level of income is low. Alesina, Spolaore and Wacziarg (2000) argues that the effects of the size of the countries on economic success is mediated by the extent of freedom of trade. Montalvo and Reynal-Querol (2005) find that ethnic diversity has a direct negative effect on growth while ethnic polarization has an substantial indirect effect through the reduction of the rate of investment and the increase in the likelihood of conflicts[1].

Most of city level studies find that heterogeneity has a positive effect on indicators such as wages or productivity.[2] Ottaviano and Peri (2003) find that US born individuals pay higher rents in heterogeneous cities which imply that diversity has a positive effect on the production and consumption of amenities. Ottaviano and Peri (2004) find that wages of white workers are higher in heterogeneously linguistic cities which can be a reflection of a positive relationship between diversity and productivity. Ottaviano et al. (2006) find that, on average, cultural diversity has a net positive effect on the productivity of US-born citizens. Sparber (2010) also finds a positive relationship between racial diversity and wages across US cities.[3] Lee (2009) uses data of growth in employment for 53 English cities for the period 1981-

[1]Recently, Goren (2014) has found similar results using a larger dataset of countries. Gören (2014) also finds an indirect positive effect of ethnic diversity through international trade.

[2]Studies at higher levels of aggregation, or analyzing variables different from wages at the city level, provide conflicting results. For instance, Dincer and Wang (2011) find a negative relationship between ethnic diversity and economic growth throughout Chinese provinces. Although ethnic diversity does not fully explain the growth differentials among Chinese coastal and inland provinces, the high level of ethnic diversity in inland China appears to be an important factor nevertheless. At city level the research of Gleaser et al. (1995) show that racial heterogeneity does not have any effect on the growth of population.

[3]Studies for other countries find similar results like Nathan (2011) for the case of the UK, Suedekum et al. (2014) for the case of Germany or Bakens et al. (2013) for the case of Netherlands.

2001 to show that cities with more diverse populations have grown faster, but that it is diversity of country of birth rather than diversity of ethnicity that drives this effect. For the period 1991-2001, neither diversity by country of birth nor ethnic diversity are significant. Yet when variables accounting for both are included together, the cities with a large number of migrants appear to witness higher employment growth in the 1990s, but ethnically diverse cities were less successful. Recently, Lee (2014) has compared the effect of diversity at the level of the firms and city diversity. He finds a positive effect of diversity in firms owners on innovation but no relationship between the fractionalization by country of birth, at the county level, and firm level innovation. Using a novel approach Alesina, Harnoss and Rapoport (2015) finds that diversity of immigration relates positively to measures of economic prosperity.

Therefore, the literature has found that diversity seems to be negative, or irrelevant, for development at high levels of aggregation, but positive at city level. The growing availability of geographically detailed information on economic variables, and the recent focus on small units of analysis, make the issue of aggregations very relevant. We argue that the answer is different depending on the size of the unit of analysis. In this paper we present a systematic examination of the effect of ethnic diversity on economic growth using units of increasing size up to the country level. This level of detailed geographical analysis is becoming increasingly popular in the analysis of ethnicity and institutions (Michalopoulos and Papaioannou 2014a, 2014b, 2014c), ethnicity and inequality (Alesina, Michalopoulos and Papaioannou 2015) and the origins of ethnolinguistic differences (Michalopoulos 2012). We find that using a very high resolution (cells of one degree by one degree) there is a positive relationship between ethnic diversity and economic growth.

Using this small areas we are able to analyze the issue of heterogeneity and growth without considering the issue of boundaries. Recently several papers have used boundaries to generate quasi-experiments to analyze the effect of national institutions on subnational development (Michalopoulos and Papaioannou 2014a) or the effect of partitioned ethnicities on conflict. Using the shape of the border to measure the artificiality of the boundaries Alesina et al. (2011) show that the partition of ethnic groups is a significant determinant of GDP per capita.

Most of the literature on the effect of diversity on economic development at the city level claims that ethnic diversity is important for innovation and productivity. These would be a potential mechanisms that explains the positive effect of ethnic heterogeneity on development. In this paper we explore a different mechanism: the increase in trade due to the specialization of ethnic groups. The issue of the impact of spatial ethnic heterogeneity on intra-national trade is an underdeveloped topic of research[4]. Several papers have analyzed mechanisms that can support trade among agents that belong to different groups. Glaeser (2005) argues that the demand of hatred depend the cost and benefits of information about the out-group. Trade can deter the spread of hatred because it creates the incentives to demand for correct information and reduces the cost of acquiring it. Greif (2000, 2006) argues that exchange tends to be personal and supported by reputation in the early stages of development. In addition, if there is lack of trust among traders of different groups there is need for a monitoring mechanism that allows for trade. Geographical closeness simplifies monitoring activities. Jha (2013)shows that riots in South Asia's medieval ports were less likely, event many years after the colonization, despite being ethnically very diverse. Jha (2013) argues that this inertia was caused by the development of institutions that supported inter-ethnic exchange.

# 3 Data

Our unit of observation is a grid-country cell. We construct a grid of one by one decimal degree, and we calculate the value of the explanatory variables and the outcome for each of this cells. We perform the analysis using grids that generate increasingly larger units of observation (cells), starting from half by half decimal degree cells and moving towards larger sizes (two by two degrees cells and four degrees by four degree cells). Finally, we increase the size of our cells to the limit of the countries to analyze the results using cross-country regressions. The basic variables for the specification are a measure of local development and ethnic diversity[5].

[4]Aker et al. (2014) show large price changes between markets immediately across national borders. They argue that this changes are lower when markets on either side of the border share a common ethnicity.

[5]For the full description of the variables included in the empirical analysis see the Data Appendix at the end of the paper.

## 3.1 Local growth

To measure the growth in each country cell we need information on economic development. At such level 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 is becoming increasingly popular for working with small geographical areas, to use satellite light density at night as a proxy for local economic activity. The satellite night light density data are available from the National Oceanic and Atmospheric Administration. These data has been used recently by Henderson, Storeygard and Weil (2012), Rohner, Thoenig and Zilibotti (2011), Michalopoulos (2012), Michalopoulos and Papaioannou (2013) or Alesina, Michalopoulos and Papaioannou 2015). There are a series of papers that 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. Chen and Nordhaus (2011) 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 they show that the weight is, in those 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 high-GDP density for any level of quality of the statistical system.[6]

All data collection is obtained from the National Geophysical Data Center, specifically the Earth Observation Group (EOG) reference to the version 4 DMSP-OLS Night-time Lights Time Series. This first study is referenced at satellite F10 - year 1992. From three image types available hawse use the stable light version which ranges from 1 to 63 values. 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 have construct measures of luminosity per capita. Our basic dependent variable is the per capita economic growth between 1992 and 2010.

[6] The cross-validation analysis in Michalopoulos and Papaioannou (2013) shows that the light night density has a high correlation with a wealth index across household in four large African countries.

## 3.2 Spatial Ethnic Diversity

We use data from GREG for the geospatial location of ethnic groups (Weidman, Rod, and Cederan, 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 a global coverage and consists of 929 groups.[7]

For each country cell we construct two diversity type of measures. For the first measure we use the percentage of land that the homeland of the ethnic group occupies in a particular cell. The second measure uses the percentage of the population living in the homeland of the ethnic group of a particular cell. We use the traditional fractionalization measure (Herfindhal index). Since data on population that lives in the particular 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,

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

where $\pi$ 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 country will not belong to the same ethnolinguistic group[8].

