



# Working Papers

[www.cesifo.org/wp](http://www.cesifo.org/wp)

## Regional Inequality, Convergence, and its Determinants – A View from Outer Space

Christian Lessmann  
André Seidel

CESIFO WORKING PAPER NO. 5322

CATEGORY 6: FISCAL POLICY, MACROECONOMICS AND GROWTH  
APRIL 2015

*An electronic version of the paper may be downloaded*

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: [www.CESifo-group.org/wp](http://www.CESifo-group.org/wp)

ISSN 2364-1428

# Regional Inequality, Convergence, and its Determinants – A View from Outer Space

## Abstract

This paper provides a new data set of regional income inequalities within countries based on satellite nighttime light data. We first empirically study the relationship between luminosity data and regional incomes for those countries where regional income data are available. We subsequently use our estimation results for an out-of-sample prediction of regional incomes based on the luminosity data, which allows us to investigate regional income differentials in developing countries as well, where official income data are lacking. Based on the predicted incomes, we calculate commonly used measures of regional inequality within countries. Investigating changes in the dispersion of regional incomes over time reveals that approximately 71-80% of all countries face sigma-convergence. Finally, we study different major determinants of the level of regional inequality based on cross-section data. Panel regressions investigate the within-country changes in inequality, i.e., the determinants of the convergence process. We find evidence for an N-shaped relationship between development and regional inequality. Geography, mobility and trade openness are also highly important.

JEL-Code: D300, E010, E230, O110, O150, O570, R100.

Keywords: regional inequality, spatial inequality, sigma convergence, panel data, luminosity data, economic development, trade, ethnic fractionalization.

*Christian Lessmann*  
*Institute of Economics*  
*TU Braunschweig*  
*Spielmannstr. 9*  
*Germany – 38106 Braunschweig*  
*lessmann@tu-braunschweig.de*

*André Seidel*  
*TU Dresden*  
*Department of Business and Economics*  
*Chair for Public Economics*  
*Germany - 01062 Dresden*  
*andre.seidel1@tu-dresden.de*

---

## 1 INTRODUCTION

In recent decades, the regional distribution of incomes within countries has attracted considerable interest among academics and policy makers. The following are important research questions, among others: What are the consequences of regional inequality? What are the determinants? Are regional inequalities transient or permanent? How do interregional inequalities relate to interpersonal income inequalities, ethnicity, and geography? Because these questions are obviously important for the economy in particular and the entire society in general, many empirical studies on these issues have been carried out with interesting and instructive results. However, all these studies have in common that they are limited to a particular country sample with a general bias toward middle- and high-income economies. The major difficulty with this research is the availability of regional income data. While it is easy to obtain regional data for developed countries through the regional statistics of the OECD and other public sources, doing so becomes difficult if less developed regions of the world are under study. Recent studies by Gennaioli et al. (2013, 2014) and Lessmann (2014) have made great progress in this field providing cross-country data sets for up to 110 countries and panel data for 83 countries and 56 countries, respectively. However, the poorest regions of the world are still blank spots on the map, which limits the informative value of the results to the particular country samples. The aim of our study is to fill this gap by using satellite nighttime light data as a proxy for regional incomes.

Our approach is based on Henderson et al. (2012), who show a striking relationship between changes in nighttime light intensities and economic growth at the country level. The idea is the following: Most economic activities in form of consumption and production, which take place in the evening or at night, require light. We can expect that the higher the nighttime light intensity is, the higher the level of economic activity will be, i.e., the higher the income. Therefore, the luminosity data measured by satellites can be used as a proxy for income in those parts of the world for which we have no reliable statistical data. This issue may be minor for country-level income data (see also Johnson et al. 2013) but is certainly major at the regional level. Regional data are often lacking in developing countries, which have low capacities and standards in their statistical authorities. For this reason, Chen and Nordhaus (2011, 2015) show how nighttime luminosity data can be used to improve estimates of output per grid cell ( $1^\circ$  latitude  $\times$   $1^\circ$  longitude), particularly in this group of countries. Consequently, recent studies, such as Besley and Reynal-Querol (2014) or Hodler and Raschky (2014), make use of these sorts of data to proxy regional income levels. We follow this literature to construct new data on regional income inequality within countries.

The key innovation of our study is the prediction of regional inequality based on satellite nighttime light data. We first use the luminosity data to predict regional incomes per capita at a subnational level for 180 countries for the period 1992–2012. Based on the predicted regional incomes, we then calculate measures of regional inequality within countries. Our prediction of regional incomes with nighttime light data refers to Gennaioli et al. (2014), who provide data on regional incomes and regional light intensities for 82 countries. Based on the statistical relationship between light and income in this data set, we predict regional income for all the countries in the world. Different robustness tests imply that our income predictions are informative.

Using the regional incomes, we calculate different measures of regional inequality within countries. We provide the coefficient of variation of regional GDP p.c., the Gini coefficient, and the population-weighted coefficient of variation.

---

The latter has been suggested by Williamson (1965) for cross-country comparisons because it takes into account the heterogeneity of regions in terms of differences in the sizes of regional populations. The inequality measures are used in several ways. We first describe the differences in the level of regional inequality across countries, finding that the development level is relevant. Our data suggest that the most heterogeneous countries are located in East Asia, Latin America, and Africa, while industrial economies usually have much lower regional inequalities. The data of Sub-Saharan Africa are particularly interesting because this region has been a white spot in the existing literature<sup>1</sup> We find that very poor countries in the center of the continent have relatively low regional inequalities at the subnational level. Higher-income countries, such as Namibia or South Africa, have relatively low inequalities. However, those countries that could be called middle-income countries in their respective country groups, such as Namibia or Zambia, have relatively high regional inequalities. This result supports previous evidence of an inverted U-shaped relationship between economic development and regional inequality [see Williamson (1965), Barrios and Strobl (2009), and Lessmann (2014)]. Importantly, our data on the very poorest countries in the world allow us to estimate the upward sloping part of the Kuznets curve in regional inequalities.

Next, we use the data on regional inequality to analyze sigma-convergence within countries. While beta-convergence focuses on the (better) growth performance of initially poor regions within a country, sigma-convergence refers to the decrease in the dispersion of regional incomes. By comparing the changes in our inequality measures between the periods 1992–2001 and 2002–2011, we find that more than 70% of all countries face sigma-convergence. However, a significant number of countries have increasing inequalities – among both developing countries (e.g., Mozambique and Bangladesh) and industrial economies (e.g., Hungary and South Korea).

