The Role of Economic Geography in African Subnational Development
(Preliminary Version. Please, do not quote.)

Sebastian Seiffert
University of Hohenheim
Stuttgart, Germany
Phone: +49 711 459 23498
Email: seiffert@uni-hohenheim.de

August 14, 2015

Abstract

This contribution investigates the role spatial agglomeration of economic in explaining difference in comparative development of the African hinterland. In order to overcome the paucity with regards to disaggregated data concerning economic activity as well its spatial distribution, this approach relies on geo-referenced satellite data. Using information about the intensity of nocturnal lights at a spatial resolution of one square kilometre, it integrates the fields of spatial economics and the research concerning the fundamental causes of economic growth. It is shown that introducing measures of spatial dispersion in economic activity can help explaining a considerable amount of unobserved heterogeneity. I show that higher levels of spatial agglomeration are associated with significantly higher levels of local development in the African hinterland. These findings are robust to controlling for both national legal institutions as well as unobservable cross-ethnicity differences.

JEL classification: O10, O40, R12

Contents

1 Introduction

The transition from nomadic hunter and gatherer societies which were characterised by small groups with a high degree mobility into sedentary societies was set of by the adaption of agriculture as main source of livelihood. One highly important feature of this process was that those emerging sedentary societies transformed their natural environment to a large scale. This implies among others deforestation as well as the establishment of artificial irrigation systems in order to enhance the productivity of crop cultivation. This period also marked the emergence of permanent settlements in the form of solid structures and the increased utilisation of food-producing as well as food-storing technologies such as grain-mills and granaries. What alternation of the environment may it making the area more suitable for food-production or the installation of solid structures for food-storage or processing have in common is that they are costly, indivisible investment which are logically characterised by increasing returns to scale. This in turn implies that the aforementioned transition to agricultural societies should be characterised by an increase in the degree of co-location, i.e. densely populated systems. Such densely populated systems then in turn built the foundation for the emergence of the division of labour and specialisation which ultimately led to the development of complex hierarchical systems, institutions as well as enhanced knowledge generation and diffusion. All of which can be attributed to the category of the so-called "agglomeration economies" which have been found to be an important factor in explaining differences in comparative development in a short to medium term perspective by a number of studies[1]. Thus, it is natural to assume that such agglomeration economies (1) were already at play during the emergence of sedentary agriculture, (2) as settlements have been proven to be persistent, are persistent until today and as a consequence are an important aspect in explaining deeply rooted differences in comparative development in the spirit of Galor (2005) and others[2]. Today, countries differ largely in the extend to which they show patterns of agglomeration in many dimensions. Agglomeration patterns may occur in the context of population distribution (Brülhart and Sbergami (2009)) as well as in the context of industry clusters (Duranton and Overman (2005)). Both of which have been shown to have a significantly positive impact on comparative development via the channels of the aforementioned agglomeration economies. Accordingly, it is natural to ask which deeply rooted factors might have been shaping modern spatial organisation of economies as this could clearly contribute to a better understanding of the long term

[1] For a review, see Baldwin and Martin (2004).

[2] Some examples are: Michalopoulos and Papaioannou (2013a), Michalopoulos and Papaioannou (2013b), Alsan (2015) and Galor and Ozak (2014).

determinants of comparative development.

In this contribution, I propose an approach to explaining the emergence of different levels of spatial agglomeration based on the spatial distribution of land suitable for the cultivation of high starch, i.e. high caloric, crops. The main determinant in this context should be the relative spatial distribution of suitable land. Corollary, areas which are characterised by a more unevenly distributed potential for crop-cultivation should exhibit more uneven distributions of the local population and thus have profited more from the aforementioned agglomeration economies. This in turn should then manifest in higher levels of comparative development.

