Review of
ECONOMICS
and
INSTITUTIONS

Review of Economics and Institutions
ISSN 2038-1379 DOI 10.5202/rei.v5i2.164

Vol. 5 – No. 2, Spring-Fall 2014 – Article 5

www.rei.unipg.it

Historical Ethnic Homelands and Income Convergence in Africa

Dimitris K. Christopoulos
Panteion University

Angelos Mimis
Panteion University

Gregorios Siourounis✉
Panteion University

Abstract: This paper tests the cross-sectional income convergence in historical African ethnic homelands proxied by per capita CO2 emissions between 1850 and 2005 using both parametric and non-parametric tests of cross-sectional income distribution modality. We report that the cross- sectional income distribution in historical African ethnic homelands exhibits two very persistent steady states: one very low and one medium-to-high. Excluding from the analysis those areas that had no CO2 emissions throughout the sample period – although they were inhabited – we find that ethnic homeland areas still share two distinct steady states after the 1940s. Our study contributes to the literature on income convergence in ethnically divergent areas and more specifically in the historical ethnic homelands in the African continent.

JEL classification: C01; D24; F35; O43;
Keywords: historical ethnic homelands in Africa, CO2 emissions, income distribution

We thank Andrea Presbitero, Stelios Michalopoulos, Miguel-Leon Ledesma, Nathan Nunn Thomas Raddatz, Clive Richardson, Helga Stefansson and two anonymous referees for very valuable comments and suggestions. All errors remain ours.

✉ Corresponding author. Address: Department of Economic and Regional Development, Leof. Syngrou 136, 17671, Athens, Greece. (Phone: +30 210 9224948. Email: siourounis@grigorios-siourounis.info)

Recommended Citation
Christopoulos, D.K., Mimis, A., Siourounis, G. (2014). Historical Ethnic Homelands and Income Convergence in Africa. *Review of Economics and Institutions*, *5*(2), Article 5. doi: 10.5202/rei.v5i2.161. Retrieved from http://www.rei.unipg.it/rei/article/view/164

1 Introduction

In many parts of the world, history has played a significant role in the structure of economic norms, cross-border relationships, and economic development. Using fine units of analysis and new methods from the Geographic Information Systems (GIS) toolbox, a new strand of the literature emphasizes predominately the role of local historical characteristics in order to examine the roots of economic behavior with respect to local norms, ethics, ethnic and land peculiarities, as well as traditions shaped perhaps over centuries (see Spolaore and Warciarg 2013, and Herbst 2000).

A much more well researched strand of the literature that stems from the Neoclassical growth paradigm studies world income distribution and more specifically the issue of income convergence among countries in a specific time stamp. Much of this research has focused on the modality of world income per capita distribution in a given year. In a series of papers Quah (1993, 1996, 1997) and, most recently Henderson et al. (2008) document that traditional parametric methods (see for example Barro 1991, and Islam 1995) are not adequate to address the issue of per capita income convergence. This literature stresses that what is important is not whether a country is close to its steady state but what happens to the entire cross sectional distribution of economies. They conclude that world distribution is bimodal, indicating the existence of two clubs of countries: rich and poor ones.

In this paper we bring together those two strands of the literature and study income per capita convergence over time in African ethnic groups. We do so by combining information on the spatial distribution of African ethnic homelands even before European colonization in the mid/late 19th century with regional variation in economic performance proxied by per capita CO2 emissions from 1850 to 2005. Pre-colonial information on the location of African ethnic homelands is taken from George Peter Murdock's (1959) ethnographic map.

We follow Henderson et al. (2008), Silverman (1981, 1986) and Holzmann, and Vollmer (2008) and employ modality tests to investigate the properties of the cross-sectional regional ethnicity level income distributions as proxied by the per capita CO2 emissions calculated for each historical ethnic homeland for all the decades starting from 1850 until 2005. Calibration methods are also employed to avoid small sample biases. Our results indicate that persistent bimodality is a feature of the data throughout the sample. Excluding from the analysis the historical African ethnic homelands that are inhabited but do not have any noticeable CO2 emissions for the entire sample period, two out of three modality tests indicate that African tribes shared one low steady state level up until the 1940s (income distributions are unimodal), whereas for all decades afterwards bimodality cannot be rejected at any conventional significance level leading to the conclusion that some African tribes escaped the low steady state to approach a medium level of development.

Related literature. Recent country level studies of income convergence in Africa find little evidence of per capita income convergence (see for example Sperlich and Sperlich 2012). We advance this literature by documenting that, using regional-level data, the per capita income distribution across African ethnic homelands exhibits two persistent peaks: one low and one high. One major concern is those ethnic homelands that never recorded any CO2 emissions regardless of being populated. In fact, 74 out of the 834 ethnic homelands that are inhabited have no CO2 emissions throughout the sample period. Dropping those from the analysis, cross-sectional distributions continue to exhibit persistent bimodality after the 1940s.

On a broader scale, our work relates to the literature on income convergence across countries (see Sala-I-Martin 2006) and more specifically on the modality of global income distribution. Quah (1997) detected two modes in all the four years that he analyzed (1961, 1970, 1980, and 1988): one mode at a very low level of income per capita and a second richer one becoming more evident over time. Moreover, when income was weighted by population, he detected three modes in each of these periods, with a diminishing 'middle class' over time, a result that was visually confirmed by Goerlich Gisbert (2003). Using unweighted data and different samples, Bianchi (1997), Jones (1997), and Kumar and Russell (2002) detected a shift over time from unimodality to bimodality. Using fine units of analysis on a regional level we validate the persistent bimodality of the cross-sectional income distributions for African ethnic homelands.

Our study also relates to a growing literature on the historical origins of African development. For example, recent studies of Africa have shown that historical slave traits have shaped today's interpersonal beliefs and trust (Nunn and Watchenkon 2012), that the artificial drawing of borders dating back to the 1880's which neglected the local idiosyncrasies and tribal land distribution constitute the primary causes of the subsequent within-country and across-country conflicts (Michalopoulos and Papaioannou 2012) and that regional development in Africa depends crucially on the pre-colonial, ethnic, and political centralization characteristics (Michalopoulos and Papaiοnanou 2012). Finally, recent studies have shown the effects of colonial investments and tax collection systems on development (Huillery 2009, Berger 2009, and Arbesu 2011). Our study moves in the same directions and uses fine units of analysis to show that some African ethnic homelands remained at a low level of development for a very long time, whereas some others moved towards more plausible levels of development.

Paper structure. The rest of the paper is organized as follows: Section 2 describes the data we propose to use as a proxy for income on a regional level in Africa going back in time. Section 3 describes in detail the modality tests we use. Section 4 reports the results and section 5 draws conclusions.

2 Data

Historical ethnic homelands The innovation of the present study is the use of fine regional data to proxy for development on the ethnic homeland level as described in George Peter Murdock's (1959) ethnographic map that reports the spatial distribution of 844 ethnicities across Africa. Of those, 8 areas (Cape Verde, Libyan Desert, Madeira, Mauritius, Reunion, Sao Tome and Principe, Seychelles, West Saharan and Desert) are classified as uninhabited upon colonization and have, therefore, not been taken under consideration in our analysis. We propose the use of per capita CO2 emissions maps from 1850 until 2005.

Figure 1 - George Peter Murdockâs (1959) Map of Pre-Colonial African Ethnicities

CO2 emissions data. Our study needs detailed spatial data on economic activity dating several decades back in time. To the best of our knowledge, high resolution spatial data of economic activity spanning all of Africa dating back to the 1800s is not available. Ample research in environmental economics, however, has shown that country per capita CO2 emissions correlate strongly with development (see, for example, Holtz-Eakin and Selden 1995). We propose the use of per capita CO2 emissions to proxy for local economic activity. Recent studies have shown that grid level data on CO2 emissions can be used to study local economic activity, like agricultural activity, land use, livestock and others (see for example Goldewijk 2001). The emissions data from the Max-Planck Institute for Meteorology described thoroughly in Jones *et. al* (2011), are free, and consist of annual maps of anthropogenic carbon emissions due to land use changes (including wood harvest) measured in g(C)m-2s-1. The data is based on Houghton (2008) and is scaled so that the sum of the ten regions that it includes exactly equals the global emissions. Within each region, the emissions are weighted with the population densities (person/km2) from Goldewijk (2000), which are linearly interpolated between the years 1850, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, and 1990. The 1990 population density is assumed to have remained constant.