Finally, we use our data to investigate the determinants of regional inequality and sigma-convergence. We start with the empirical model by Lessmann (2014), who analyzes the inverted U relationship between regional inequality and development. Using cross-sectional regressions with long period averages, we find support for the inverted U but with increasing inequalities at very high levels of economic development. Therefore, the relationship is ultimately N-shaped in our data, i.e. an inverted U with another increase in inequality after the inverted U pattern has been completed. This is in line with the findings of Amos (1988) for the case of U.S states. We also investigate several geographic, political, international, and social determinants finding many robust results, which we cannot discuss here in detail. The determinants considered are related to territorial fragmentation, mobility, openness, institutions, and other factors. Concerning sigma-convergence, we apply panel fixed-effects regressions, which focus on the within-country variations in the data. Again, we find an N-shaped relationship between economic development and changes in regional inequality. Moreover, increased trade openness is significantly related to the divergence of regions. Concerning this issue, a conflict between income growth and distribution might occur. In contrast, democratic institutions promote sigma-convergence. In many respects, our results support earlier studies in the field. Based on our data set, we are also able to analyze regions of low-income countries, which helps us generalize these findings. Our empirical results do not suffer from a potential sample selection bias. Note, however, that we conduct only OLS regressions in this part of the analysis; therefore, our results only document statistical correlations, not necessarily causal relationships. Hence,

---

<sup>1</sup> An interesting study on African regions provides Mveyange (2015) who uses the dispersion of grided nighttime lights as a proxy for the dispersion of grid income at the district level.

---

we abstain from policy recommendations and explicitly encourage researchers to investigate this and other relevant issues in more detail.

The remainder of the paper is organized as follows. In section 2, we explain the methodology that we apply to construct regional income proxies from luminosity data. Thereby, we also explain several important measurement issues that are relevant when working with satellite data in a regional context. In section 3, we calculate different measures of regional inequality within countries that are commonly used in the literature. We compare differences across countries, and we analyze sigma-convergence within countries. In section 4, we regress the inequality measures on selected explanatory variables to investigate the determinants of the level of regional inequality and the changes in inequality over time. In section 5, we summarize our main findings and conclude.

## 2 PREDICTING REGIONAL INCOMES WITH LUMINOSITY DATA

The main focus of our study is the construction of measures of regional inequality for those countries where reliable regional income data are lacking. The first step of our analysis is the prediction of regional incomes. For this purpose, we use satellite nighttime light data. We proceed as follows: (1) We explain the underlying data sets and measurement issues, (2) we explain our methodology and estimation results, and (3) we discuss our findings by comparing predicted regional incomes with actual incomes. Note that we use income, output, and GDP synonymously because it is impossible to distinguish between consumption and production with the underlying data. We assume that all the relevant variables are highly correlated. Hence, our constructed variables should be a fairly good proxy for the level of regional economic development.

### 2.1 DATA

#### *NIGHTTIME LIGHT*

Various studies use nighttime light emissions to measure socioeconomic variables, particularly in developing countries. Elvidge et al. (1997) show that light intensities are highly correlated with the GDP at the country level ( $R^2=0.97$ ). More recently, Henderson et al. (2012), in what is likely the most prominent study in this area of research, relate changes in nighttime light to economic growth at the country level, finding a correlation of approximately 70%. The findings of Chen and Nordhaus (2011, 2015) support the hypothesis that luminosity can be even more informative as a proxy for output than standard output proxies from sources such as G-Econ and World Bank. They show, on the national and subnational levels (grid cells of  $1^\circ$  latitude  $\times$   $1^\circ$  longitude), that this is likely the case in countries with low-quality statistical systems and no recent population or economic censuses. Furthermore, for small samples of developed countries studies have shown that light can also be a good proxy for regional development (Ebener, et al., 2005; Ghosh et al., 2010; Sutton et al., 2007). These findings support our approach of using luminosity data as a proxy for output on a regional level where standard sources are not available or of poor quality.

The data that are used to measure nighttime light intensities come from meteorological satellites of the US Air Force. The satellites orbit the earth 14 times a day measuring Earth lights between 8:30 and 10:00 pm. Scientists at the

---

National Oceanic and Atmospheric Administration (NOAA) and National Geophysical Data Center (NGDC) process the satellite data and distribute them to the public. Several manipulations of the raw data are necessary to obtain comparable values [see Henderson et al. (2012) for details]. The adjustment of the raw data is necessary to compensate, for example, for the local coverage with clouds, dust in the atmosphere, and changes in satellite and sensor technology. The aim of these operations is to measure only man-made lights as precisely as possible. The final grid datum is a digital number between 0 (no light) and 63 for every 30 arc-second output pixel, which is approximately 0.86 square kilometer (at the equator). The final data are available from 1992 onward on an annual basis.

The literature notes several measurement problems that we must take into account when using nighttime light intensities as a proxy for output. One point of concern that is often discussed is the censoring of the data at 63 for rich and densely populated areas. This is only the case for a small fraction of pixels. However, it is noteworthy that in a few areas there is no variation in the luminosity data, while there are significant differences in socioeconomic variables. For example, the metropolitan areas of New York City and Mexico City are top-coded, although there are large income differences between both regions. Not least for this reason, we may expect that the luminosity data is a less appropriate proxy for income in rich countries than in poorer ones. Of course, one might also expect that very rich countries have better technologies that allow for less energy-intensive production processes. We compensate for these effects within countries by considering national output as a development proxy when predicting regional output. There are also measurement problems at the other end of the distribution of light intensities. In deserts or mountains, many pixels are coded zero, although a small number of people with positive incomes may still live there. We follow Hodler and Raschky (2014) by adding 0.01 in those regions where the average nighttime light per pixel would be zero otherwise. Another problem concerns changes in satellite and sensor technology that affect the sensitivity of the light measurement. We address this issue – because it is common in the literature – by controlling for satellite fixed effects.<sup>2</sup> Apart from the data manipulations by NOAA scientists and our slight adjustments, we do not make any changes to the raw data.<sup>3</sup>

#### *SUBNATIONAL BOUNDARIES*

The Global Administrative Areas (GADM) project is a spatial database of the locations of administrative areas (or administrative boundaries), which we use to aggregate gridded data to sub-national regions. Thereby, we refer to the 1st subnational administrative level, which are, e.g., states, provinces, cantons, or Bundesländer. In most instances, the territorial level is similar to OECD TL2 regions or EUROSTAT NUTS1 regions. However, the regions are quite heterogeneous in terms of area, population size, political power, climate, and geography. This point becomes critical when calculating country-level measures of regional inequality (see section 3 for details). Our final data set is based on 180 countries and 3,163 regions.

---

<sup>2</sup> We refrained from using annual dummies because doing so would result in a bias in our GDP forecasts for the years 2011 and 2012, for which we do not have all the relevant data.

<sup>3</sup> We use QGIS Desktop 2.4.0 with the plugin Zonal Statistics 0.1 to analyze the data.