My analysis shows that the degree of unevenness in the spatial distribution of suitable land is significantly correlated to the modern spatial distribution of population as well as the current level of economic measured by the density of nocturnal lights emission. The results are robust to controlling for a wide array of potentially confounding variables such as colonial institutions, the occurrence of natural resources and the degree of complexity and hierarchy of pre-colonial ethnic institutions which have been shown to be an important factor in the underlying context by Michalopoulos and Papaioannou (2013 a & b).

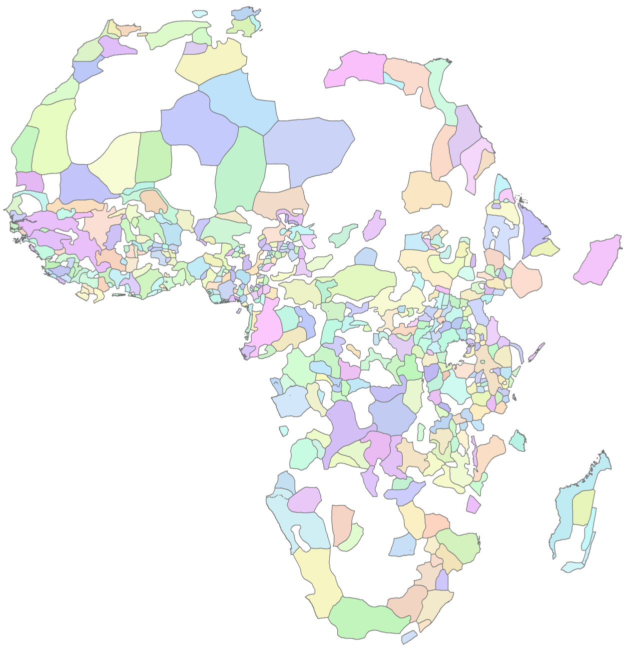
2 Data

As their contribution built the ideal setting for the pursuit of my research question, I largely rely on the data used in Michalopoulos and Papaioannou (2013a).

2.1 Identifying Ethnicities

As the aim of this study is to identify the long-run impact of an uneven spatial distribution of agricultural suitability. relying on the national borders of the African continent is not suitable. The reason is that these borders were basically drawn exogenous to ethno-cultural differences as they relied mainly on the areas of influence of former colonial powers (Michalopoulos and Papaioannou (2013a)). The authors among others have shown that national institutions on average have no significant explanatory power regarding comparative development once controlling for ethnicity-specific (un-)observables. In a companion paper Michalopoulos and Papaioannou (2013b) show that differences in economic development show a strong positive relationship to pre-colonial political centralisation. Therefore, I rely on the ethnographic Atlas by Murdock (1959) and take a sub-sample of 431 African ethnic groups[3]. Figure 1 gives an overview.

Figure 1: Ethnic Areas on the African continent.



2.2 Soil Suitability Data

As one of the main points of interest in this contribution is the spatial distribution of crop suitability, I need a fine-grained geo-refferenced data set. This is provided by the Food and Agriculture Organisation's (FAO) Global Agro-Ecological Zones project (GAEZ). The data is distributed in the form a 5'x5' (ca. 100 km^2) grid size raster data set. Each grid cell is assigned a value between 1 (very high suitability) to 8 (not suitable) Namely, I rely on data about the potential suitability for cereal crops under a low-input (i.e. low agricultural technology), rain-fed regime averaged over the years 1961 - 2000. It is important to note that the data is not about actual yields but the potential yields which are determined by agro-ecological constraints which should have been fairly constant at least over the last 200 years. Figure 2 depicts the raster data set.

[3]Note that this subsample does not entail all ethnic areas on the African continent. The sample will soon be completed

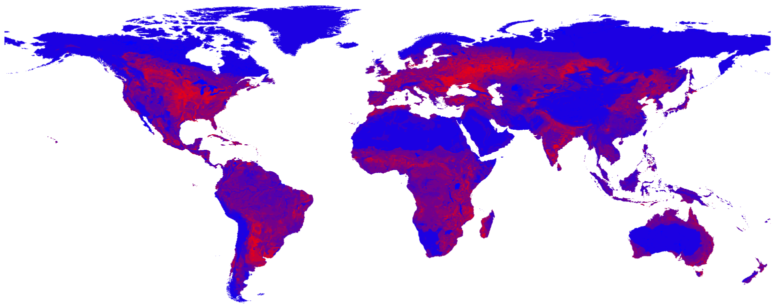


Figure 2: Global Distribution of Potential Yields from Cereal Crops
(Red indicates High Potential Yields).