The data is on a regular grid with a resolution of 0.5 degrees (which is equivalent to 50 kms at the equator) in NetCFD format. We note that that these land use change emission maps were constructed to provide a sufficient accurate forcing for global scale Earth System models. In fact, annual continental scale land use change emission estimates from Houghton (2008) were just weighted by gridded population estimates. At least for the last 20 years (satellite era) there is locally more precise information about land use change available that unfortunately is not suitable for our analysis. This data serves better when used in aggregate comparison exercises. Aggregation in the ethnic homeland level moves in this direction significantly. In order to construct per capita CO2 emissions per tribe for each decade starting from 1850, we take the sum of the reported per capita emissions of all cells that fall within the historical ethnic homeland of each ethnic group as reported in George Peter Murdock's (1959) ethnographic map.

Thus, we construct each homeland specific value as:

$$per\ \ capita\ \ CO2_{i,t,j} = CO2_{i,t,j} * W_{i,t,j}$$

where:

$CO2_{i,t,j}$is the sum of CO2 emissions of all cells i that fall within the same ethnic homeland j, in year t.

W is a weighting factor that equals 1 over the sum of population density of all cells i that fall within the same ethnic homeland j, in year t.

i is a cell of dimension 0.5×0.5.

t stands for year.

j stands for homeland as describe in Murdock's (1959) ethnographic map.

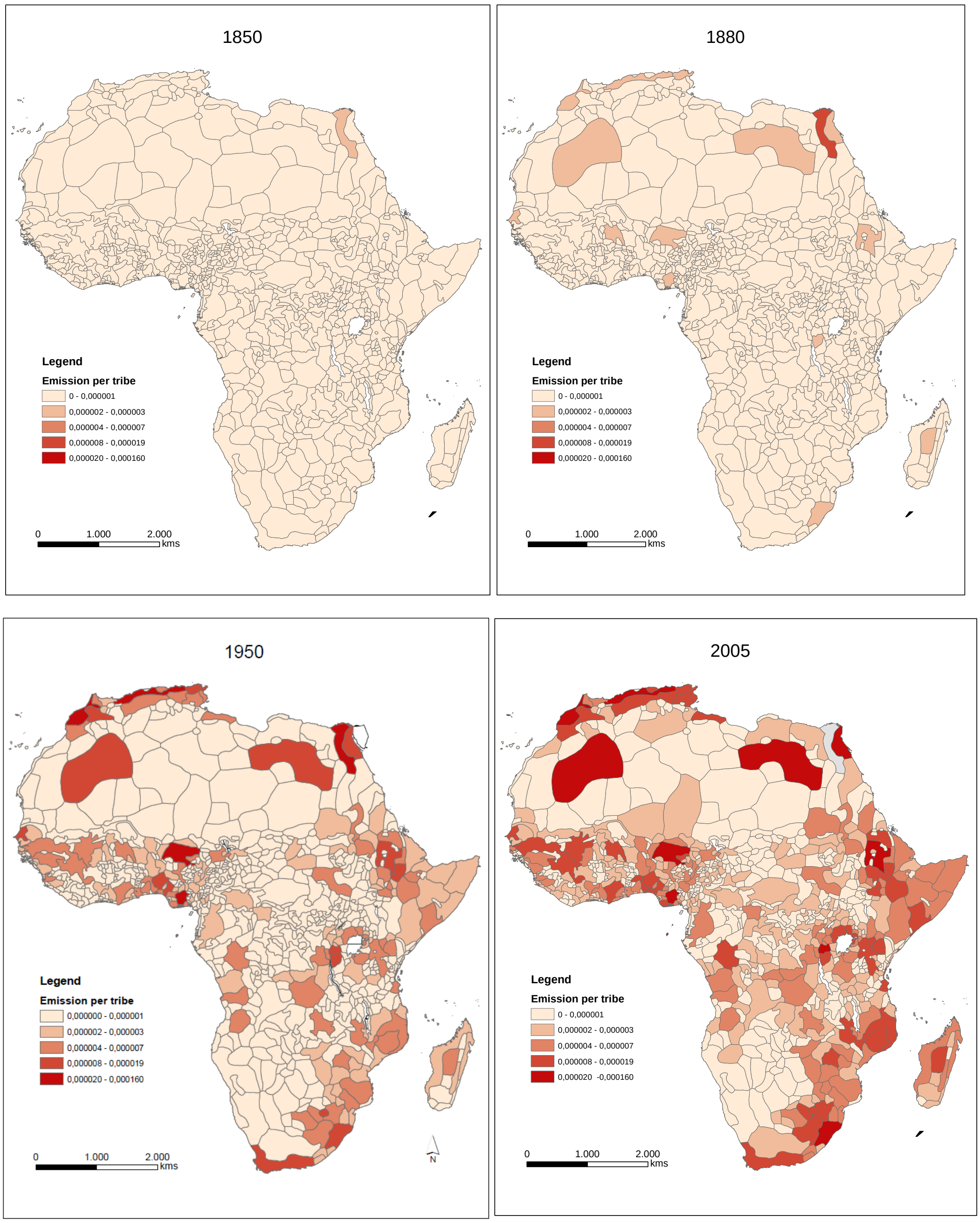
To get a visualization of the data, Figure 2 plots 4 different years back in time: for the starting date of the sample (1850), the year of the Berlin Conference (1880-1) that marked the start of the European colonization in Africa, the year 1950 that is just around the eruption of massive independence movements in Africa (for example, in 1941 Ethiopia inaugurated a vast wave of 50 independence movements from Colonial occupation), and finally 2005 that marks the last year of available data.

Although estimated per capita CO2 emissions within ethnic homelands vary significantly across time, it should be noted that the cross-sectional distribution does not vary at the same rate. Areas that record high values in 1880, record high values in 1950 and 2005 as well. To further illustrate the coverage of the data, in 1850 only one ethnic homeland located along the northern part of the Nile River, namely the Egyptians, had any noticeable per capita CO2 emissions whereas in 1880 this rose to fourteen ethnic homelands or 1.8% of the total populated regions. Those were: the Algerians , the Egyptians , the Hausa, the Ibo, the Morocca, the Tunisians, the Zulu, the Wolof, the Rundi, the Maaza, the Merina, the Mossi, the Malinke and the Amhara. From all populated historical ethnic homelands, in 1850 there were 168 without any CO2 emissions recorded. This figure declined to 114 in 1880, to 91 in 1950 and finally to 74 in 2005. We next present evidence of the suitability of the proposed data to proxy for regional development at the ethnic homeland level.

Cross-validation 1: per capita CO2 emissions and per capita light density. In the absence of any reliable regional measure for economic development in Africa, researchers in recent studies have used the log of light density. Light density refers to the intensity of light in a nocturnal satellite map of the continent. Michalopoulos and Papaioannou (2012 and 2013), Ghosh *et al.* (2010) and Ciccone and Jarocinski (2010) document that this measure adequately approximates development in Africa especially considering the population. There are two reasons why we propose the use of per capita CO2 emissions instead: the first is that light density dates back only to 1992 and the second is that there are a lot of completely dark areas which leaves out considerable information about the regional variation among ethnic homelands. To compare our proxy with light density we average light density across cells that fall within the same 0.5 x 0.5 cell, the grid of per capita emissions map, for each ethnic group in 2005 and divide by the corresponding population density (persons/km^2). Light density at night comes from the National Geophysical Data Center, whereas population data comes again from Goldewijk (2000). This resulted in 10,231 comparable cells-observations. From those, 6773 cells are completely dark, which means that 66.2% of the African continent has no light activity (a number comparable to that reported in previous studies that use lights as a proxy for local activity). The comparable number for no CO2 emissions is 2910, from which 2552 have no lights. Thus, only 358 observations have no emissions but do have lights. This, of course, does not tell us anything about the quality of the CO2 data set. It only tells us that the coverage is wider.