# 4 Results

The basic specification is

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

[7] Desmet, Ortuño-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 the computational size of the exercises in this paper recommends to let the discussion on the degree of ethnic aggregation for future research.

[8] The Data Appendix describes in detail the source of the variables used in the empirical exercises.

where $i$ and $j$ refer to a cell and a country respectively, and $y_{ijt}$ and $y_{ij0}$ are night light density per capita in 2010 and 1992 respectively. FRAC is the level of ethnic fractionalization at each country-cell. Using this arbitrary geographical area we reduce the concerns about the possible endogeneity of the political boundaries confronted by 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 fixed effects in most of the regression. This is another advantage with respect to cross-country regressions. In fact, when we increase the size of this country-cells we reach the size of each country and, therefore, we can analyze the impact of the boundary versus the effect of the size of the cell.

Table 1 shows the basic results. Columns 1 to 3 include all the observations.[9] Column 1 presents the outcome of the regression without geographical and climate controls, and without country fixed effects. The estimation shows a positive relationship between ethnic fractionalization and growth controlling for the initial level of development[10]. Column 2 includes country fixed effects. In this case we find also a positive correlation of ethnic fractionalization on economic growth. This result is still found when we use robust standard errors clustered at country level (Column 3).[11] Columns 4 to 6 reproduce the previous regressions but using only the cells that are populated. Column 4 does not include country fixed effects. Column 5 includes country fixed effects and column 6 present the results using a robust standard error clustered at country level[12].The positive association between ethnic fractionalization and growth is robust to using the sample of the populated areas. These results indicate that there is a positive correlation between local ethnic diversity and local growth at a very high degree of 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 points.

[9]For cell that are not populated we use the transformation in Michalopoulos and Papaioannou (2014). See Data Appendix.

[10]In the regressions of table 1 the speed of convergence implied by the coefficient of the log of the initial level of development ranges from 1.6% to 2.4% which is very similar to the typical value found for the speed of convergence across regions or countries. This result indicates that night light density generates similar results to the ones found with other indicators of economic development.

[11]The conclusion is unaffected if we also use a correction for spatial correlation.

[12]The standard errors corrected for spatial correlation using Conley's approach are shown between parenthesis.

## 4.1 Controlling for observable variables

Table 2 shows the results of the basic regression when we add many additional controls.[13] Column 1 includes controls for geographic and climate controls to the specification in the last row of Table 1. The estimated coefficient implies that an increase in two standard deviations of ethnic heterogeneity increases average annual per capita growth in 1.3 percentage points. Column 2 adds to the list of explanatory variables the (logged) population density. Ethnic fractionalization continues having a positive and significant effect on growth.[14] In column 3 we show that the results are not driven by the proportion of fertile soil or natural resources. Ethnic fractionalization still has a negative and significant effect on economic growth. Column 4 adds the distance to the closest river and the distance to the closest lake to the set of explanatory variables without affecting the basic message of the previous columns. Column 5 includes a dummy variable for country cells that contain part of the border of a country[15]. The basic result is supported also in this specification.

## 4.2 The influence of the grid generating coordinates

Did we get a lucky grid? It is unlikely that any initial coordinates could produce a lucky grid since it generates a large number of cells. However, in this section we consider this possibility and produce 100 grids with random initial coordinates. Basically, we take our initial coordinates (longitude -180; latitude -89) and add to both a random number generated by two uniform distributions.[16] Figure 1, Definition of Grid, locates the random origins of the one hundred grids generated by the random coordinates.

Figure 2 (Random initial coordinates) shows the results of running the basic regression (column 1 of Table 2) one hundred times using cells generated by grids with the initial random coordinates shown in Figure 1. Figure 2 shows that all the parameter estimates are statistically significant no matter

[13]See Data Appendix for a detailed description of each of the variables.

[14]Population refers to the year 1990. Results are also robust to the inclusion of the log of population and area size.

[15]Column 5 includes between parenthesis the standard error corrected by spatial correlation following Conley (1999). The results are unaffected by this correction.

[16]See Data Appendix for a detailed explanation of the process of generation of those grids.

the initial coordinates of the grid. In addition, the estimates move mostly in a close range between 0.7 and 0.9.

## 4.3 Addressing urban agglomeration

The positive relationship between ethnic diversity and local growth could be capturing just 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 just the urban premium what drives our results? Table 3 addresses this issue. In columns 1 we include a dummy for the national capital while in column 2 we add also dummies for provincial capital. Column 3 adds dummies for urban agglomerations, considering as such urban areas that concentrate more than 500.000 inhabitants. In all the cases the basic results of tables 1 and 2 are maintained: ethnic diversity has a positive and significant effect on growth.

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 5 indicate that the relationship between ethnic diversity and local growth are not driven by the agglomeration effect tied only to the presence of urban areas.

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

[17] See Data Appendix.

robust to the elimination of urban centers from the sample.

We have ran many other robustness analysis.[18] 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 results are also robust to the use of the level of night light per capita instead of the growth rate. However, the use of the growth rate has some advantages since we can control for some omitted variables by using the initial level of development.[19] The empirical findings are basically unaffected if we use other sources of ethnic diversity like Ethnologue, or we use Murdock's data in the analysis of the African case, or we include other measures of diversity as the number of ethnic groups. The results are also robust to using the level interaction between ethnicity and the degree of decentralization of the countries or the intereaction of ethnicity and its institutional development.

We have also run some instrumental variables estimations (Table 5). As we mentioned at the beggining, the purpose of the paper is not to find causality but to notice some interesting facts about ethnic diversity and the size of the areas of reference. We do not want to emphasize this IV strategy since, as it is common with instrumental variables' techniques, it is difficult in general to find good instruments. We have consider two potential instruments for ethnic fractionalization in Africa: the ethnic fractionalization derived from Murdock's 1800 mapping of ethnic groups and the degree of ethnic diversity of the contiguous cells (within 200 km). In both cases we find a positive and statistically significant effect of ethnic diversity.

# 5 Changing the resolution

The literature on the influence of diversity on development has analyzed different units of observation, from countries to cities. However, in all those cases, the observations were subject to political or administrative boundaries that could compromise the exogeneity of the country/county/city degree of diversity. In this section we show the effect of a progressive change of the size of the units of observation on the relationship between ethnic diversity

[18]See Table 4.

[19]In addition it is easy to show that a simple model of growth with a measure of the variety of goods depending on diversity generates a growth rate that is funcion of ethnic fractionalization

and growth.[20] Figure 3 shows that for cells of small size there is a positive and statistically significant effect of ethnic diversity on growth, as we have seen in previous sections. However, increasing the size of the observations reduce the effect and turns it statistically insignificant.