2.3 Population Count Data

The Gridded Population of the World project (GPW) provides data about the estimated population count at a spatial resolution of 2.5 arc minutes (ca. 5 km^2). The data is constructed from national and subnational inputs. Figure **??** represents the data set for Africa. At this point, it is important to note that in the mapping process, none of the other data used in this study is involved. Accordingly, there is no source of multicollinearity.

2.4 Measuring Regional Economic Activity in Africa

While there exist some national accounts data on economic activity at the national level for the African continent, it is often of dubious quality. Once moving to lower levels of spatial aggregation, there are hardly any reliable sources which one could utilise. Accordingly, analyses of regional economic activity in African countries hinges on the existence of appropriate alternatives to national accounts data. Following Henderson et al. (2011) and Michalopoulos and Papaioannou (2013a), I use satellite images in nocturnal light

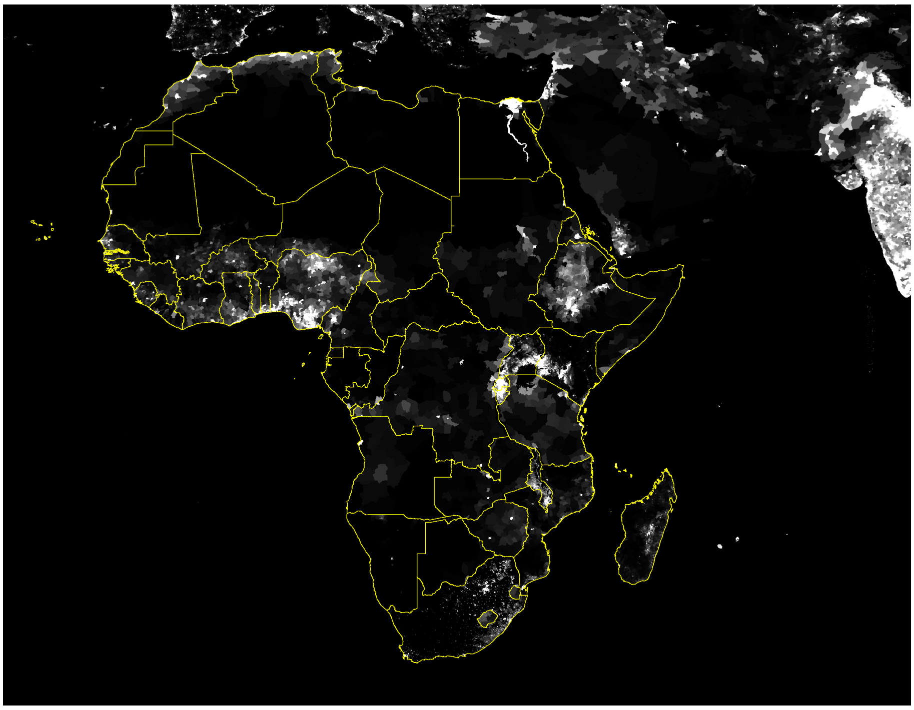
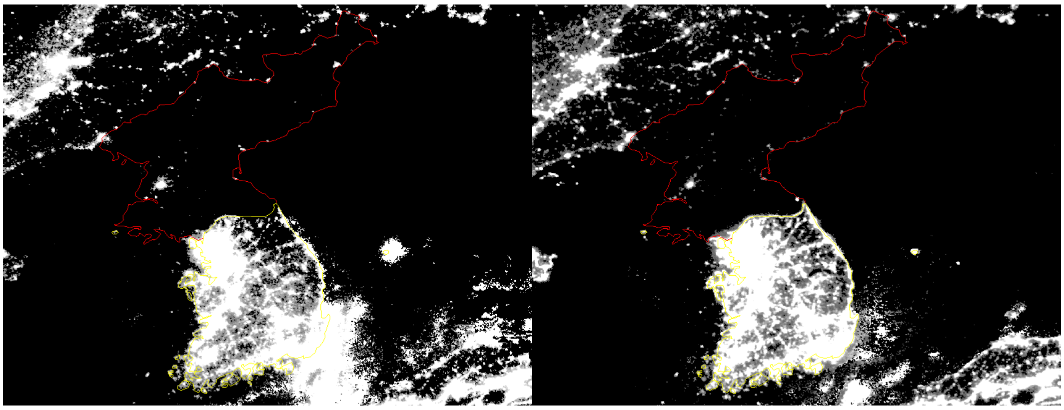