We use population data from Gridded Population of the World (GPW) v.3 provided by the Center for International Earth Science Information Network (CIESIN). Similar to the luminosity data, the global population density is published in TIF maps. Using the population density and the size of regions, we calculate the total regional population. Note that we round the results up at the regional level; therefore, the minimum population of a region is one. The frequency of the original data is in 5-year waves; therefore, we interpolate missing values to obtain annual data. Hence, short-term fluctuations in the population, for example, caused by conflicts and natural disaster-induced migration, are not covered by the data.

Table A.1 in the appendix provides summary statistics for all the variables generated from the geo-coded data. Note that we refer only to those countries that are ultimately included in our empirical analysis.

## 2.2 PREDICTION OF REGIONAL INCOMES

In this section, we investigate the relationship between regional income levels and regional nighttime lights. For this purpose, we use the data set provided by Gennaioli et al. (2014), which contains information on both average light emission per region and regional income. The data set includes 82 countries (1,503 regions) and covers an average time span of 32 years. Note that the panel is highly unbalanced. Income data refer to the GDP per capita (in constant PPP US\$) and are collected from different sources, including international official statistics (e.g., OECD Stat.), national statistics, and single country reports, such as the human development reports. Therefore, the quality of the regional data varies across countries. In particular, in less developed countries, we might expect a larger measurement error in the regional data [see Chen and Nordhaus (2011, 2015) for a detailed discussion of the quality of regional data]. The luminosity and income data are collected at the state or province level; however, there are a few cases in which the administrative boundaries differ from the GADM data, which we use in our later analysis. For example, in the GADM data, Great Britain is separated into four sub-national units: England, Northern Ireland, Scotland, and Wales, while the data provided by Gennaioli et al. (2014) use the EUROSTAT NUTS1 / OECD TL2 definition, where England is subdivided into nine regions totaling 12 regions for Great Britain. We address this issue in the forthcoming regression analysis by controlling for the number of regions by country and country size (in sq. km). Table A.2 in the appendix provides the summary statistics of the relevant variables in the Gennaioli et al. (2014) data.

### 2.2.1 METHODOLOGY

We exploit the variations in the data within and across countries and regions to obtain an estimate for the relationship between light and income. We need these results as inputs in our forecasting model of regional income for those countries where the regional income data are not available but satellite data as a proxy are. For this purpose, we estimate in a first step the following random effects model:

$$\ln(y_{i,j,t}) = \alpha + \beta \cdot \ln(l_{i,j,t}) + \gamma \cdot \ln(y_{j,t}) + \delta \cdot \ln(r_j) + \zeta \cdot \ln(a_j) + \eta_k + \theta_t + v_i + \varepsilon_{i,j,t}, \quad [1]$$

where  $y_{i,j,t}$  ( $i = 1, 2, \dots, n_j, j = 1, 2, \dots, N, t = 1, 2, \dots, T$ ) is the GDP per capita of region  $i$  in country  $j$  at time  $t$ ,  $l_{i,j,t}$  is the average nighttime light within a region,  $y_{j,t}$  is the country-level income,  $r_j$  is the number of regions within a country,  $a_j$  is the overall size of a country in square kilometers,  $\eta_k$  are time-invariant fixed effects for  $k$  different lending group regions of the world as defined by the World Bank (North America is used as reference group),<sup>4</sup>  $\theta_t$  is the satellite configuration fixed effects, which change over time (but not necessarily on an annual basis),  $v_i$  is a regional random effect with  $E[v_i] = 0$  and  $\varepsilon_{i,j,t}$  is the error term. We use all the variables in logarithmic transformation; therefore, the coefficients can be interpreted as elasticities. Note that we always report clustered standard errors. We cluster standard errors at the country level because the errors are correlated within countries.

The GDP p.c. is measured in country-level real power purchasing parities (PPP) in constant 2005 US\$.<sup>5</sup> An important measurement problem may arise from differences in regional price levels. The question is whether the regional inequalities measured based on the country-level prices are real in terms of the regional price levels. Regional prices for non-tradable goods, services, housing, and other commodities will adjust to the respective regional income levels. Therefore, Gennaioli et al. (2014) analyze the differences in income with differences in housing prices and durable goods. They find that the regional price adjustments dampen differences in nominal regional GDP but that “real GDP per capita is far from being equalized across regions” [Gennaioli et al. (2014), p. 280]. Furthermore, in our analysis, the potential source of bias is even less severe because we focus on the differentials in regional incomes predicted from the luminosity data. The variation in the predicted regional incomes within countries comes entirely from the variation in regional light emissions, which we use as a proxy for real regional income.

The estimated parameters  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\gamma}$ ,  $\hat{\delta}$ ,  $\hat{\zeta}$ ,  $\hat{\eta}$ , and  $\hat{\theta}$  are used in a second step to predict regional incomes  $\hat{y}_{i,j,t}$ :

$$\ln(\hat{y}_{i,j,t}) = \hat{\alpha} + \hat{\beta} \cdot \ln(l_{i,j,t}) + \hat{\gamma} \cdot \ln(y_{j,t}) + \hat{\delta} \cdot \ln(r_j) + \hat{\zeta} \cdot \ln(a_j) + \hat{\eta}_k + \hat{\theta}_t. \quad [2]$$

We stress that all the variables that are necessary to predict regional incomes are available for literally all the countries in the world. Our predicted regional incomes  $\hat{y}_{i,j,t}$  ultimately depend on the different country-level variables, the fixed effects, and the regional luminosity  $l_{i,j,t}$ .

Finally, predicted country-level income is calculated based on the predicted regional income by using

$$\hat{y}_{j,t} = \frac{1}{p_{j,t}} \sum_{i=1}^{n_j} p_{i,t} \hat{y}_{i,j,t}, \quad [3]$$

<sup>4</sup> The country groups are: East Asia and the Pacific (EAP), Europe and Central Asia (ECA), Latin America and the Caribbean (LAC), Middle East and North Africa (MENA), North America (NA), South Asia (SA), and Sub-Saharan Africa (SSA).

<sup>5</sup> The national GDP reported by Gennaioli et al. (2014) was aggregated from regional data. In some cases, the aggregated GDP differs from the GDP reported by the World Bank; therefore, the regional GDPs have been rescaled to fit with the national levels. Unfortunately, we cannot reproduce these adjustments, which would be necessary to compare our income predictions with national levels. Thus, if we compare our predicted income data with observed data, there is a small difference in levels, which, importantly, does not affect the distribution of income within countries.



---

where  $P_{j,t}$  is the total population of country  $j$ , and  $p_{i,t}$  is the population of region  $i$ . This is an important reference value that we will need later to compare our predicted incomes with observed country-level data.