Within analysis. In Figure 3 below, we plot the within ethnic homelands light density and per capita CO2 emissions in 2005. Calculated within ethnic homelands conditional correlations (R^{2}s from regressing emissions

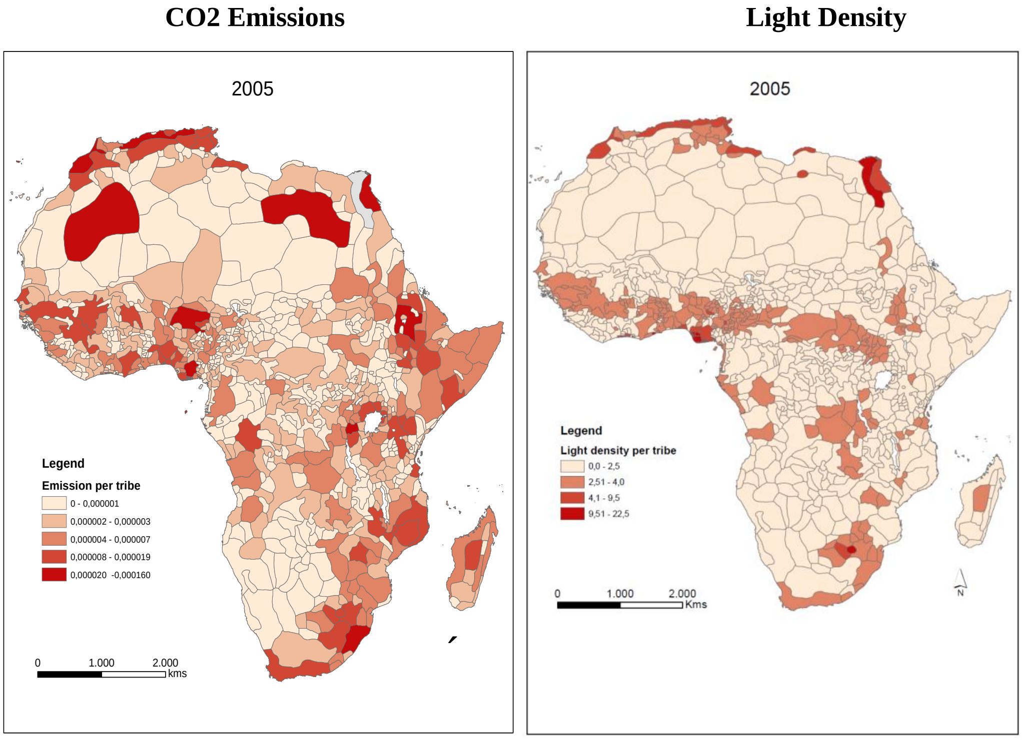
Figure 2 - Regional Per Capita CO2 Emissions by African Historical Ethnic Homelands



on lights) of per capita light density and per capita CO2 emissions vary between 0.30 and 0.65. When the top and bottom 1% are dropped, correlations become more concentrated around 0.6. As an example, the Hausa and the Zulu ethnic homeland within conditional correlations are 0.45 and 0.65 respectively.

Between analysis. To investigate the between ethnic homelands correlation of emissions and light, we aggregate per capita CO2 emissions in the ethnic homeland level and we contrast it with light density for the same year as the one reported in Michalopoulos and Papaioannou (ECMA 2013), namely 2000. To use the maximum available observations we take the ln (values + 0.01) to avoid very low (close to zero) emission areas and/or very low or completely dark areas, dropping all the uninhabited ones (total 8). Unconditional correlation stands at 64.16% and is significant at any conventional significance level. A Spearman rank test also confirms a significant association (41.14%). These results, however, are driven heavily by outliers. To show this, we drop the top and bottom 1% and redo all the calculations. Unconditional correlation falls but remains very significant to 27.72% (Spearman stands at 23.09%). Doing the same and dropping the top and bottom 5% marginally reduces the unconditional correlation to 25.86% (Spearman stands at 19.8%). Thus, there is a significant bias generated by very few observations that record extremely high or low emissions and lights but, in general, the two series are statistically significantly positively correlated.

Figure 3 - Per Capita CO2 Emissions and Per Capita Light Density in the Zulu Historical Ethnic Homeland in 2005



CO2 Emissions

Light Density

Cross-validation 2: per capita CO2 emissions and pre-colonial characteristics. To further strengthen our argument on the suitability of per capita CO2 emissions to proxy for local development on the ethnic homeland level within the African continent, we next calculate unconditional correlations between calculated CO2 emissions in 1880 and pre-colonial ethnic specific characteristics per African ethnic group obtained from Murdock's (1967) Ethnographic map (as reported in Michalopoulos and Papaioannou 2013). We report those correlations in Table 3. We note that CO2 emissions capture primarily land use change (including wood harvest), thus we expect it to correlate negatively with areas that depend less on land. Interestingly enough, per capita CO2 emissions in 1880 are statistically significantly negatively correlated with gathering (-15.86%) and hunting (-13.49%). Since urban areas in Africa are concentrated close to the sea, we should also expect them to correlate negatively with such activities as the ones recorded in the CO2 dataset, which is indeed the case since emissions are negatively correlated with distance from the capital (-27.58%) and distance from the sea (-21.23%) measures. Finally, emissions are statistically significantly positively correlated with agriculture (23.86%), settlement patterns (30.43%), subsistence economy (23.32%), jurisdictional hierarchy beyond the local level (18. 08%), and previous presence of slavery (12.67%), all indications of more stable local settlements around the use of land for farming or animal feeding.[1] The positive correlation with the previous presence of slavery is interesting on its own since it shows that in such areas land use has played an important role in the local economy for too long. Overall we believe that the proposed proxy for development in African ethnic homelands since the 1850s copes well with survey data dating back to that time.

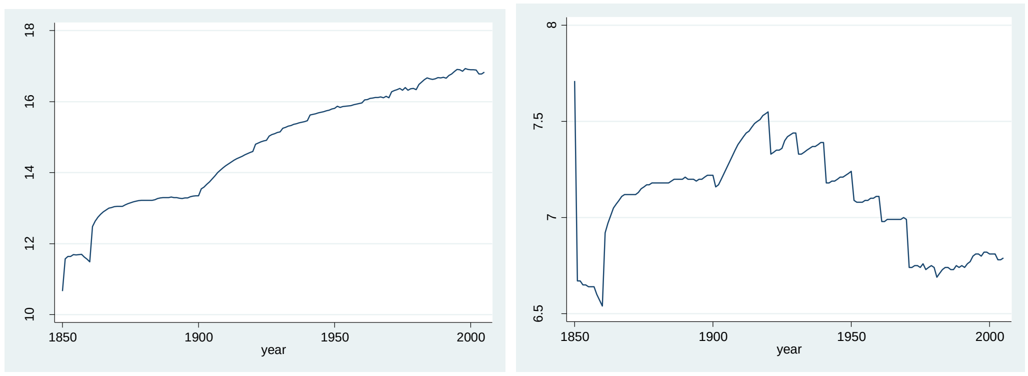
Descriptive statistics of cross-sectional per capita CO2 distributions. In Figure 4 we plot the yearly calculated first four moments of the cross-sectional distribution of the log per capita CO2 emissions plus 0.01 to avoid losing any non-negative observations from areas that are inhabited but have no emissions at all. An immediate observation is the high persistence of all moments (the autoregressive coefficient varies between 0.6 for the cross-sectional yearly standard deviation to 0.9 for the cross-sectional yearly average). In terms of trends, per capita CO2 emissions rise steadily for the last 150 years. Volatility also seems to fall monotonically. The cross-sectional distribution loses its tail of very poor ethnicities over time and this contributes to the uniform fall in skewness although per capita CO2 emissions become more and more concentrated around the mean value as depicted by the steady increase in calculated cross-sectional kurtosis.

To further get a feeling of the cross-sectional shape of the income distribution as proxied by CO2 emissions, Figure 5 reports Gaussian kernel density estimates of the cross sectional per capita CO2 emissions per ethnic homeland in 1850, 1881, 1950 and 2005. It is apparent that the distribution is bimodal, exhibiting two peaks: one very low and one high. In the next section we first discuss and then apply formal tests for modality.