Table 6 shows that, at the limit (country level), the effect of ethnic diversity on growth is not statistically significant. This result is consistent with previous research using cross country data. In column 1 we use our measure of ethnic fractionalization constructed using GREG, the measures of economic growth constructed using night light data and all the controls (corresponding to column 5 in Table 2). In column 2 we use the measures of ethnic fractionalization constructed using the population that lives in ethnic homeland, and changes in night light data to proxy growth. In column 3 we use data from the PWT 7.1 instead of night light. And finally column 4 we use the traditional data used in country growth regression: data from Atlas Narodov Mira to construct fractionalization measure and data from PWT 7.1 to construct growth measures. In all cases ethnic diversity has no effect on economic growth.[21]

Finally we analyze what happen when we look at a higher resolution. Increasing resolution below 0.5 by 0.5 degrees could be problematic if we use indices of fractionalization because very small pixels close to ethnic borders would appear homogenous when they are very close to an ethnic border. In order to analyze the effect of diversity on economic growth at very high resolution level we perform a pixel level analysis. We generate 5km x 5km (0.045 by 0.045 degrees) grid cells spanning from -180 degrees longitudes and -89 degrees latitude. For each cell of this size we construct measures of economic growth in the same way as we did for larger cells. We then compute the distance of each centroid of the cell to the closest ethnic border inside the country. Figure 4 shows that cells that are closer to ethnic borders grow faster than cells far away. For the regression analysis we construct a dummy that has value 1 if the distance of the cell to the closest ethnic border is less that 5 km, and zero otherwise. Our data are cells that are inside a buffer of 50km on each side of the ethnic border. This implies certain similarity in the characteristics of the cell inside the buffer. Table 7 show the regressions of Table 1 using cells at a very high resolution, and a buffer of 50km each side

[20]We use the baseline specification of column 1 of Table 2.

[21]This is additional evidence of the usefulness of nighlight for the type of exercises that we are performing.

of the ethnic border. The results are in line with our previous findings, and indicate that cells very close to ethnic border have a higher economic growth than cells far away. In table 8 we use the specification of the last column of Table 7 but changing the definition of buffer, and also the definition of the dummy that capture "being close to an ethnic border". From column 1 to 4 we still have the 50 km buffer. In column 1 we use a dummy that has value 1 if the cell distance to the closest ethnic border is less than 2 km. Column 2 to 4 we construct similar dummies for less that 3km, 4km and 10km. In all cases the basic findings are unchanged. In column 5 to 7 we change the size of the buffer, that instead of 50km is 10km, 25km and 75km. In all cases the basic results are unaffected, which is that being close to an ethnic border, is correlated with more economic growth.

# 6 Interpretation: a potential mechanism

Why diversity is good for units of small size and irrelevant at larger sizes? In the second section we have 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. Following the comments in Section 2, we analyze the possibility that ethnic diversity can increase trade when observed at high resolution. Assuming that members of different ethnic groups have less trust in each other than members of the same group, trade across ethnic groups imply the need to monitor, and be able to retaliate in case of non-fulfilment of the conditions of the contracts. Therefore, trade across ethnic groups requires proximity. It is not possible to find data on trade across ethnic groups and, therefore, to provide some evidence in 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 also find evidence of ethnic specialization. We find that the variance of the proportion of different crops is positively correlated with ethnic diversity conditional on geography and climate. And this result is unaffected by the inclusion of the initial level of development. Moreover, high variability in the proportion of crops in an area is associated with high growth. Therefore, one potential story that because of the specialization of ethnic groups they can derive a large improvement on welfare by trading with other ethnic groups. This implies that if this effect is larger than the transaction cost associated with lower levels of trust or problems of communication, we should find local markets at the ethnic borders.[22]

Assuming that trade al local level usually take place in markets we look for data on markets. Porteous (2015) has gathered information on the location of 223 markets in Africa and he shows a map of Africa with the location of these local markets. The distribution of markets by the size of the population results in 60% of them located in cities of more than 100.000 inhabitants and 40% in small rural villages.

We have also overlapped Porteous (2015) map with our map of the spatial distribution of ethnic groups in Africa, Figure 5. It is easy to see that, as a first look, many of the markets are very close to ethnic border. In fact the average distance of all theses market to the closest ethnic border is just 27 km, which seems to indicates that trading markets are located close to the ethnic borders.In order to show how far is the actual distribution of markets with respect to a random geographic distribution we have run a simulation with 500 random samples of 223 locations in Africa (as many as market in Porteous 2015). We first consider continental sub-Saharan Africa, which is the area covered by Porteous (2015). We estimate the Harvesine 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 are reported in Figure 6 that shows that the average distance of 27 km is at the 1% of the distribution. This indicates that markets are much closer to ethnic borders than randomly generated locations.

Figure 7 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 simulation of markets' distribution in Sub-Saharan Africa. For the placebo analysis we randomly generate the location of 223 "virtual" markets, which is the actual number of markets identified by Porteous (2015), in each simulation and use a buffer of 50km to calculate their index of ethnic diversity. Figure 7 shows that the ethnic diversity of the actual markets is in the tail of the distribution of heteregeneity indices of the "virtual" markets. Figure 8 runs a similar exercise to Figure 7 but relating market location and growth. It shows that the growth rate around the markets is much higher than the growth rate on the "virtual" markets. Table 9 presents the regressions of the presence of a market on the degree of ethnic fractionalization. It shows that the higher is ethnic fractionalization, the higher is the probability of finding a market. In addition,

[22]Unfortunately, there is no information on trade across ethnic groups.

Table 9 shows that the existence of a market lead to a higher growth rate conditional on the initial level of development.

The previous evidences show that areas which have more ethnic diversity have also more markets. A possible explanation is based on the specialization of ethnic groups. The geographic proximity of ethnic groups may increase trade if they are highly specialized in the production of specific agricultural products or services. It is true that the level of trust among different ethnic groups is, in general, smaller than trust intragroup but the fact that they are close to each other can facilitate monitoring and, therefore, counterbalance the potencial lack of trust. Jha (2013) shows that medieval Hindus and Muslim could provide complementary services and a mechanism to share gain from trade which increased tolerance between these groups. The development of these practices into formal institutions generated inertia in the degree of ethnic tolerance.[23] The location of local markets in Africa seems to support this interpretation. The idea is that heterogeneity is good for trade and exchange, as suggested by our paper, and homogeneity is good to have better institutions as supported by most of the literature. Therefore, for a given institutional framework, diversity seems to explain why some areas are more regionally richer than others.

# 7 Concluding remarks

The relationship between ethnic heterogeneity and development is complicated. Empirical research working with cross section data finds a negative, or null, relationship. However, research at the city level finds usually a positive relationship between diversity and wages and/or productivity. In this paper we perform a systematic analysis of the effect of the size of geographical units on the relationship between ethnic diversity and growth. We find that small areas tend to generate a positive relationship while for larger areas there is no relationship. We argue that a possible mechanism to explain the positive relationship between diversity and growth is the increase trade in the boundaries across ethnic groups due to ethnic specialization. The idea is that heterogeneity is good for trade and exchange and homogeneity has a positive impact on the quality of institutions, as argued by most of the literature. Therefore our results indicate that, conditional on the institutional

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

framework, diversity is good for local development.

# References

[1] Acemoglu, D., Naidu, S., Restrepo, P., J. Robinson (2015), "Democracy does cause growth?", mimeo.

[2] Aker, J., M. Klein, S. O'Connell and M. Yang (2014), "Borders, ethnicity and trade," *Journal of Development Economics*, 107, 1-16.