Figure 3: Population Count and African National Borders.

emission.

The underlying data is made available by the *Defense Meterological Satellite Program's Operational Line Scan System (DMSP-OLS)* of the United States Airforce. The pictures cover almost the entirety of the world's surface. They entail information about illumination intensity recorded between 20:00 and 21:30 local time. While the pictures are taken every night, the final products display year averages. The data is distributed as geo-referenced raster data with a grid resolution of 30 arc seconds per pixel. Due to the ellipsoidal nature of the globe, this equivalent to approximately one square kilometre depending on the exact location. The recorded illumination intensity is coded in integer values from 0 (no light emission) to 63 (sensor satiation).

Besides man-made light from human settlements, the satellites also record light emission from natural events such as northern lights or forest fires as well as gas flares in areas of crude oil extraction. This is a source of unfavourable noise if one is aiming at using nocturnal lights as a proxy for economic activity. While temporal events like forest fires do not impose a problem as they are removed during the per year averaging process, the persistent noise has to be accounted for prior to using the data. I did this using a global map of gas flares and replaced all illumination values of the respective pixels by zero values. Once the unfavourable noise has been removed, one is left with a com-



(a) North and South Korea in 1992 (b) North and South Korea in 2006

Figure 4: North Korea versus South Korea

prehensive, high-resolution map of economic activity. Figure4 illustrates the situation using the North and South Korea as an example. While the stable relationship between national-level GDP figures and nocturnal lights has been proven by Henderson et al. (2011) and others, Michalopoulos and Papaioannou (2013a) use individual survey data from the Afrobarometer to show that night lights strongly correlate with public goods provision in the nature of access to sewage systems or electrification at all levels of spatial aggregation for the African continent. This further underlines the appropriateness illumination data in the underlying context.

As suggested by Michalopoulos and Papaioannou (2013a), I use average illumination intensities of the years 2007 and 2008. This allows to compute the average light intensity per pixel for the 431 ethnic homelands described in section 2.1.

2.5 Further Variables

As the aim of this study is to assess in how far the spatial distribution of suitable soil may have shaped modern population distributions and with it comparative development, I control for an array of potential confounding factors. These are measured by the additional variables listed in table 1.