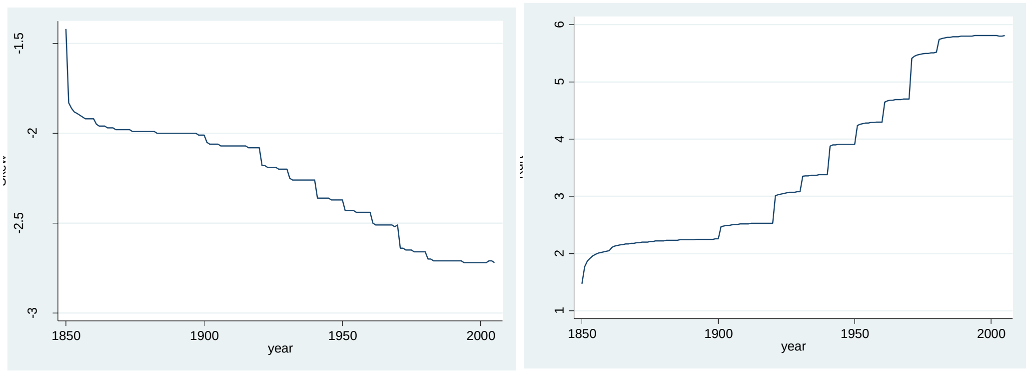
---

[1] Changing the year of consideration, a decade before or afterwards for per capita CO2 emissions does not change those correlations in any significant way.

Figure 4 - Yearly Cross Sectional Moments of Ethnic Homelands Log (Per Capita CO2 Emissions +0.01) for the Entire Sample Period (1850-2005)



Average **Standard Deviation**



Skewness **Kurtosis**

3 Modality Tests

To formally assess whether ethnic homelands converge towards two income levels we employ two tests: The Silverman (1981,1986) test and the Dip test (Hartigan and Hartigan 1985).

The Silverman test Let (Z_i), $i = 1, 2, ...m$ denote a sample Z of size m from a distribution with unknown density f. A non parametric estimate of this density $\tilde{f}(z)$ is as follows,

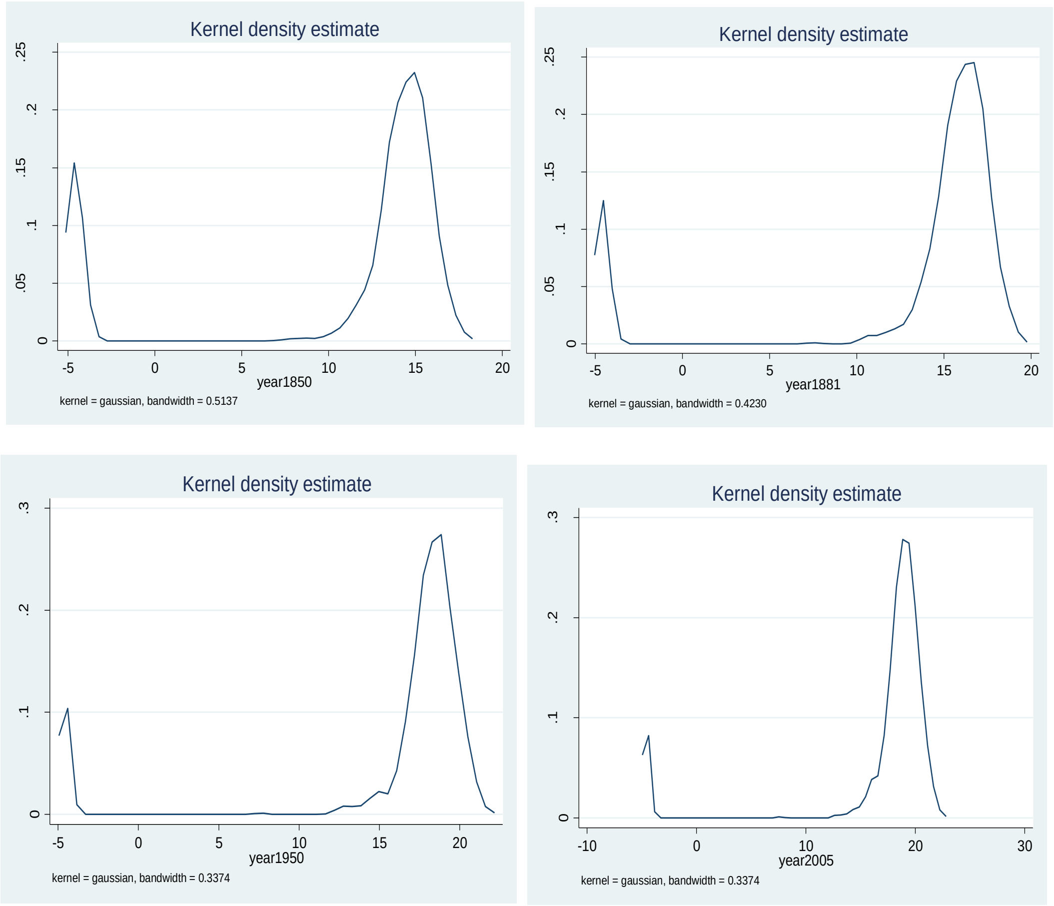
$$\tilde{f}(z,h) = m^{-1}\, h^{-1} \sum_{i=1}^{m} K[(1/h)(z_i - z)]$$

where K is a kernel normal function while h is a smoothing parameter, also called bandwidth parameter. Large values of h are associated with a few modes (q) in the estimated density.

A test statistic (Silverman 1981) can then be written as,

$$\tilde{h}^{q}_{crit} = \inf\left\{ h : \tilde{f}(z,h)..has.at..most..q.modes \right\}$$

Figure 5 - Cross-Sectional Gaussian Kernel Density Estimates of Ethnic Homelands Log (Per Capita CO2 Emissions + 0.01) for Years 1850, 1881, 1950, and 2005



which is used to test the null hypothesis, that is, f has q modes against the alternative that f has more than q modes. Large values of $\tilde{h}^q_{crit}$ show evidence against the null hypothesis.

A bootstrap procedure is employed to compute the $\tilde{h}^q_{crit}$ statistic. This procedure is given by,

$$y_i = [1 + (\tilde{h}^q_{crit})^2/\sigma^2)^{-0.5}(Z_i + (\tilde{h}^q_{crit})^2 e_i)$$

where Z_i is sampled uniformly, with replacement, from the data $z_1, ...z_m$, σ^2 is the sample variance of the data, and e_i is a normal random variable. In this way y_i is randomly drawn from a smooth conditional distribution. The conditional kernel density for a bootstrap sample $Y = \{y_1, ...y_m\}$ is given by,

$$\tilde{f}_*(z,h) = m^{-1}(\tilde{h}^q_{crit})^{-1}\sum_{i=1}^{m} K\left[(1/\tilde{h}^q_{crit})(y_i - z\right]$$

Acceptance or rejection of the null hypothesis can be based on the following expression,

$$\tilde{P} = P[\tilde{h}^{q*}_{crit} \geq \tilde{h}^{q}_{crit}]$$

where $\tilde{h}^{q*}_{crit}$ is associated with the conditional kernel density $\tilde{f}_*(z, h)$ using the bootstrapped sample $Y = [y_1, ...y_m]$.

Finally, the Hall and York (2001) method was applied to Silverman's test to obtain the correct critical values.

The Dip test. The Dip statistic has been proposed by Hartigan and Hartigan (1985) and measures the unimodatility of a sample as the maximum difference between the empirical distribution function and the unimodal distribution function that minimizes that maximum difference. In particular, the Dip test considers that a distribution function F is unimodal with mode q if F is convex to the left $(-\infty, q)$ and concave to the right $(q, -\infty)$. The test is based on the idea of a minorant. Within this context we define the greatest convex minorant of F in $(-\infty, \alpha)$ as the supremum of all convex functions $G(z)$ [*sup* $G(z)$ for $z \leq \alpha$] that are nowhere greater than F and the least concave majorant of F in $[\alpha, \infty]$ as the infimum of all concave functions $R(z)$ [*inf* $R(z)$ for $z \geq \alpha$] that are nowhere less than F. The Dip test of a distribution function $F$is given by

$$D(F) = \inf_{R \in U} \sup_{-\infty < z < \infty} [F(z) - R(z)]$$

where U is the class of all unimodal distribution functions.