[3] Alesina, A., J. Harnoss and H. Rapoport (2015), "Birthplace Diversity and Economic Prosperity," mimeo.

[4] Alesina, A., S. Michalopoulos, and E. Papaioannou (2015): "Ethnic Inequality," *Journal of Political Economy*, forthcoming.

[5] Alesina, A., A. Devleeschauwer, W. Easterly, W., S. Kurlat, and R. Wacziarg (2003), "Fractionalization," *Journal of Economic Growth*, 8, 155-194.

[6] Alesina, A., W. Easterly and J. Matuszeski (2011), "Artificial states," *Journal of the European Economic Association*, 9 (2), 246-277.

[7] Alesina, A. and E. La Ferrara (2005), "Ethnic diversity and economic performance," *Journal of Economic Literature,* 43 (3), 762-800.

[8] *Atlas Narodov Mira (Atlas of the People of the World)*. Moscow: Glavnoe Upravlenie Geodezii i Kartografii, 1964.

[9] Barret, D. ed. (1982), *World Christian Encyclopedia*. Oxford University Press.

[10] Cameron, C. A., J. Gelbach, and D. Miller (2011): "Robust Inference with Multi-Way Clustering," *Journal of Business and Economic Statistics*, 28(2), 238—249.

[11] Chen, X., and W. D. Nordhaus (2011): "Using Luminosity Data as a Proxy for Economic Statistics," *Proceeedings of the National Academy of Sciences*, 108(21), 8589—8594.

[12] Desmet K, Ortuño-Ortin I, Weber S (2009), “Linguistic Diversity and Redistribution, ”*Journal of the European economic Association*, 7 (6), 1291-1318.

[13] Desmet, K., Ortuño-Ortin I. and R. Wacziarg (2012), ”The Political Economy of Linguistic Cleavages,” *Journal of Development Economics*, 97, 322-338.

[14] Dincer O.C., and F. Wang (2011), “Ethnic diversity and economic growth in China, ”*Journal of Economic Policy Reform*, 14 (1), 1-10. ELIMINAR

[15] Easterly, W., and Levine (1997), “Africa’s growth tragedy: Policies and Ethnic divisions, ” *Quarterly Journal of Economics*, 112 (4), 1203-1250.

[16] *Encyclopedia Britannica* (2000). Chicago: Encyclopedia Britannica.

[17] Fafchamps, M. (2000), “Ethnicity and credit in African manufacturing,” *Journal of Development Economics* 61, 205-235

[18] Gören, E. (2014), ”How Ethnic Diversity Affects Economic Growth,” *World Development*, 59, 275-297.

[19] Glaeser, E. (2005), ”The political economy of hatred,” *Quarterly Journal of Economics*, 45-86.

[20] Greif, A. (2000), ”The fundamental problem of exchange: A research agenda in Historical Insitutional Analysis, *European Review of Economic History*, 4, 251-284.

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

[22] Jha, S. (2013), ”Trade, Institutions and Ethnic Tolerance: Evidence from South Asia,” A*American Political Science Review*, 107 (4), 806-832.

[23] Lee, N. (2014), ”Migrants and ethnic diversity, cities and innovation: Firm effect or city effects?,” *Journal of Economic Geography*, 1-28.

[24] Lee, N. (2011), "Ethnic diversity and employment growth in English cities," *Urban Studies*, 48, 407-425.

[25] Michalopoulos, S., and E. Papaioannou (2013), "Pre-colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 81(1), 113—152.

[26] Michalopoulos, S., and E. Papaioannou (2014), "National Institutions and Subnational Development in Africa," *Quarterly Journal of Economics*, 129(1), 151—213.

[27] Montalvo, J. G. and M Reynal-Querol (2005a), "Ethnic polarization, potential conflict and civil wars," *American Economic Review*, 95 (3), 796-816.

[28] Montalvo, J. G. and M Reynal-Querol (2005b), "Ethnic Diversity and Economic Development," *Journal of Development Economics*, 76, 293-323.

[29] Ottaviano G. P., and G. Peri (2004), "Cities and Cultures, "*Journal of Urban Economics*, 58(2), 304-307.

[30] Ottaviano, G. P., and G. Peri (2005), "The economic value of cultural diversity: evidence from US cities, " *Journal of Economic Geography*, 6, 9–44.

[31] Ottaviano G. P. and G. Prarolo (2009), "Cultural identity and knowledge creation in cosmopolitan cities, "*Journal of Regional Science*, 49 (4), 647-662.

[32] Porteous, O. (2015), "High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Agricultural Storage and Trade," mimeo.

[33] Sparber C. (2010), "Racial Diversity and Macroeconomic Productivity across US States and Cities, ", *Regional Studies*, 44(1), 71-85.

[34] Taylor, C. and M.C. Hudson (1972), *The World Handbook of Political and Social Indicators*, 2nd ed., New Haven, CT: Yale University Press.

# Table 1: Ethnic Diversity and Growth

*Dependent Variable: Growth*

| | All Sample | All Sample | All Sample | Only Populated Areas | Only Populated Areas | Only Populated Areas |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Log night light 1992 | -0.252*** | -0.350*** | -0.350*** | -0.264*** | -0.370*** | -0.370*** |
| | [0.005] | [0.007] | [0.022] | [0.005] | [0.007] | [0.022] |
| | | | | | | (0.014) |
| Ethnic Fractionalization | 0.501*** | 0.690*** | 0.690*** | 0.425*** | 0.617*** | 0.617** |
| | [0.094] | [0.096] | [0.240] | [0.098] | [0.100] | [0.287] |
| | | | | | | (0.154) |
| Country FE | No | Yes | Yes | No | Yes | Yes |
| Observations | 25693 | 25693 | 25693 | 21514 | 21514 | 21514 |
| R-squared | 0.171 | 0.287 | 0.287 | 0.159 | 0.294 | 0.294 |

*Notes* - In columns 1,2,4 and 5 we report robust standard error in brackets. In columns 3 and 6, robust standard error in brackets clustered at country level are reported in brackets. In column 6, Conley standard errors in parenthesis (Spatial correlation kernel cutoff = 200km). * Significant at 10%, ** significant at 5%, and *** significant at 1%.