Our empirical model needs further discussion. In particular, we need a justification for the application of a random-effects model instead of a country (or regional) fixed-effects model. The fixed-effects model is a reasonable approach when the differences between countries (or regions) can be viewed as parametric shifts of the regression function. This usually applies to cross-country panel data models, where the respective dummy variables capture the country-specific unobserved heterogeneity between countries. Also in our context, it is reasonable to expect that political factors, history or geography have a constant impact on the relation between country-level incomes and average luminosity. However, the regression coefficients received from the fixed-effects model depend on the individual intercepts. In our forecasting model used to predict regional incomes (equation [2]), we cannot consider these variables because we have no reliable proxy for the individual country fixed effects. The advantage of the random-effects model is that the expected value of the country-specific effect is zero; therefore, we need not apply any arbitrary data imputation procedure for the missing intercepts.<sup>6</sup> This approach may, however, come at the cost of founding the predictions on a slightly biased coefficient. We show below that the major coefficient of interest,  $\hat{\beta}$ , is not sensitive to applying a fixed-effects model or a random-effects model with additional country and region information. Our specification is a compromise between random and fixed effects: we control in our random effects model for several country-level fixed factors (national income, number of regions, and area) and fixed effects for different country groups.

## 2.2.2 ESTIMATION RESULTS

The results of different specifications of eq. [1] are reported in table 1. Each column reports the results we obtain when we progressively add fixed effects and other variables to our model. Column (1) shows the results of a random-effects model, where we simply regress the regional income  $y_{i,j,t}$  on average regional nighttime lights  $l_{i,j,t}$ . Column (2) adds region fixed effects and satellite fixed effects. Column (3) considers country fixed effects and satellite fixed effects. Column (4) adds the country-level GDP p.c. Columns (5)-(10) consider country-group fixed effects and satellite effects instead of country fixed effects. Column (5) shows the results that we obtain when only considering luminosity and fixed effects. Column (6) adds the country-level income  $y_{i,t}$ . Column (7) factors in the (country-level) number of regions and area. In the last specifications reported in columns (8) – (10), we split our sample into low-, middle- and high-income countries following the current World Bank definition for 2015.<sup>7</sup>

---

<sup>6</sup> Note that we also estimated fixed-effects models, where we proxy the individual intercepts with the “average intercept” of those countries, which were comparable in terms of income and geography.

<sup>7</sup> Low-income countries:  $y_j < 4.086 \text{ US\$}$ ; middle income countries:  $4.086 \text{ US\$} \leq y_j < 12.615 \text{ US\$}$ ; high income countries:  $y_j \geq 12.615 \text{ US\$}$ .

**Table 1:** Regression results of equation [1] using panel data provided by Gennaioli et al (2014)Dependent variable:  $\log(\text{GDPpc}_i)$ 

	Pooled, Region FE, Country FE				Country Group FE					
	Full sample				Full sample		LIC	MIC	HIC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\log(\text{Light}_i)$	0.370*** (0.010)	0.181*** (0.014)	0.173*** (0.010)	0.117*** (0.010)	0.252*** (0.011)	0.106*** (0.009)	0.115*** (0.009)	0.118*** (0.017)	0.173*** (0.016)	0.008 (0.007)
$\log(\text{GDPpc}_i)$				0.772*** (0.028)		0.887*** (0.015)	0.883*** (0.015)	0.949*** (0.055)	0.715*** (0.045)	0.975*** (0.023)
$\log(\# \text{ Regions}_i)$							-0.137*** (0.020)	-0.072* (0.039)	-0.197*** (0.034)	-0.045** (0.018)
$\log(\text{Area}_i)$							0.054*** (0.009)	0.068*** (0.015)	0.078*** (0.012)	-0.011 (0.011)
Constant	8.592*** (0.027)	8.916*** (0.017)	8.746*** (0.087)	1.870*** (0.261)	10.278*** (0.044)	0.980*** (0.157)	0.645*** (0.186)	-0.630 (0.502)	1.523*** (0.422)	0.545* (0.292)
# Observations	5,198	5,198	5,198	5,198	5,198	5,198	5,198	1,036	2,279	1,883
# Regions	1,484	1,484	1,484	1,484	1,484	1,484	1,484	337	1,060	1,219
# Countrys	80	80	80	80	80	80	80	25	30	25
R-squared within	0.258	0.525	0.522	0.658	0.521	0.656	0.657	0.667	0.537	0.811
R-squared-between	0.317	0.317	0.887	0.885	0.570	0.865	0.872	0.688	0.326	0.697
Region FE	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO
Country FE	NO	NO	YES	YES	NO	NO	NO	NO	NO	NO
Country Group FE	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
Satellite FE	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Standard errors are clustered at the country level.

---

In nine of ten specifications, the coefficient of the luminosity variable is statistically significant and positive as expected. Nighttime light emissions at the regional level are highly correlated with regional incomes. In our first specification, reported in column (1), we find a coefficient of 0.370 implying that a 10% increase in luminosity is associated with a 3.7% increase in regional income. The results reported in columns (2) and (3) show, however, that the coefficient decreases significantly when we add region or country fixed effects combined with satellite fixed effects. If we use country fixed effects, the coefficient is only 0.173. Note that we have not considered any further potential determinants of regional incomes until now. In column (4), we add the country-level income, which decreases  $\hat{\beta}$  to 0.117.

As discussed above, a country fixed-effects specification would likely yield the least biased estimates for the relationship between regional nighttime light and income. However, for our forecasting model, this specification is useless because we cannot use country fixed effects in an out-of-sample prediction. Therefore, we considered country-group fixed effects instead of country fixed effects in the results reported in column (5). The coefficient of interest is much larger here with a value of 0.252. If we neglect country-specific factors, we overestimate the effect of regional light on regional income. Therefore, we add the country-level income in column (6), which decreases  $\hat{\beta}$  to 0.106. In column (7), we add further country-level variables: country size and the number of regions. The coefficient  $\hat{\beta}$  is 0.115 there, which is very close to the country fixed-effects estimate discussed above. Obviously, our country level (time-invariant) control variables and the country-group fixed effects yield results, which are not different from a country fixed effects estimation. The results reported in column (7) provide the parameters that we use in our forecasting model.