The null hypothesis that the distribution function F has a unimodal density f is tested against the alternative that it has more than one. Hartigan and Hartigan (1985) replaced the theoretical distribution F with the empirical one $(\tilde{F})$ of a random n-sample suggesting thus the $D(\tilde{F})$ statistic. The null hypothesis that the function F is unimodal is rejected against the alternative when $D(\tilde{F})$ exceeds the $\alpha-$critical level. In other words, the reference distribution for estimating the Dip statistic is the uniform unimodal distribution. Following Cheng and Hall (1998), $P-$values are calculated by comparing the Dip statistic obtained with those for repeated samples of the same size from a uniform distribution.

3.1 Main Results

Silverman's Dip statistical test results for all decades starting from 1850 till 2005 are depicted in Table 1. For all decades, we examine the hypothesis of whether unimodality is either rejected (Dip test) or cannot be accepted (Silverman test) at conventional statistical levels. The results show that in Africa there are two well separated clusters of rich and poor ethnic homelands that are very persistent over time. This result is to be expected, given the two peaks in the kernel density estimates of the cross sectional distributions described in Figure 6 above. This is due to the fact that a significant part of the sample, namely 74 out of the 834 ethnic homelands or 9% of the sample, has zero CO2 emissions, even in 2005, although they are inhabited.

Our results are in line with Proto (2007), who shows that an excessive number of poor individuals competing for scarce land within a specific region, like an ethnic homeland here, drive up the rental price of land, which then slows the process of accumulation and forces the economy to converge to a less efficient dual equilibrium. Here, very poor but populated homelands might find it difficult to move up the development ladder if available regional resources are scarce. They are also consistent with Gyawali *et al* (2008), which shows that although income convergence in smaller, mainly rural geographic areas may be consistent with neoclassical growth theories, especially when more discrete, micro level data are analyzed, ethnically fragmented areas within a broader political entity, like a state in US, tend to diverge in terms of per capita income. Clearly, across African ethnic homelands, there is a persistent pattern: some ethnic homelands remain very poor throughout the sample without any noticeable development and share a common very low steady state and the rest enjoy a higher steady state.

We next focus on the cross-sectional distribution of those ethnic homelands that had positive per capita CO2 emissions throughout the sample, including those that had zero CO2 emissions in 1850 but gradually turned positive over the sample period. Thus we exclude 74 tribes (shown in Figure 6) from the modality tests and study the cross-sectional income convergence in the remaining African ethnic homelands.

In Figure 7 below we replicate the Gaussian Kernel density estimates for the new sample for the same years as in Figure 6. Although in 1850 there were a number of ethnic homelands without any economic activity (168 ethnic homelands had no emissions at all) their numbers have declined over time to vanish in 2005. This shows clearly that not all ethnic homelands experienced the same historical and/or economic conditions in the last 150 years, resulting in a considerable variation within the cross-sectional income distribution regardless of the persistence of its shape over time.

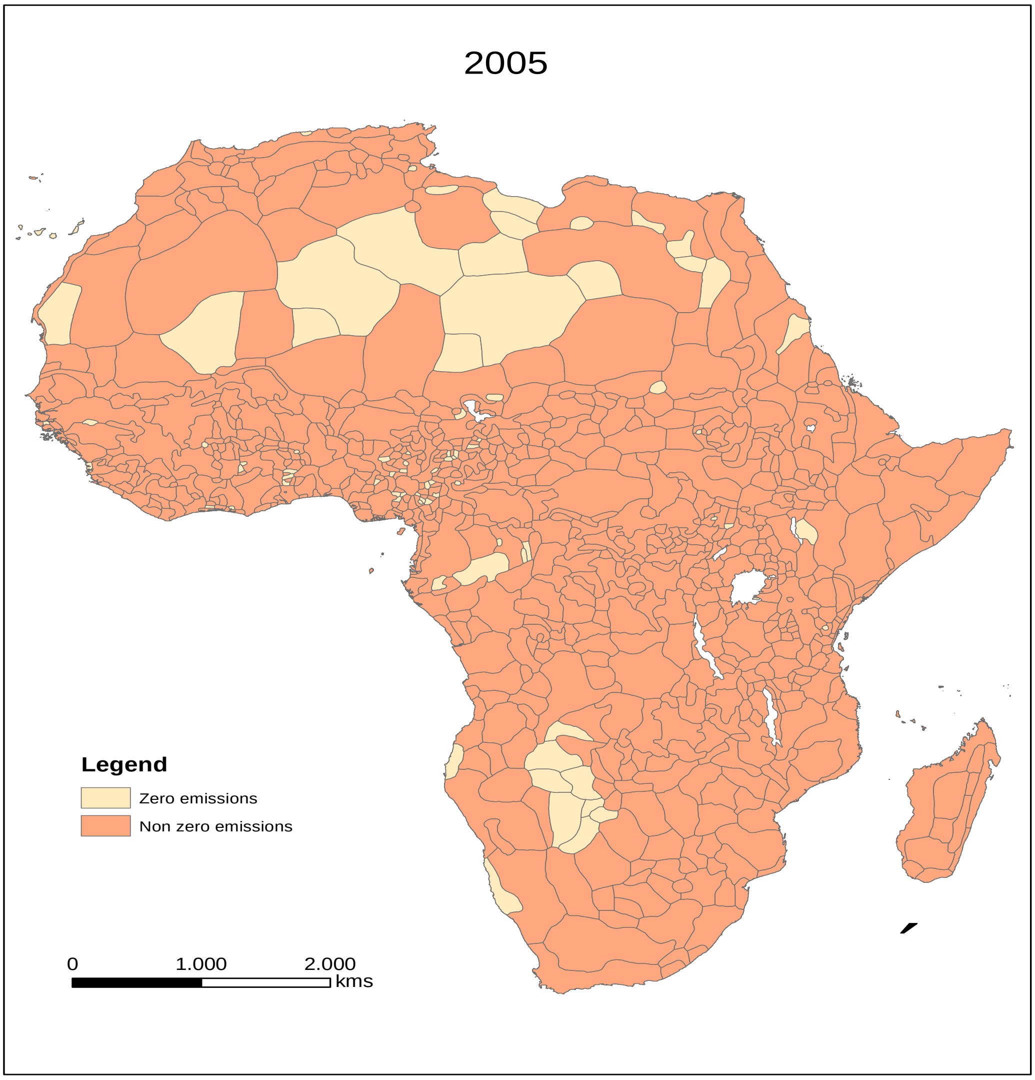
To formally assess this claim we recalculate the two tests and report the results in Table 2. The Silverman test shows that the distribution is still bimodal for all of the decades we have examined whereas the Dip test indicates that the distribution is bimodal until the 1920s and then it becomes unimodal. This means the Dip test shows that after the 1930s, the African Ethnic homelands converge towards a common steady state of development. Since we get two conflicting results we next use a parametric bimodality test in a search for a confirmation of the Silverman or the Dip test.

Following Holzmann, and Vollmer (2008) we consider $f(z;\theta)\ \theta \in \Theta \subset R^d$ be a parametric family of densities allowing for the two component mixture family namely,

$$f(z;\theta_1,\theta_2,p) = pf(z;\theta_1) - (1-p)f(z;\theta_2)$$

where $(\theta_1,\theta_2,p) \in \Theta \times \Theta \times [0,1] = \Theta_{Mix} \subset R^{2d+1}$.