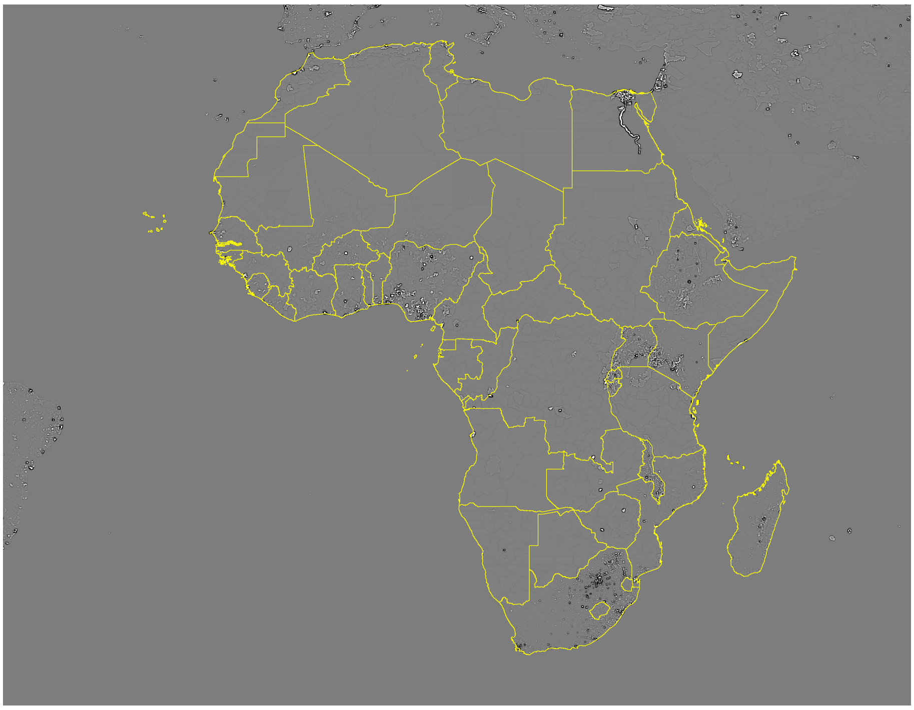


Figure 5: Slope Position Index of Population Count in Africa

3 Identification

3.1 Introducing a Novel Measure of Spatial Concentration.

Commonly used measures to assess difference in the degree of equality in distributions are Lorenz-curve based such as the Gini coefficient. Those indices measure inequality in the distribution of some parameter across a population.

For the underlying application, those measures are not completely adequate as they give some impression about the overall inequality in the respective values but do not really measure whether it may be concentrated in some subset of contingent cells of space. In order to identify spatial concentration adequate to the goal of this study, I need a measure which describes the unevenness in the value of interest relative to the neighbouring grid cells. Therefore, I rely on a method commonly used in the field Geomorphology to classify landscapes with respect to their topographic structure: the Slope Position Index introduced by De Reu et al. (2013). This index measures a grid cells topographical level relative to its neighbouring area. Accordingly, positive values indicate local peaks while negative values indicate local valleys. Using the standard deviation of this index gives the variability in the slope of a terrain. Hence, areas with a high standard deviation in the slope position index are characterised by a high degree of concentration, and vice versa. Figure 5 gives an example for the African continent. I compute this index for both the population count as well as the suitability for cereal crops and use them as a measure of spatial concentration.

3.2 Estimation Framework

The empirical analysis will take place in three steps: First, I will assess the relationship between the level of comparative development and the spatial concentration of population as well as the relationship between the degree of spatial concentration of population and the degree of spatial concentration in suitability for cereal crops. In a final step, I will then use the degree of spatial concentration of cereal suitability as instrument for the degree of spatial concentration of population in order to estimate its effect on comparative development.

In order to assess the relationship between economic development and spatial concentration of population, I will estimate the following equation:

$$Y_i = \alpha PSI_i + X'\beta + +u_i \tag{1}$$

PSI_i denotes the level of spatial concentration in population, X' is a vector of controls. u_i is the error term. As it is possible that the level of economic development influences the spatial arrangement of population, there exists a source for potential reverse causality. In order to overcome that issue, I will run an IV regression where I instrument PSI_i with the degree of spatial concentration of cereal suitability as well as the average level of suitability across the respective ethnic homeland.

Table 2 presents the results of assessing the relationship between the different concentration measures and comparative development measured by illumination intensity. In both cases, we can observe a significant positive relationship between concentration and development. The relationships are robust to a variety of locational and geographic controls listed in table 1 as well as the inclusion of both regional and language family fixed effects.

4 Conclusion

References

Alsan, M. (2015). The Effect of the TseTse Fly on African Development. *American Economic Review*, 105(1):382–410.