# Table 2: Ethnic Diversity and Growth: for observables

*Dependent Variable: Growth*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Log night light 1992 | -0.369*** | -0.367*** | -0.369*** | -0.369*** | -0.370*** |
| | [0.021] | [0.021] | [0.021] | [0.021] | [0.021] |
| | | | | | (0.013) |
| Ethnic Fractionalization | 0.720*** | 0.733*** | 0.725*** | 0.700*** | 0.735*** |
| | [0.258] | [0.243] | [0.239] | [0.247] | [0.244] |
| | | | | | (0.149) |
| Geographic Variables | Yes | Yes | Yes | Yes | Yes |
| Climate Variables | Yes | Yes | Yes | Yes | Yes |
| Population Density | No | Yes | Yes | Yes | Yes |
| Share Mining and Fertile Soil | No | No | Yes | Yes | Yes |
| Distance to River and Lake | No | No | No | Yes | Yes |
| Border | No | No | No | No | Yes |
| Observations | 21514 | 21514 | 21514 | 21514 | 21514 |
| R-squared | 0.298 | 0.298 | 0.299 | 0.299 | 0.301 |

*Notes* - Robust standard error clustered at country level are reported in brackets. In column 6, 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 and Ruggedness Index. Climate Controls: Average temperature from 1961-1980 and average precipitation from 1961-1980. Population Density = ln(Total Population in 1990/ Area in $Km^2$). Share Mining and Fertile Soil:Share Mining, % Fertile Soil. Distance to River and Lake: Distance to River (km) and Distance to Lake (km). Border= Border (yes=1)

# Table 3: Ethnic Diversity and Growth: Robustness to Agglomeration Effect

*Dependent Variable: Growth*

| | (1) | (2) | (3) | Drop all-urban center (4) | Drop 10% richest (5) | Drop 20% richest (6) | Drop 10% most densely (7) | Drop 20% most densely (8) |
|---|---|---|---|---|---|---|---|---|
| Log night light 1992 | -0.370*** | -0.375*** | -0.375*** | -0.367*** | -0.373*** | -0.374*** | -0.375*** | -0.359*** |
| | [0.021] | [0.021] | [0.021] | [0.022] | [0.015] | [0.016] | [0.022] | [0.023] |
| Ethnic Fractionalization | 0.736*** | 0.697*** | 0.699*** | 0.730** | 0.710*** | 0.924*** | 0.697** | 0.561* |
| | [0.245] | [0.242] | [0.246] | [0.291] | [0.269] | [0.183] | [0.275] | [0.291] |
| Nat. Capital (yes=1) | -0.211 | -0.516*** | -0.504*** | | -0.545*** | -0.589*** | -0.644*** | -1.409*** |
| | [0.140] | [0.116] | [0.088] | | [0.081] | [0.078] | [0.215] | [0.316] |
| Prov. Capital (yes=1) | | 0.512*** | 0.524*** | | 0.545*** | 0.533*** | 1.044*** | 1.416*** |
| | | [0.130] | [0.127] | | [0.122] | [0.122] | [0.179] | [0.212] |
| Urb. Agglom (yes=1) | | | -0.043 | | -0.218 | -0.467*** | | |
| | | | [0.180] | | [0.163] | [0.125] | | |
| Controls from Table 2 | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 21,514 | 21,514 | 21,514 | 18,174 | 19,151 | 17,694 | 18,674 | 16,091 |
| R-squared | 0.301 | 0.303 | 0.303 | 0.293 | 0.292 | 0.307 | 0.297 | 0.286 |

*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 C6 Table 2 include: distance to coastline, Ruggedness Index, Average temperature from 1961-1980 , average precipitation from 1961-1980,Log Population Density, Share Mining, % Fertile Soil, Distance to River, Distance to Lake and Border (yes=1).

# Table 4: Ethnic Diversity and Growth: Robustness to Alternative Data on Ethnicity

*Dependent Variable: Growth*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Log night light 1992 | -0.382*** | -0.382*** | -0.384*** | -0.384*** |
| | [0.022] | [0.022] | [0.034] | [0.033] |
| Ethnic Fractionalization (Ethnologue) | 0.674*** | 0.686*** | | |
| | [0.159] | [0.155] | | |
| Ethnic Fractionalization 1800 (Murdock) | | | 1.091*** | 0.914*** |
| | | | [0.258] | [0.233] |
| Geographic Variables | Yes | Yes | Yes | Yes |
| Climate Variables | Yes | Yes | Yes | Yes |
| Population Density | No | Yes | No | Yes |
| Share Mining and Fertile Soil | No | Yes | No | Yes |
| Distance to River and Lake | No | Yes | No | Yes |
| Border | No | Yes | No | No |
| Observations | 19822 | 19822 | 3654 | 3654 |
| R-squared | 0.307 | 0.310 | 0.261 | 0.269 |

*Notes* - Robust standard error clustered at country level are reported in brackets. In column 6, 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 and Ruggedness Index. Climate Controls: Average temperature from 1961-1980 and average precipitation from 1961-1980. Population Density = ln(Total Population in 1990/ Area in $Km^2$). Share Mining and Fertile Soil:Share Mining, % Fertile Soil. Distance to River and Lake: Distance to River (km) and Distance to Lake (km). Border= Border (yes=1)

# Table 5: Ethnic Diversity and Growth: IV approach for Africa.

*Dependent Variable: Growth*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Log night light 1992 | -0.422*** | -0.422*** | -0.421*** | -0.421*** |
| | [0.042] | [0.042] | [0.042] | [0.042] |
| Ethnic Fractionalization | 2.335** | 2.382** | 2.373*** | 2.402* |
| | [1.044] | [1.166] | [0.913] | [1.325] |
| Ethnic Fractionalization 1800 (Murdock) | | | | -0.019 |
| | | | | [0.460] |
| Controls from Table 2 | Yes | Yes | Yes | Yes |
| First Stage. Dept. Var=Ethnic Fractionalization | | | | |
| Sample | Africa | Africa | Africa | Africa |
| Ethnic Fractionalization 1800 (Murdock) | 0.285*** | | 0.227*** | 0.227*** |
| | [0.038] | | [0.035] | [0.035] |
| Ethnic Fractionalization 200km | | 0.897*** | 0.802*** | 0.802*** |
| | | [0.047] | [0.054] | [0.054] |
| Observations | 3040 | 3050 | 3040 | 3040 |
| Partial-R2 | 0.270 | 0.270 | 0.270 | 0.269 |

*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 C6 Table 2 include: distance to coastline, Ruggedness Index, Average temperature from 1961-1980 , average precipitation from 1961-1980, Log Population Density, Share Mining, % Fertile Soil, Distance to River, Distance to Lake and Border (yes=1).

# Table 6: Ethnic diversity and growth: country level results

*Dependent Variable: Growth*

| | Growth (1) | Growth (2) | Growth PWT7.1 (3) | Growth PWT7.1 (4) |
|---|---|---|---|---|
| Ethnic Fractionalization | -0.117 | | 0.000 | |
| | [0.175] | | [0.134] | |
| Ethnic Fractionalization POP | | -0.111 | | |
| | | [0.182] | | |
| Ethnic Frac. based on Soviet Atlas | | | | 0.128 |
| | | | | [0.113] |
| Controls from Table 2$^{+}$ | Yes | Yes | Yes | Yes |
| Observations | 233 | 233 | 167 | 139 |
| R-squared | 0.232 | 0.232 | 0.138 | 0.200 |

*Notes* - Robust standard error clustered at country level are reported in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%. We include initial level of growth and GDP in columns 1-2 and 3-4, respectively. Controls from C6 Table 2 *** include: average distance to coastline, average Ruggedness Index, Distance to Average temperature from 1961-1980 , average precipitation from 1961-1980, Log Population Density, Share Mining, % Fertile Soil, average Distance to River and average Distance to Lake.