Columns (8)-(10) split the sample by average income. The estimates for low-income countries (column 8) are similar compared with the full sample (column 7). In middle-income countries (column 9), the relationship between light and income is a bit stronger ( $\hat{\beta} = 0.173$ ) but with a higher uncertainty of the point estimate. In high-income countries (column 10), the coefficient is smaller compared with the full sample ( $\hat{\beta} = 0.082$ ) and not statistically significant. On the one hand, this result may be related to the problem of top-coded luminosity data. One might think of the Netherlands or other densely populated rich countries where we observe a high incidence of top-coded regional luminosity data, and therefore too little variation in the data. On the other hand, the relationship between luminosity and income might differ across countries within this particular income group. This is relevant if the technologies that generate income vary; think, for example, of tax havens compared with non-tax havens or countries with a large financial sector compared with manufacturing-based economies. The country-group fixed effects certainly cannot account for all of these differences. The result of a less stable relationship between light intensity and income in high-income countries is consistent with previous studies; see Chen and Nordhaus (2011, 2015). These measurement problems are obviously less relevant in middle- and low-income countries than in high-income economies. This is important for our analysis because we aim to construct regional data particularly for low- and middle-income countries.

Next, we briefly discuss the regression diagnostics. At this stage of our analysis, we refer to the data provided by Gennaioli et al. (2014). We can ultimately use the data of 1,484 regions in 80 countries, providing a maximum of 5,198 region-year observations. In that specification, which considers regional and satellite fixed effects (column 2), we can explain approximately 53% of the variation in regional incomes within regions and over 32% between them. The

---

model, which uses country fixed effects and country-level income (column 4), explains approximately 66% of the variation within regions and 89% between them. In the specification reported in column (7), which provides the coefficients for our forecasting model, we are able to explain over 66% of the variation in regional income within regions and more than 87% between regions. Therefore, we are confident that our regional income estimates based on luminosity data are reliable.

### 2.3 PREDICTION OF REGIONAL INCOMES

Our regression results show a robust relationship between regional nighttime lights and regional incomes. The model that we ultimately use to predict regional incomes considers country-group fixed effects, satellite fixed effects, and country-level control variables. The estimated parameters  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\gamma}$ ,  $\hat{\delta}$ ,  $\hat{\zeta}$ ,  $\hat{\eta}$ , and  $\hat{\theta}$ , as reported in column (7) in table 1, enter equation [2], yielding the functional relationship that we use to predict regional incomes  $\hat{y}_{i,j,t}$ :

$$\ln(\hat{y}_{i,j,t}) = 0.645 + 0.115 \cdot \ln(l_{i,j,t}) + 0.883 \cdot \ln(y_{j,t}) - 0.137 \cdot \ln(r_j) + 0.054 \cdot \ln(a_j) + \hat{\eta}_k + \hat{\theta}_t. \quad [4]$$

We calculate the luminosity variable  $l_{i,j,t}$  for all sub-national regions in the world using the 1<sup>st</sup>-level administrative boundaries defined by the GADM project (section 2.1). All the remaining country-level data are taken from the World Development Indicators and the CIA World Factbook. The variation in regional income within countries is therefore determined by  $\hat{\beta}$  and  $l_{i,j,t}$ . The other parameters only affect the regional income level and not the dispersion within countries.

We emphasize that we change the data sets between equations [1] and [2 and 4]. The light—income relationship is estimated using the Gennaioli et al. (2014) data (equation [1]). This estimation gives us the parameters for our prediction of regional incomes for all countries (equations [4]). In doing so, we assume that all countries in the world, particularly those where regional income data are unavailable, have a similar relationship between light and income. This assumption needs discussion. The regional data provided by Gennaioli et al. (2014) consider many countries at different levels of economic development. However, there may still be systematic differences between those countries where regional data is available and those where we must rely on our constructed measures, which may cause bias in our forecast model. The largest number of countries with missing regional income data are located in Sub-Saharan Africa. We ultimately consider 47 SSA countries in our analysis, of which only 6 are included in the Gennaioli et al. (2014) data. The average country-level income in the 41 countries without any regional income data is 3,320 US\$ (standard deviation 5,040 US\$), while it is 3,086 US\$ (standard deviation 3,335 US\$) in those 6 countries where we could compare our predicted regional incomes with reported regional incomes.<sup>8</sup> Obviously, there is a selection bias in the Gennaioli et al. (2014) data toward countries with higher income levels, which is likely to be correlated with the capacity of statistical authorities to provide regional data. Our estimations reported in table 1 imply that the low-income countries (GDP p.c. < 4,086 US\$) have similar light—income relationships compared with the full sample. We are therefore confident that our predictions of regional income are also informative in out-of-sample predictions. However,

---

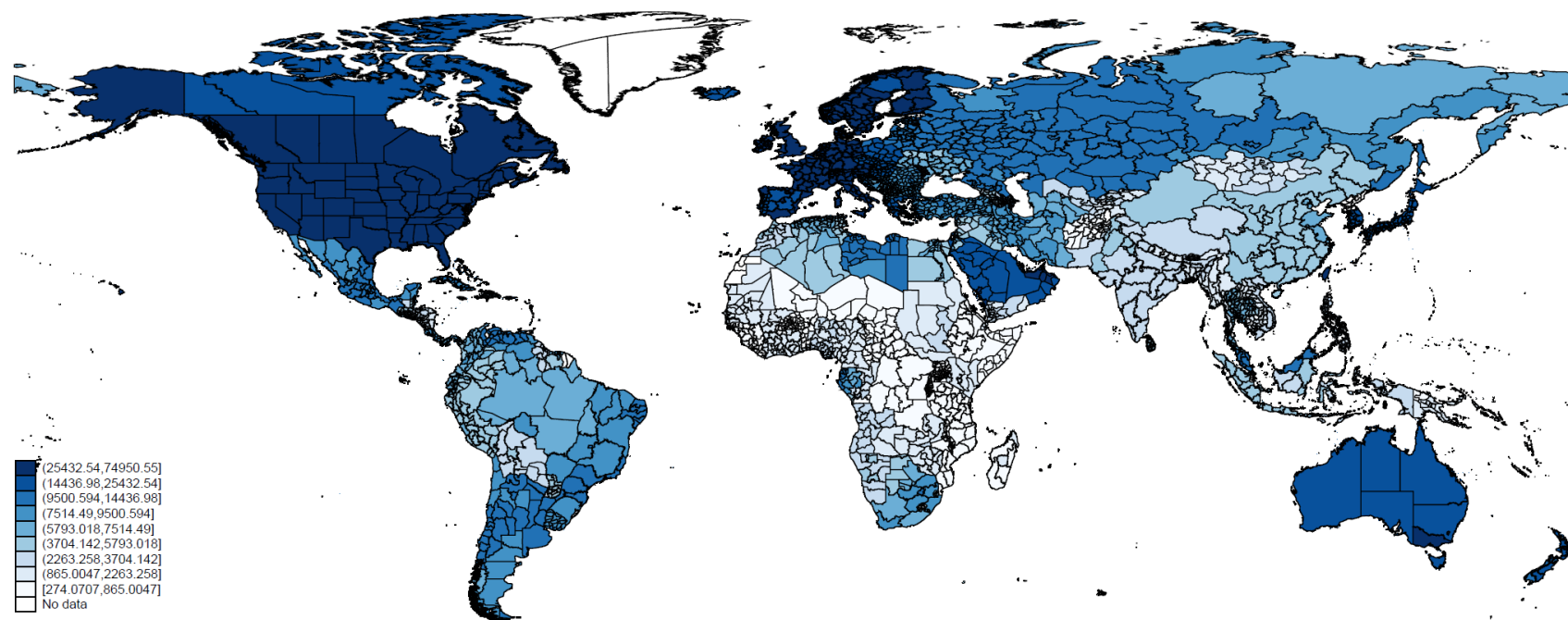
<sup>8</sup> The 6 SSA countries in the Gennaioli et al. (2014) data are Benin, Kenya, Lesotho, Mozambique, Nigeria, and Tanzania.