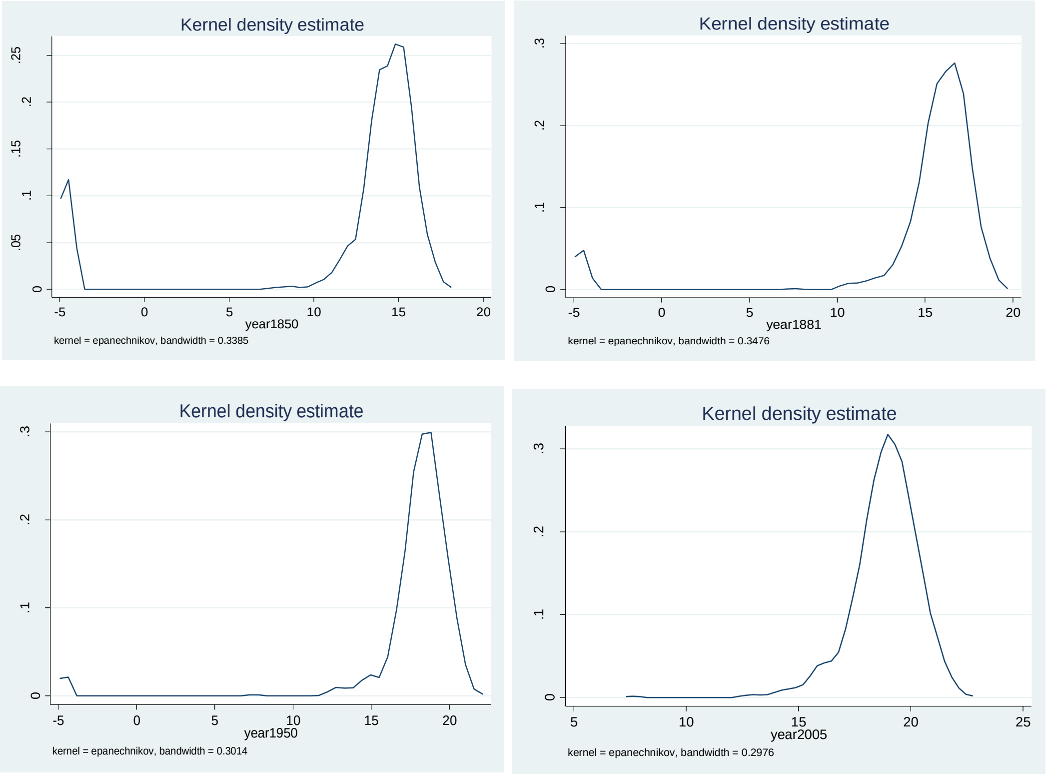
Figure 6 - African Historical Ethnic Homelands Without Any CO2 Emission Throughout the Sample Period (1850-2005)



To test for equal variances of the two component densities we assume that $E_{Mix} \subset \Theta_{Mix}$, where $E_{Mix} \subset R^q$ for a minimal $q \leq 2d+1$. By assuming that the mixture density is at most bimodal we can break up the set E_{Mix} disjointly into $E_{Mix} = E_{UModal} \cup E_{BModal}$, where E_{UModal} is the unimodal part while E_{BModal} is the bimodal one. The likelihood function is then given by:

$$L_n(\theta_1, \theta_2, p) = \sum_{k=1}^{n} \log f(z_k; \theta_1, \theta_2, p)$$

Figure 7 - Gaussian Cross-Sectional Kernel Density Estimates of Ethnic Homelands Log (Per Capita CO2 Emissions + 0.01) for Years 1850, 1881, 1950, and 2005 Excluding Zero Emissions Areas Throughout the Sample Period (1850-2005)



A likelihood ratio (LR) is employed

$$R_n : 2\left(\underset{(\theta_1,\theta_2,p)\in E_{Mix}}{Sup} L_n(\theta_1,\theta_2,p) - \underset{(\theta_1,\theta_2,p)\in E_{UModla}}{Sup} L_n(\theta_1,\theta_2,p)\right) \xrightarrow{D} (\chi_0^2 + \chi_1^2)$$

where χ_0^2 is the point measure at zero and χ_1^2 is the chi-square distribution with 1 degree of freedom. The null hypothesis is bimodality. Based on simulation results Holzmann, and Vollmer (2008) conclude that the LR test outperforms slightly the Silverman's test while both tests outperform the Dip test when the density is symmetric and bimodal. When the density is asymmetric and bimodal the LR test outperforms Silverman's test. In this case the Dip test has no significant power. The fourth column of Table 2 reports the p-values for the null hypothesis of bimodality. This test shows that until the 1930s the distribution was unimodal but after the 1940s the African ethnic homelands have two steady states: some converge towards a low level of income whereas others converge to a higher level of income, which is consistent with what the Silverman test shows. This means that even if we exclude all those ethnic homelands that never managed to reach any positive CO2 emissions from the analysis, there are still two peaks in the cross-sectional income distribution of the remaining ones: one low and one medium-to-high after the 1940s.

3.2 *Regional Robustness*

In the above analysis, we have not considered regional variations that might be important in such a diverse continent. To address this issue we run both the tests (Silverman and Dip) for the five regions of Africa namely: Eastern, Middle, Northern, Southern and Western again. To visualize the data, we plot regions, countries and ethnic groups in Figure 8 for the year 2000. The results for the two tests for the same year (2000) are reported in Table 4. Results for all the regions except Western Africa accord with the reported results in the latter part of Table 2, which is that the results are conflicting using Silverman and Dip tests. In an effort to resolve this conflict we also apply the Bimodality test (third column in Table 4), which shows that the distribution for all regions except Western Africa is unimodal. This might be due to the fact that ethnic groups within those areas use common resources and face similar landscape and weather conditions, which are paramount for agricultural production, thus ethnic homelands within those regions tend to converge to a common steady state if development is measured with per capita CO2 emissions from land use. Western Africa is the only region for which the Bimodality test shows multimodality. This might be due to the fact that this region is characterized by large discrepancies with regard to the landscape of the countries it encompasses. On the one hand, it includes large countries with few ethnic groups that occupy the significant but very poor land of the Sahara desert (like Mauritania, Mali, Niger and Chad), and on the other hand it includes countries with dense emissions and large numbers of ethnic groups (as for example, Senegal, Guinea, Cote d' Ivore, Ghana, Togo, Benin, and Nigeria).[2]

3.3 *Data Robustness*

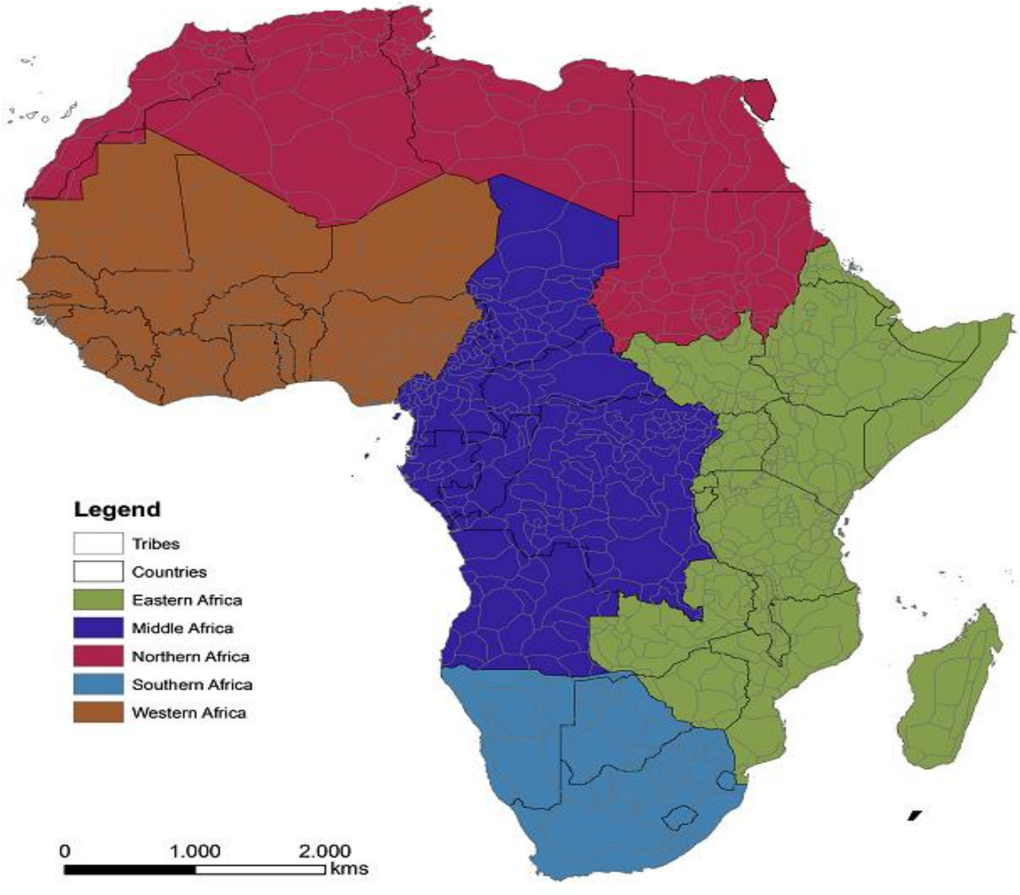
The proposed measure to proxy for regional development includes CO2 emissions due to agricultural activity and forest harvesting only, something that might lead to serious biases for cross-sectional comparisons particularly for the last quarter of the 19th century when urbanization intensified globally. This however, is less of a problem with regard to the African continent since serious urbanization started much later. In the 1900s, less than 5% of Africans lived in urban areas and this figure rose to only about 15% in the 1950s, which means that the results are not driven by population variation in large cities.