# Table 7: Ethnic Diversity and Growth, pixel analysis (5kmx5km)

*Dependent Variable: Growth*

| | All Sample | All Sample | All Sample | Only Populated Areas | Only Populated Areas |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Log night light 1992 | -0.108*** | -0.230*** | -0.230*** | -0.256*** | -0.256*** |
| | [0.000] | [0.001] | [0.017] | [0.001] | [0.016] |
| Closest GREG Border within Country (nocoastline) $<= 5km$ (yes=1) | 0.125*** | 0.042*** | 0.042** | 0.062*** | 0.062*** |
| | [0.004] | [0.004] | [0.020] | [0.004] | [0.021] |
| Country FE | No | Yes | Yes | Yes | Yes |
| Observations | 2,780,089 | 2,780,089 | 2,780,089 | 2,314,029 | 2,314,029 |
| R-squared | 0.034 | 0.203 | 0.203 | 0.209 | 0.209 |

*Notes* - In columns 1 and 2 we report robust standard error in brackets. In columns 3, robust standard error in brackets clustered at country level are reported in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

# Table 8: Ethnic Diversity and Growth, 5kmx5km Analysis

*Dependent Variable: Growth*

| | 50km Buffer | | | | | 10km Buffer | 25km Buffer | 75km Buffer |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log night light 1992 | -0.256*** | -0.256*** | -0.256*** | -0.256*** | -0.256*** | -0.265*** | -0.256*** | -0.255*** |
| | [0.016] | [0.016] | [0.016] | [0.016] | [0.016] | [0.013] | [0.015] | [0.016] |
| Closest GREG Border within Country (nocoastline) <= 1*km* (yes=1) | 0.098*** | | | | | | | |
| | [0.023] | | | | | | | |
| Closest GREG Border within Country (nocoastline) <= 2*km* (yes=1) | | 0.081*** | | | | | | |
| | | [0.021] | | | | | | |
| Closest GREG Border within Country (nocoastline) <= 3*km* (yes=1) | | | 0.072*** | | | | | |
| | | | [0.021] | | | | | |
| Closest GREG Border within Country (nocoastline) <= 4*km* (yes=1) | | | | 0.069*** | | | | |
| | | | | [0.020] | | | | |
| Closest GREG Border within Country (nocoastline) <= 10*km* (yes=1) | | | | | 0.039** | | | |
| | | | | | [0.019] | | | |
| Closest GREG Border within Country (nocoastline) <= 5*km* (yes=1) | | | | | | 0.035*** | 0.052*** | 0.055** |
| | | | | | | [0.010] | [0.016] | [0.024] |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,314,029 | 2,314,029 | 2,314,029 | 2,314,029 | 2,314,029 | 673,235 | 1,501,718 | 2,801,968 |
| R-squared | 0.208 | 0.208 | 0.208 | 0.209 | 0.208 | 0.218 | 0.211 | 0.207 |

*Notes* - In columns 1 and 2 we report robust standard error in brackets. In columns 3, robust standard error in brackets clustered at country level are reported in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%.

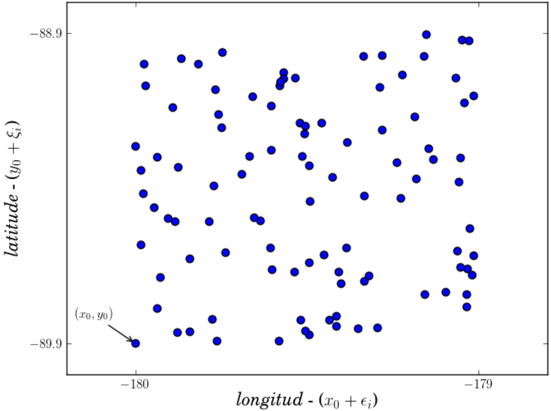
# Table 9: Ethnic Diversity and Growth: Market Analysis

*Dependent Variable: Growth*

| | Dep. Var: Presence of Market (yes=1) | Dep. Var: Growth |
|---|---|---|
| | (1) | (2) |
| Log night light 1992 | | -0.459*** |
| | | [0.025] |
| Ethnic Fractionalization | 0.066*** | |
| | [0.023] | |
| Presence of Market (yes=1) | | 1.336*** |
| | | [0.327] |
| Controls from Table 2 | Yes | Yes |
| Observations | 3103 | 3103 |
| R-squared | 0.095 | 0.307 |

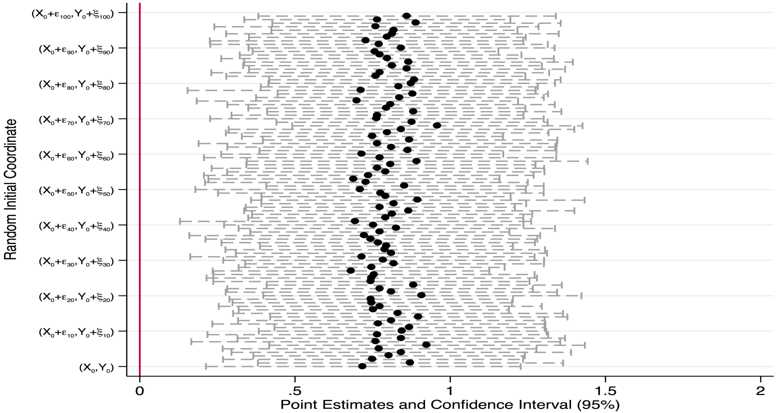
*Notes* - Robust standard error clustered at country level are reported in brackets. * Significant at 10%, ** significant at 5%, and *** significant at 1%. Country fixed effects are included. Controls from C6 Table 2 include: distance to coastline, Ruggedness Index, Average temperature from 1961-1980 , average precipitation from 1961-1980, Log Population Density, Share Mining, % Fertile Soil, Distance to River, Distance to Lake and Border (yes=1).

# Figure 1: Definition of Grid - Simulation



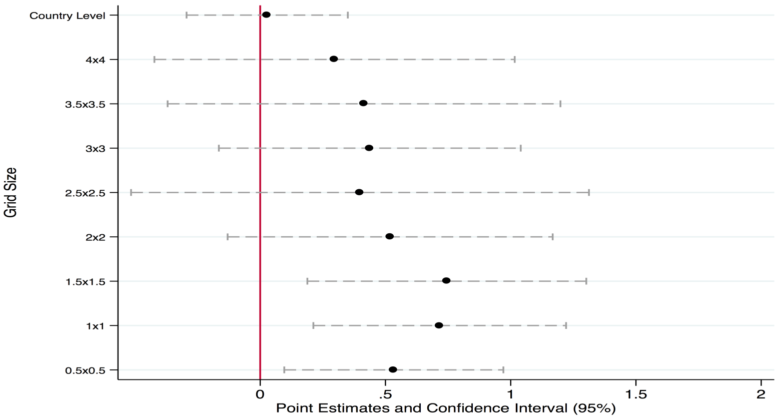

*Notes* - This graph shows the random initial coordinates generated. Our baseline starting points are: $X_0(longitude) = -180$ and $Y_0(latitude) = -89$. $\epsilon_i, \xi_i \sim Uniform(0, 1), \forall(i = 1, \ldots, 100)$