Baldwin, R. E. and Martin, P. (2004). Agglomeration an Regional Growth. In *Handbook of Urban and Regional Economics*, volume 4, chapter 60, pages 2672–2709.

Brülhart, M. and Sbergami, F. (2009). Agglomeration and Growth: Cross-country Evidence. *Journal of Urban Economics*, 65(1):48–63.

De Reu, J., Burgeois, J., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P., Van Merveinne, M., Verniers, J., and Crombe, P. (2013). Application of the topographic position index to heterogeneous landscapes. *Geomorphollogy*, 186:39 – 49.

Duranton, G. and Overman, H. G. (2005). Testing for Localization Using Micro-Geographic Data. *Review of Economic Studies*, 72(4):1077–1106.

Galor, O. (2005). From Stagnation to Growth: Unified Growth Theory. In *Handbook of Economic Growth*, volume 1, Part A, pages 171–293. Elsevier.

Galor, O. and Ozak, O. (2014). The Agricultural Origins of Time Preference. *Working Papers*.

Henderson, V., Storeygard, A., and Weil, D. N. (2011). A Bright Idea for Measuring Economic Growth. *American Economic Review*, 101(3):194–99.

Michalopoulos, S. and Papaioannou, E. (2013a). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics*, 129(1):151–213.

Michalopoulos, S. and Papaioannou, E. (2013b). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, 81(1):113–152.

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

Table 1: Summary Statistics

Variable	(1) Obs.	(2) Mean	(3) St. Dev.	(4) p50	(5) Min	(6) Max	(7) Source
Light Density	431	0.413	1.896	0	0	16.58	NOAA
Ln(Lights)	431	-3.486	2.281	-3.486	-6.907	2.962	NOAA
Population Slope Index (PSI)	431	298	749	74.6	0.322	1237	GPW
Ln(PSI)	431	4.273	1.867	4.312	-3.176	9.327	GPW
Cereal Suitability Slope Index (CSI)	431	0.165	0.258	0.055	0	1.861	FAO GAEZ
Ln(CSI)	431	-2.063	0.762	-1.932	-4.411	-0.629	FAO GAEZ
Average Cereal Suitability (ACS)	431	4.975	1.066	4.864	2.975	8.845	FAO GAEZ
Ln(ACS)	431	1.582	0.209	1.581	1.090	2.179	FAO GAEZ
Average Altitude	431	656	481	523	19	2242	NOAA
Terrain Ruggedness (TR)	431	656	481	523	19.10	2242	NOAA
ln(TR)	431	6.127	0.997	6.260	2.950	7.715	NOAA
Average Annual Precipitation (1961-2000)	431	1133	588	1128	24.85	3148	FAO GAEZ
Average Temperature (1961 -2000)	431	24	3.289	24	15	29	FAO GAEZ
Distance to Closest Seashore (in km)	431	585	428	545	10	1651	NOAA
Water Area	431	861	2134	205	0	2163	ESRI
Area	431	39.56	67.99	15.18	1.214	604	Murdock (1959)
Ln(Area)	431	2.835	1.271	2.720	0.194	6.405	Murdock (1959)
Malaria Stability Index	431	0.751	0.313	0.910	0.001	1	M+P (2013b)
Diamond Deposit (Dummy)	431	0.136	0.343	0	0	1	M+P (2013b)
Oil Deposit (Dummy)	431	0.111	0.314	0	0	1	M+P (2013b)

Table 2: Spatial Concentration and Comparative Development across Ethnic areas

	(1) PSI_i	(2) CSI_i
Agglomeration	0.705***	0.299*
Robust SE	(0.061)	(0.163)
R-squared	0.613	0.463
Observations	431	431
Region FE	Yes	Yes
Language Family FE	Yes	Yes
Precol. Centralisation	Yes	Yes
Location Controls	Yes	Yes
Geographic Controls	Yes	Yes