---

the sample of low-income countries in the Gennaioli data, which is our basis, is still too small and rich compared with the entire population of low-income countries. If the light—income relationships within this group are completely different, then our regional income predictions are biased. However, it is not possible to test this issue directly. Instead, we aggregate the constructed regional incomes to country-level incomes using equation [3] and compare these results with reported country-level incomes.

Our prediction of regional incomes yields a panel data set of 3,163 regions in 180 countries for the period (1992-2012). The (unweighted) average regional income is 9,380 US\$ (standard deviation 10,464). The richest region is Ad Dawhah in Qatar; the poorest region is Grand Cape Mount in Liberia. The data cover 99% of the arable global surface and 99% of the gross world product. Figure 1 illustrates the predicted regional incomes on a world map. The darker the color of a region is, the higher the regional GDP p.c. is. All the countries are rated on the same scale; therefore, we observe only a few variations in regional incomes within countries but large differentials between them. Unsurprisingly, the Northern Territories in Canada and the Northern parts of Russia appear significantly poorer than the rest of the countries. We could also observe significant income differentials across the regions of South America. However, more informative are those maps that focus on the regional incomes within country groups. Figures A.1 and A.2 in the appendix show regional income data for Sub-Saharan Africa and Liberia, respectively. Here, we observe much more variation in the data because the scale is more sensitive to income variations in these particular groups of countries.

**Figure 1:** : Predicted regional income (mean 2001-2010)



---

## 2.4 DISCUSSION OF RESULTS

In this section, we compare the predicted regional incomes with the observed data. Figure 2 illustrates the results. The abscissa reports the regional income predictions as calculated from equation [4]; the ordinate shows reported regional incomes. We also add the information on the country-level income distinguishing between the categories low, middle and high income. Note that we build averages for the period 2001-2010; therefore, each bullet in the figure reflects one region. The figure includes a bisecting line where predicted regional income and reported income is equal ( $\hat{y}_i = y_i$ ).

The illustration suggests a highly positive relationship between the predicted regional incomes and reported incomes. The Pearson correlation coefficient is 0.90, which deviates only slightly from the ideal value of 1. The figure also suggests that we do not systematically over- or underestimate regional incomes in the overall level because the distribution of observations around the bisecting line is symmetric. However, the quality of our estimates depends on the income level because the deviation of observations from the bisecting line is particularly large in middle-income countries. This point is highlighted by the use of different symbols for the corresponding country-level incomes: triangles represent low-income countries, crosses middle-income countries, and dots high-income countries. Our constructed regional incomes are close to the reported incomes in high-income countries; however, there are several outliers in lower middle- and middle-income countries. Obviously, our model underestimates regional income in some cases; these are exceptionally rich regions in the relatively poor countries, which sometimes correspond to the capital region. Here, the scaling of the luminosity data becomes important again because metropolitan areas in developing countries are also top-coded.

Finally, we briefly discuss two country examples. Figure 3 compares predicted state income with reported income for the United States of America (panel a) and Vietnam (panel b). In the case of the USA, our prediction is very close to the reported data. Only Alaska is a clear outlier, where we significantly underestimate the regional income. Again, this underestimation has to do with the luminosity data because the satellite sensors are not sensitive enough to capture the few sources of light. More than 99% of the pixels in Alaska are coded with a zero. In Vietnam, our prediction is much poorer. We systematically overestimate the income of poor regions (dots below the bisecting line) and underestimate the income of rich regions (dots above the bisecting line). Given that the variation between regions within a country is determined purely by nighttime lights in our model, the common coefficient  $\hat{\beta}$  is obviously not suitable to capturing the light—income nexus in the case of Vietnam, at least as long as we compare income levels. If, as we demonstrate later, we concentrate on changes in the regional distribution of incomes over time, this lack of suitability should be less of a concern because we include country fixed effects in the regressions.

A scatter plot showing the relationship between the logarithm of observed regional GDP per capita (y-axis) and the logarithm of predicted regional GDP per capita (x-axis). The plot includes data points for three categories: LIC (Least Income Countries, blue triangles), MIC (Middle Income Countries, red plus signs), and HIC (High Income Countries, green circles). A solid black line represents the identity line (y=x). The data points are clustered along this line, indicating a strong positive correlation between observed and predicted regional GDP per capita. The x-axis ranges from approximately 5.5 to 12.5, and the y-axis ranges from approximately 5.5 to 12.5. The legend is located in the bottom right corner.