# Figure 2: Random Initial Coordinates



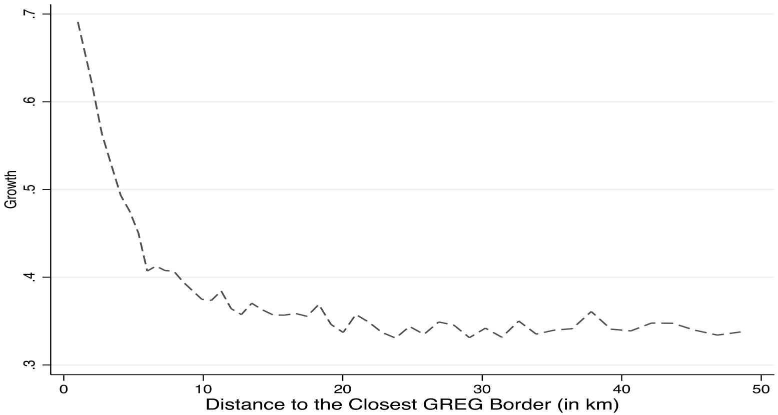

*Notes* - This graph shows the point estimate for *Ethnic Fractionalization* and its Confidence Interval for 100 simulated grids. We estimate the specification of Table 2 column 1. Our baseline starting points are: $X_0(longitude) = -180$ and $Y_0(latitude) = -89$. $\epsilon_i, \xi_i \sim Uniform(0, 1), \forall (i = 1, \ldots, 100)$

# Figure 3: Grid Size Analysis



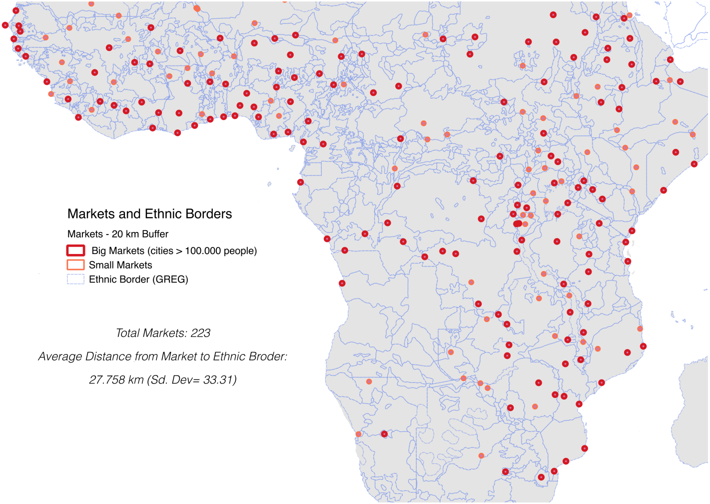

*Notes* - This graph shows the point estimate for *Ethnic Fractionalization* and its Confidence Interval for the different Grid Size. We estimate the specification of Table 2 column 1.

# Figure 4: Grid Size Analysis



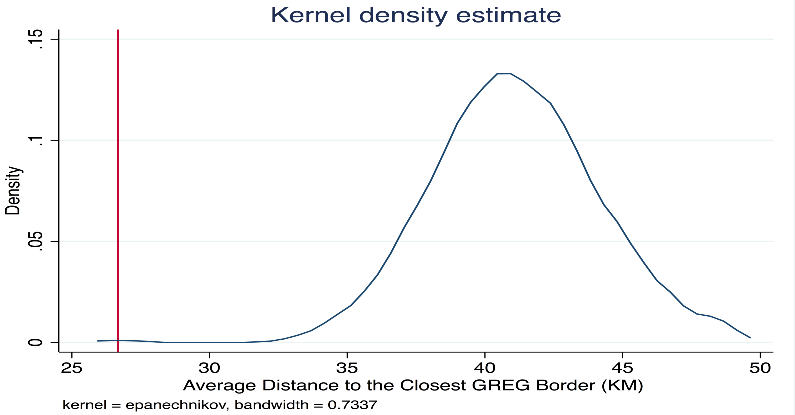

*Notes* - This graph shows the distance to the GREG border (in km) of the 5kmx5km grid. We estimate the Harvesine Distance.

# Figure 5: Market Location and Ethnic Borders



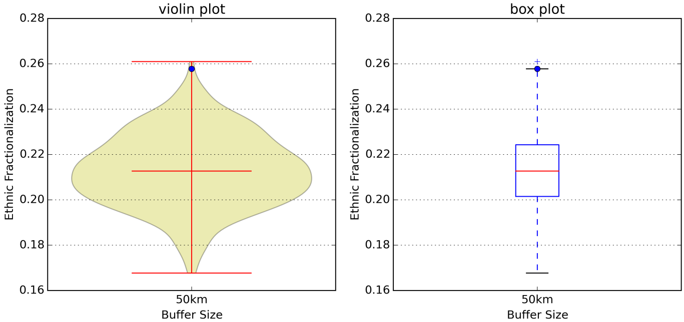

*Notes* - This graph shows the location of 223 markets in Sub-Saharan Africa identified by Porteous (2015). Markets are defined as: (i) towns and cities which have a population of at least 100,000 people and are at least 200 kilometers apart (if two towns with over 100,000 people are closer than 200 kilometers I include the larger one). (ii) smaller towns that are located at important road junctions or ports. (iii) major towns (some closer than 200 kilometers apart) in countries which still have high population-to-market ratios after my ratios two steps.

# Figure 6: Market Location - Simulation



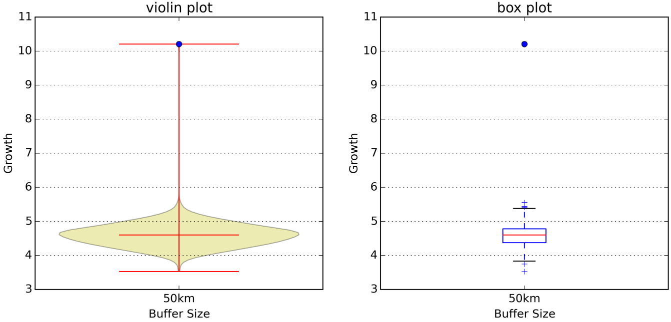

*Notes* - This graph shows the average distance to the GREG border using the actual location and 500 simulated location market in Sub-Saharan Africa. In each simulation, we randomly generated the location of 224 markets, which are the number of actual number markets located in Africa by Porteous (2015). In all cases, we estimate the Harvesine Distance. On average, whereas the actual location of markets are located at 26km to to closest GREG border (located at 1 percentile), the simulated markets are 41 KM.

# Figure 7: Market Location and Ethnic Diversity - Simulation




*Notes* - This graph shows the average Ethnic Diversification Index for the actual location and 500 simulated location market in Sub-Saharan Africa. In each simulation, we randomly generated the location of 224 markets, which are the number of actual number markets located in Africa by Porteous (2015). Then, we create a 50km buffer around each market (or simulated market) and estimate the Ethnic Diversity Index. The overall Ethnic Diversity Index mean is 0.212 (actual market = .237 - located at 100 percentile)