Serious mis-measurement might also arise for the $CO2_{i,t,j}$ variable, due

---

[2] Furthermore, we obtained per capita GDP data for all African countries in 2000 and summed up all per capita CO2 emissions in the country level to compare convergence tests. Both tests indicate multimodality in accord with Table 1.

Figure 8 - Per Capita CO2 Emissions for 2000 per Ethnic Homeland, Country and African Region



the fact that the data for the entire continent is estimated with the use of only two data points: North Africa and the Middle East and Tropical Africa (Houghton 2008). Regarding North Africa and Middle the East, there is considerable variation, but overall CO2 emissions remained low over time. The same is not true for Tropical Africa where there is a constant increase in CO2 emissions to comparable levels from other parts of the world. Although two (2) data points are sufficient to describe the entire cross-sectional distribution across ethnic homelands under the assumption of normality, time invariance across time due to lack of true CO2 emissions variation limits the scope of our exercise in that modality tests are driven entirely from the cross-sectional variation in population density. Although this has been proven to be a very good proxy for regional development (see for example Goldewijk 2001) we need to be cautious in the interpretation of our results.

Finally, the constructed per capita CO2 emission series use population data dating from a time before any census was ever conducted, thus $W_{i,t,j}$ might be poorly estimated, begging the question of how reliable it is for all years prior to the 1960s. Goldewijk (2001) reports population estimates dating back to the 1700s from 22 studies (Table 2, page 420) that show remarkable consistency with his own estimates. We measured the cross-sectional correlations across different years in the past. For example, for 2000, unconditional correlation between CO2 emissions and population is 0.41, whereas for 1880 it is 0.63. This is an indication that CO2 data from earlier dates are more dependent on the population data. To further assess the significance of this concern, in our modality tests we recalculated the two main tests (namely the Silverman and the Dip tests) utilizing per capita CO2 emissions at the ethnic homeland level constructed using population densities from a restricted sample that limits the population metric to a maximum of 20 persons per square Km. Thus, cross sectional variability, if any, comes mainly from constructed CO2 emissions. Using this data set, results do not change in any material way, which means that cross-sectional convergence is not driven by the quality of the population data but by the shape of the cross sectional distribution.

4 Conclusions

This paper proposes the use of geographically referenced per capita CO2 emissions as a proxy for regional development and uses parametric and non parametric tests to assess whether African ethnic homelands, as defined in George Peter Murdock's (1957) ethnolinguistic map, share a common steady state in terms of income per capita from 1850 until 2005. The main result of the analysis is that the distribution of log per capita CO2 emissions is characterized by bimodality indicating two prevailing steady states among African ethnic groups: one low and one high. Removing from the analysis the 74 ethnic homelands that never reach a noticeable level of CO2 emissions throughout the sample, two out of the three tests we employ show that the remaining ethnic homelands diverge towards two different steady states. We see our work as a contribution towards a better understanding of the cross-sectional regional income distribution in Africa using a fine unit of analysis – that of geo-referenced CO2 emissions – as a development proxy.

Table 1 - Values for (Calibrated) Modality Tests

Year	*Silverman*	*Dip*
1850	0.451	0.001***
1860	0.112	0.001***
1870	0.234	0.001***
1880	0.336	0.001***
1890	0.657	0.001***
1900	0.837	0.001***
1910	0.690	0.001***
1920	0.693	0.001***
1930	0.360	0.001***
1940	0.419	0.001***
1950	0.373	0.001***
1960	0.484	0.001***
1970	0.765	0.001***
1980	0.523	0.001***
1990	0.669	0.001***
2000	0.658	0.001***
2005	0.659	0.001***

*,**,*** Indicate significance at 10%, 5%, and 1% respectively. For the Dip tests, the null hypothesis is that the distribution function F has a unimodal density f against the alternative that it has more than one. Low p- values show rejection of the null hypothesis. For the Silverman test, the null hypothesis is that f has 2 modes against the alternative that f has more than 2 modes. Large p- values show acceptance of the null hypothesis.

Table 2 - Values for (Calibrated) Modality Tests Excluding African Ethnic Homelands with Always Zero Emissions

Year	*Silverman*	*Dip*	*Bimodality test*
1850	0.473	0.001 ***	0.001 ***
1860	0.101	0.004***	0.269
1870	0.258	0.005***	0.001***
1880	0.453	0.005***	0.001***
1890	0.666	0.005***	0.001***
1900	0.855	0.001***	0.008***
1910	0.673	0.003***	0.001***
1920	0.676	0.004***	0.145
1930	0.382	0.145	0.001***
1940	0.418	0.278	0.501
1950	0.349	0.279	0.504
1960	0.444	0.483	0.506
1970	0.681	0.699	0.508
1980	0.354	0.817	0.520
1990	0.279	0.807	0.518
2000	0.281	0.812	0.589
2005	0.274	0.822	0.590

*,**,*** Indicate significance at 10%, 5%, and 1% respectively. For the Dip tests, the null hypothesis is that the distribution function F has a unimodal density f against the alternative that it has more than one. Low p-values show rejection of the null hypothesis. For the Silverman test, the null hypothesis is that f has 2 modes against the alternative that f has more than 2 modes. High p- values show acceptance of the null hypothesis. For the bimodality test the null hypothesis is that f has two modes (the empirical distribution is bimodal) against the alternative that f is unimodal. Low P values show rejection of the null hypothesis.

Table 3 - Correlation of Per Capita CO2 Emissions in 1880

Pre-Colonial Activity	
Gathering	-0.1586
p-value	0.0008
Obs	441
Hunting	-0.1349
p-value	0.0045
Obs	441
Agriculture	0.2386
p-value	0.0000
Obs	441
Settlement Patern	0.3043
p-value	0.0000
Obs	441
Jurisdictional Hierarchy Beyond the Local	0.1808
Level	0.0001
	441
Subsistence Economy	0.2332
p-value	0.0000
Obs	441
Former Presence of Slavery	0.1267
p-value	0.0077
Obs	441
Distance to the Capital	-0.2758
p-value	0.0000
Obs	441
Distance to the Sea	-0.2123
p-value	0.0000
Obs	441

Table reports pair wise correlations between CO2 emissions and a number of pre-colonial measures form Murdock's 1967 Ethnographic Atlas retrieved from Michalopoulos and Papaioannou (2013). P-values and number of usable observations are also reported below calculated correlation.

Table 4 - P-Values of (Calibrated) Modality Tests per Africa Region

	Silverman	*Dip*	*Bimodality Test*
Eastern Africa	0.532	0.991	0.001
Middle Africa	0.802	0.439	0.001
Northern Africa	0.174	0.611	0.001
Southern Africa	0.225	0.953	0.001
Western Africa	0.061*	0.020**	0.500

*,**,*** Indicate significance at 10%, 5%, and 1% respectively. For the Dip tests, the null hypothesis is that the distribution function F has a unimodal density f against the alternative that it has more than one. Low p- values show rejection of the null hypothesis. For the Silverman test, the null hypothesis is that f has 2 modes against the alternative that f has more than 2 modes. High p- values show acceptance of the null hypothesis. For the bimodality test the null hypothesis is that f has two modes (the empirical distribution is bimodal) against the alternative that f is unimodal. Low P values show rejection of the null hypothesis.