(a) United States of America

(b) Vietnam

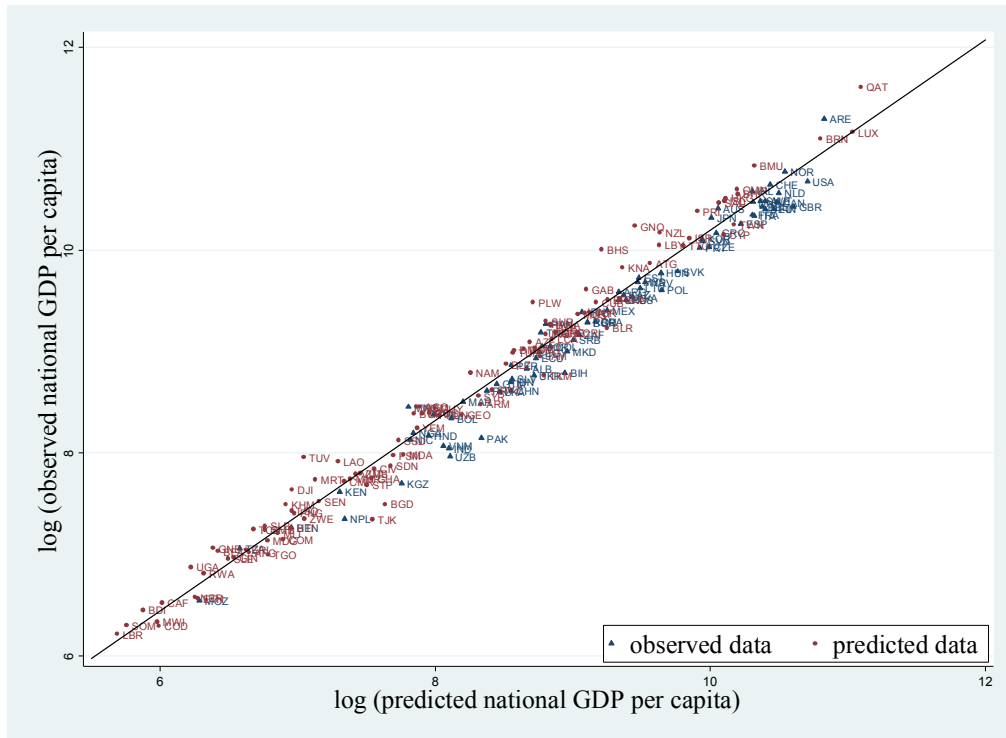
15



a fitted line to the graph, which helps us identify those countries where we over- or underestimate aggregate incomes by trend. If we add the bisecting line instead, we observe that the predicted income systematically underestimates observed incomes by approximately 2%. Please see figure A.5 in the appendix for the respective illustration. This underestimation is caused by the different base years used in the different data sets, changes in the PPP calculations between 2005 and 2011, and adjustments by Gennaioli et al. (2014) to fit the aggregated national GDP to WDI data. These changes are, however, common across countries; therefore, they cannot affect the income distribution within countries.

In general, our predicted national incomes are close to the officially reported data. The correlation between predicted national income and observed national income is 0.97. This value is satisfying, given that we aggregate predicted regional incomes from regional lights to country-level incomes. However, there are some outliers where our income predictions deviate significantly from the observed data. This is particularly true for maritime states, such as Tuvalu, Palau, and the Bahamas. Here, we significantly underestimate incomes, which is likely a low coding issue.<sup>9</sup> We also significantly overestimate aggregate income in some cases, e.g., in Pakistan, Tajikistan, and Uzbekistan. Finally, the graph also illustrates the differences between the observable data and our full sample. With our approach, we can also provide regional income data for those low- and middle-income countries where no reliable data exist.

**Figure 4:** All countries: Predicted versus observed GDP per capita at the national level (mean 2001-2010)



<sup>9</sup> Note that the following sections exclude very small countries from the analysis because it makes little sense to analyze regional inequality in those cases.

---

## 3 REGIONAL INEQUALITY AROUND THE WORLD

### 3.1 REGIONAL INEQUALITY

Regional income data are relevant for many studies in economics and geography. A major field of research is concerned with regional convergence. The literature distinguishes between two types of convergence: beta-convergence and sigma-convergence. Beta-convergence refers to the situation where poor regions grow faster than rich regions; i.e., the poor regions are catching up. Sigma-convergence refers to cases in which the dispersion of income decreases over time. While one might think that both approaches are equal, they are not. Quah (1993, 1996) and Young et al. (2008) show that beta-convergence is a necessary but not sufficient condition for sigma-convergence. If, for example, we observe beta-convergence in a particular period but some former poor regions within a country overtake previously rich ones, then the dispersion of income may increase rather than decrease.<sup>10</sup> Hotelling (1933) and Friedman (1992) already noted that a real test of a tendency to convergence should concentrate on measures of income dispersion. We intend to follow that suggestion in the present section.

The most influential empirical studies on convergence are likely Barro and Sala-i-Martin (1991, 1992). Gennaioli et al. (2014), who provide us with regional income and luminosity data, also analyze beta-convergence in a wider data set. Those studies find evidence of beta-convergence between regions with a typical speed of convergence between 1 and 2% per year. If we take this result seriously, we should expect all income differentials between regions to vanish over time and must simply wait. However, studies that concentrate on sigma-convergence, i.e., the dispersion of income within countries, have contradicting results [see e.g., Quah (1993, 1996) and Young et al. (2008)]. Moreover, the relationship between income (growth) and income dispersion within a country is not necessarily linear. Williamson (1965) adapts the idea of Kuznets (1955) to the case of regional inequality.<sup>11</sup> In the original work, Kuznets (1955) states that interpersonal income inequality first increases in the course of economic development, then peaks, and then decreases. This sort of relationship is often called inverted U-shaped, which may also apply to regional inequality. One reason for this pattern could be, for example, that a single region in a homogeneous country faces a positive technological shock or a discovery of natural resources. This region will grow faster, and regional inequality will rise, until a convergence process begins that reduces inequality. Empirical evidence for the inverted U hypothesis in regional inequality exists, e.g., Williamson (1965) and Lessmann (2014). Moreover, Lessmann (2014) finds some evidence that regional inequality increases again at very high levels of development in a sample of 56 countries. This finding is consistent with the results of Amos (1988), who finds that regional inequality increases across U.S. states after the inverted U pattern has been completed. Unfortunately, the study by Lessmann (2014) has a serious bias toward middle- and high-income countries; therefore, we reproduce his examination in the following section based on our total population of countries.

---

<sup>10</sup> Brezis et al. (1993) call these turnovers “leapfrogging”.

Note that we use the phrases “sigma-convergence” and “changes in regional inequality” synonymously. By studying regional inequality more generally instead of sigma-convergence, we also allow for cross-country comparisons in levels of income dispersion, as is common in economic geography. In the following, we calculate different measures of regional inequality within countries and test for sigma-convergence.