# Figure 8: Market Location and Growth - Simulation

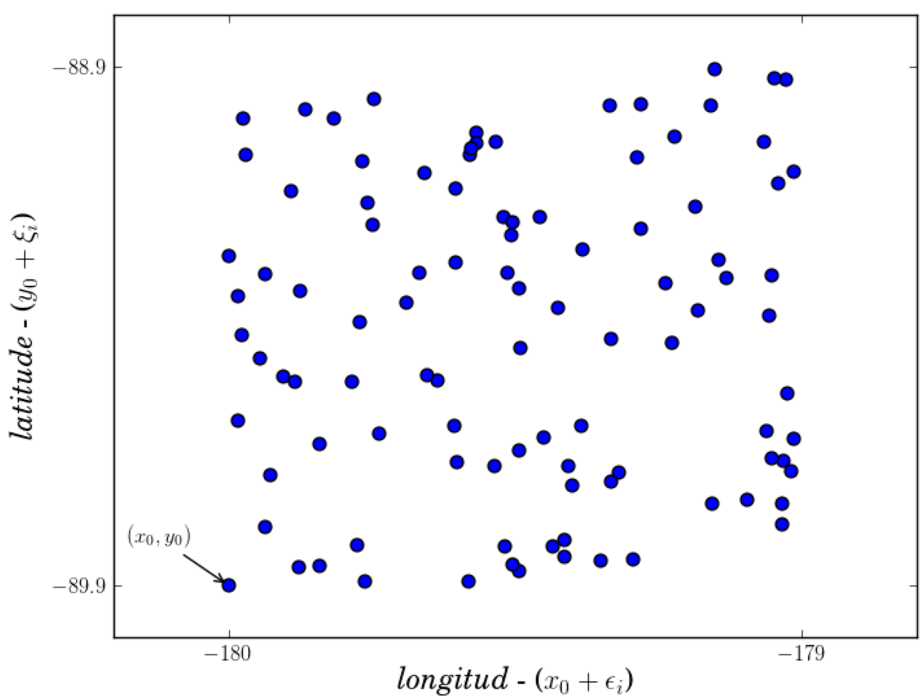



*Notes* - This graph shows the average Growth for the actual location and 500 simulated location markets in Sub-Saharan Africa. In each simulation, we randomly generated the location of 224 markets, which are the number of actual number markets located in Africa by Porteous (2015). Then, we create a 50km buffer around each market (or simulated market) and Growth. The overall Growth mean is 4.59 (actual market = 10.20 - located at 100 percentile)s

# 1 Grid Simulation

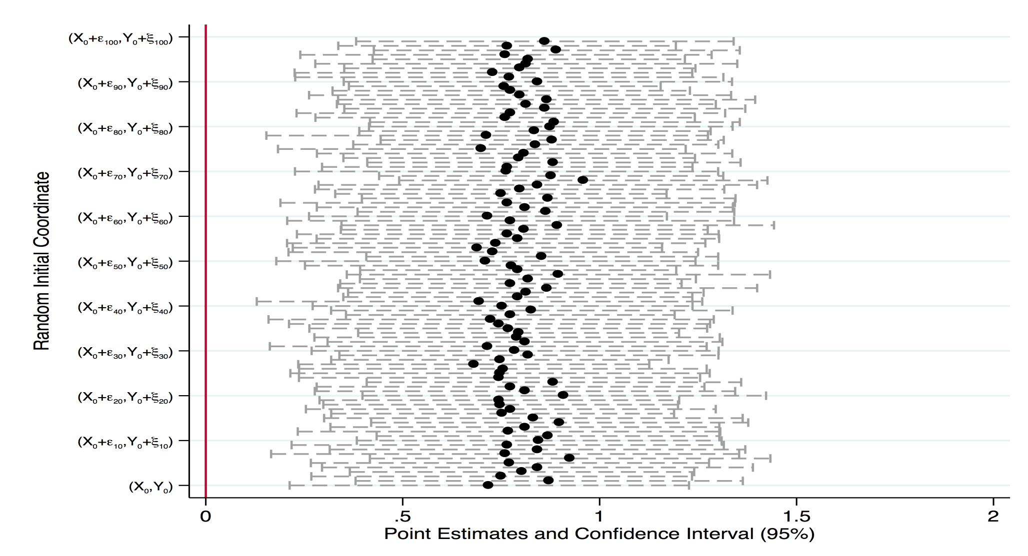
In order to determinate whether our results are driven by the geographical definition of our cell, we randomly generate 100 different cells to replicate our results. Based on our baseline starting coordinates (i.e. $X_0(longitude) = -180$,$Y_0(latitude) = -89$), we generate 100 new starting coordinate in the following manner: $(X_0 + \epsilon_i, Y_0 + \xi_i$, where, $\epsilon_i, \xi_i \sim Uniform(0,1), \forall(i = 1, ..., 100))$. As a result, we could have 100 different cells for all across the world (see Figure 1).

Figure 1: Random initial coordinate




Once we simulate the new random grids and intersect them with the political borders, we rebuild our data set for each new cell definition. Graph 2 shows the point estimate for *Ethnic Fractionalization* and its Confidence Interval for 100 simulated grids. We estimate the specification of Table 2 column 1. In general, point estimates are always significant and have the same magnitude. Therefore, we can conclude that our results are robust to the initial coordinates of the grid that produced the hundred-cells.

Figure 2: Point Estimates for the different random initial coordinate




*Notes* - T Our baseline starting points are: $X_0(longitude) = -180$ and $Y_0(latitude) = -89$. $\epsilon_i, \alpha_i \sim Uniform(0,1), \forall(i = 1, ..., 100)$

# 2 Data Description

## 2.1 Baseline OBS

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

[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.

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.

## 2.2 Economy Growth, 1992-2010

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:

$$Growth_{1992-2010} = ln\left(\frac{0.1 + nighlight_{2010}}{0.1 + population_{2005}}\right) - ln\left(\frac{0.1 + nighlight_{1992}}{0.1 + population_{1990}}\right) \tag{1}$$

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

[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).

## 2.3 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:

$$FRAC = 1 - \sum_{i=1}^{N} \pi_i^2 = \sum_{i=1}^{N} \pi \left(1 - \pi\right) \qquad (2)$$

where $\pi$ is the proportion of area/population who belong to ethnic group $i$ in a given ($\forall i = 1, ..., N$).

## 2.4 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) [8]. It provides a detailed information on the monthly average precipitation (mm/month) and the temperature (Degrees Celsius) from 1961 to 1990 [9]. We first take the overall average of the monthly information at 10-minute latitude/longitude, and then average at our cell level.

## 2.5 Share Mining

We use the Seamless Digital Chart of the World (SDCW)[10], which provides a unique information on areas where natural resources are being extracted from the earth[11]. The SDCW is based on the best currently-available global

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

[8] 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

[9]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`).

[10]`www.worldgeodatasets.com/basemaps/`

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

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:

$$ShareMining_i = \frac{\textit{Area Natural Resources Extraction (km2)}_i}{\textit{Total Area (km2)}_i} \tag{3}$$

## 2.6 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[12]. Thereby, we estimate the overall mean of the fertile cell at our cells level.

## 2.7 Urban Centers, Rivers and Lakes

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 Capital (km):** Euclidean distance from the centroid of our cell to the political capital of the country.

[12]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.

**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.

**National Capital (yes=1):** It indicates whether the political national capital is located within a giving cell.

**Provincial Capital (yes=1):** It indicates whether a provincial capital is located within a giving cell.

## 2.8 Terrain Ruggedness Index, 100m

We use the Terrain Ruggedness Index used by Nunn and Puga (2012)[13]. Each cells on a 30 arc-seconds grid across the surface of the Earth provides the Terrain Ruggedness Index, in millimeters[14]. Then, we estimate the weighted average, using as weights the area of each cell[15], 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.

[13]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

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

[15]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).