References

Andres R.J. et al., 1996. Geographic Patterns of Carbon Dioxide Emissions from Fossil-Fuel Burning, Hydraulic Cement Production and Gas Production on a One Degree by One Degree Grid Cell Basis: 1950 To 1990. ORNL/CDIAC-97, NDP058. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A. 56 pp. doi: 10.3334/CDIAC/ffe.ndp058. doi:10.3334/CDIAC/ffe.ndp058

Asongu, S. A., 2012. African Development: Beyond Income Convergence, MPRA Paper No. 36054.

Barro, R.J., 1991. Economic Growth in a Cross Section of Countries. Quarterly Journal of Economics 106, 407-433. doi:10.2307/2937943

Cheng, M.Y., Hall, P., 1998. Calibrating the Excess Mass and Dip Tests of Modality. Journal of the Royal Statistical Society B 60, 579-590. doi:10.1111/1467-9868.00141

Ciccone, A., Jarocinski, M., 2010. Determinants of Economic Growth: Will Data Tell. American Economic Journal Macroeconomics, forthcoming. doi:10.1257/mac.2.4.222

Easterly, W., R. Levine, 1997. Africa's Growth Tragedy: Policies and Ethnic Divisions. Quarterly Journal of Economics 112, 1203-1250. doi:10.1162/003355300555466

Ghosh, T., Powell, R.L., Elvidge, C.D., Baugh, K.E., Sutton, P.C., S. Anderson, 2010. Shedding Light on the Global Distribution of Economic Activity. The Open Geography Journal 3, 148-161.

Goldewijk, K., 2000. Estimating Historical Land Use Changes Over the Past 300 Years: The HYDE Database, Version 2.0. Global Biogeochemical Cycles.

Goldewijk, K., 2001. Estimating Historical Land Use Changes Over the Past 300 Years: The HYDE Database, Version 2.0. Global Biogeochemical Cycles 15, 417-433. doi:10.1029/1999GB001232

Gyawali, B, Fraser, R., Bukenya, J., Schellas, J., 2008. Income Convergence in a Rural, Majority African-American Region. The Review of Regional Studies 38, 45-55.

Hall, P., York, M., 2001. On the Calibration of Silverman's Test for Multimodality. Statistica Sinica 11, 515-536.

Hartigan, J.A., Hartigan, P.M., 1985. The Dip Test of Unimodality. Annals of Statistics 13, 70-84. doi:10.1214/aos/1176346577

Henderson, J.D., Parmeter, F.C., Russell, R.R., 2008. Modes, Weighted Modes and Calibrated Modes: Evidence of Clustering Using Modality Tests. Journal of Applied Econometrics 23, 607-638. doi:10.1002/jae.1023

Islam, N. 1995. Growth Empirics: A Panel Data Approach. Quarterly Journal of Economics 110, 1127-1170. doi:10.2307/2946651

Herbst, J., 2000. States and Power in Africa. Princeton University Press, Princeton NJ.

Holtz-Eakin, D., Selden, T.M., 1995. Stoking The Fires? CO2 Emissions And Economic Growth. Journal of Public Economics 57, 85-101. doi:10.1016/0047-2727(94)01449-X

Holzmann, H., Vollmer, S., 2008. A Likelihood Ratio Test for Bimodality in Two-Component Mixtures with Application to Regional Income Distribution in the EU, Advances in Statistical Analysis 92, 57-69. doi:10.1007/s10182-008-0057-2

Houghton, R.A., 2008. Carbon Flux to the Atmosphere from Land-Use Changes: 1850-2005. In TRENDS: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Michalopoulos, S., Papaioannou, E., 2013. Pre-Colonial Ethnic Insitutions and Contemporary Africam Development. Econometrica 81, 113-152. doi:10.3982/ECTA9613

Michalopoulos, S., Papaioannou, E., 2012. Divide and Rule or the Rule of the Divided: Evidence from colonial Africa, mimeo.

Murdock, G.P., 1959. Africa: Its Peoples and their Culture History. McGraw-Hill Book Company, New York.

Nunn, N., Wantchekon, L., 2009. The Slave Trade and the Origins of Mistrust in Africa. NBER Working Paper No. 14783.

Quah, D., 1993. Galton's Fallacy and Tests of the Convergence Hypothesis. Scandinavian Journal of Economics 95, 427-444. doi:10.2307/3440905

Quah, D., 1996. Twins Peaks: Growth and Convergence in Models of Distribution Dynamics. Economic Journal 106, 1045-1055. doi:10.2307/2235377

Quah, D., 1997. Empirical for Economic Growth and Distribution: Stratification, Polarization and Convergence Clubs. Journal of Economic Growth 27-59. doi:10.1023/A:1009781613339

Proto, E., 2007. Land And The Transition From A Dual To A Modern Economy. Journal of Development Economics 83, 88-108. doi:10.1016/j.jdeveco.2005.11.004

Silverman, B.W., 1981. Using Kernel Density to Investigate Multimodality. Journal of the Royal Statistical Society B 43, 97-99.

Silverman, B.W., 1986. Estimation of Statistics and Data Analysis. Chapman & Hall: New York. doi:10.1007/978-1-4899-3324-9

Sperlich, Y., Sperlich, S., 2012. Practical Tools for Monitoring Convergence and Development focusing on Integration Areas. Journal of Comparative Policy Analysis 14, 72-103. doi:10.1080/13876988.2011.646822

Spolaore E., Warciarg, R., 2013. How Deep Are the Roots of Development. Journal of Economic Literature 51, 1-45. doi:10.1257/jel.51.2.325

Jones, C.D., Hughes, J.K., Bellouin, N., Hardiman, S.C., Jones, G.S., Knight, J., Liddicoat, S., O'Connor, F.M., Andres, R.J., Bell, C., Boo, K.-O., Bozzo, A., Butchart, N., Cadule, P., Corbin, K.D., Doutriaux-Boucher, M., Friedlingstein, P., Gornall, J., Gray, L., Halloran, P.R., Hurtt, G., Ingram, W.J., Lamarque, J.-F., Law, R.M., Meinshausen, M., Osprey, S., Palin, E.J., Parsons Chini, L., Raddatz, T., Sanderson, M.G., Sellar, A. A., Schurer, A., Valdes, P., Wood, N., Woodward, S., Yoshioka, M., Zerroukat, M., 2011. The HadGEM2-ES implementation of CMIP5 Centennial Simulations. Geoscientific Model Development 4, 543-570. doi:10.5194/gmd-4-543-2011

Appendix

Appendix: CO2 Data Set Sources

Data used in all figures and main:

CO2 emissions: annual maps of anthropogenic carbon emissions due to land use change (including wood harvest). The unit is g(C)$m^{-2}s^{-1}$. All maps have the same regular grid with a resolution of 0.5 degree.

Source:http://www.mpimet.mpg.de/ ... /landcover-change-emission-data.html
The Max Plank Institute has constructed CO2 maps based on the original work of Houghton (2008) from The Woods Hole Research Center, 149 Woods Hole Road, Falmouth, Massachusetts 02540, U.S.A.
For a detailed description please see: http://cdiac.ornl.gov/ ... /houghton.html

Population density: annual maps of population density. The unit is person /km^2. Within each region the emissions are weighted with the population densities also used in Klein Goldewijk (2000), which are linearly interpolated between the years 1850, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, and 1990. After year 1990 population density is assumed to stay constant.

Data used in section 4.3. Data robustness:

CO2 emissions: same as the one used for the main results.

Population density: Weighting fossil fuel emissions with population density is a common method (see e.g. Andres et al. 1996). Certainly, this approach is more problematic for land use change emissions than for fossil fuel emissions, because in urban centers most of the land use change had already been occurred in the previous times. Therefore, the population density has been reduced to a maximum of 20 persons per km2.
Source:http://www.mpimet.mpg.de/en/wissenschaft/land-im-erdsystem/wechselwirkung-klima-biogeosphaere/landcover-change-emission-data.html