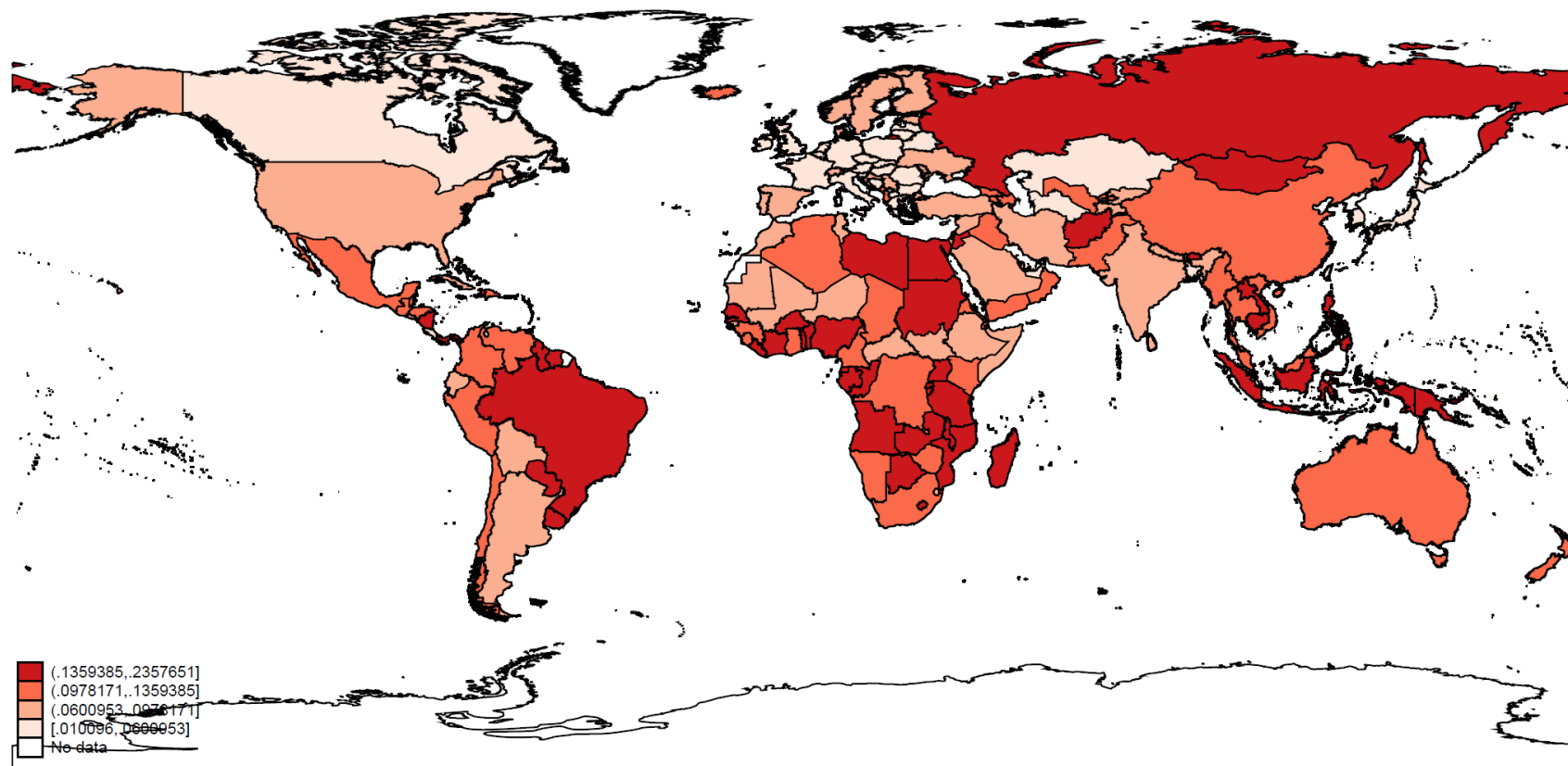
Measuring regional income inequality is more challenging than measuring personal income inequality on one important dimension – the heterogeneity of regions. The number of regions by country varies in our data set between 2 (Sao Tome and Principe) and 89 (Russia). Additionally, the size of regions is difficult to compare with Lake Sevan (Armenia) as the smallest region in our data with 0.6 sq.km and Sakha (Russia) as the largest region with 7,508,595 sq. km. An inequality measure that aims to compare income levels across countries must account for this issue. Otherwise, the different values of an inequality measure may yield a completely misleading country ranking. If, in contrast, the focus is purely on changes in inequality within countries over time, this is a minor issue because the country-level territorial heterogeneity is fixed. Based on the predicted regional incomes  $\hat{y}_{i,j,t}$  we calculate different inequality measures: the coefficient of variation ( $CV_j$ ), the Gini coefficient ( $GINI_j$ ), and the population-weighted coefficient of variation ( $WCV_j$ ). All the inequality measures satisfy the relative income principle (mean independence), population principle, and Pigou-Dalton principle. The weighted coefficient of variation also accounts for the different sizes of regions. The inequality measures are given by (omitting time subscripts for clarity):

$$\begin{aligned}
 CV_j &= \frac{1}{\hat{y}_j} \left[ \frac{1}{n_j} \sum_i^{n_j} (\hat{y}_j - \hat{y}_i)^2 \right]^{1/2}, \\
 GINI_j &= \frac{2 \sum_i^{n_j} i \hat{y}_i}{n_j \sum_i^{n_j} \hat{y}_i} - \frac{n_j+1}{n_j}, \\
 WCV_j &= \frac{1}{\hat{y}_j} \left[ \sum_i^{n_j} \frac{p_i}{P_j} (\hat{y}_j - \hat{y}_i)^2 \right]^{1/2}.
 \end{aligned}
 \tag{5,6,7}$$

The  $CV$  relates the standard variation of regional incomes to the country mean. In contrast, the  $WCV$  weights the squared deviation of regional income with the population share of that region in the respective country ( $p_i/P_j$ ) to give smaller (larger) regions a smaller (larger) weight in the overall inequality measure. Hence, highly unequal population distribution within countries is taken into account. This is important, for example, in Canada, where the Northern Territories have a significantly lower income compared with the country average but are very sparsely populated. Without considering the lower population there, Canada appears to be one of the most unequal economies, while it is in the group of the countries with the lowest regional inequality if we consider the  $WCV$ . One could also interpret the  $CV$  and  $GINI$  as pure measures of geographic inequality between different administrative regions, while the  $WCV$  measures intergroup inequality in a country, where groups of people are formed by their place of residence. The un-weighted inequality measures are commonly used in the literature on sigma-convergence, while the weighted coefficient of variation has been suggested by Williamson (1965) to compare inequality levels across countries.

We first discuss the level of sub-national regional inequality across countries. For this purpose, we draw figure 5, which illustrates the mean of the weighted coefficient of variation  $WCV$  for the period 2001-2010. The map shows the results of the inequality measure, with darker colors representing higher levels of inequality.

**Figure 5:** Regional inequality within countries (WCV, mean (2001-2010))



Industrial countries in North America and the core of Europe have the lowest levels of regional inequality, while countries in Latin America, Africa, and East Asia show significantly higher inequality. However, within the different country groups are important differences. The results for African countries are particularly interesting because there does not exist a systematic analysis for a large set of countries that uses a comparable database. The very poor countries in the Sahel zone and the landlocked countries in the center of the continent have relatively low regional inequalities. The countries with access to the sea, particularly in the South-West of Africa and in the South East, have relatively high inequality. Additionally, the richer countries, e.g., Namibia and South Africa, have relatively low regional inequalities. This observation supports the results of Williamson (1965) and Lessmann (2014) noting an inverted U-shaped relationship between regional inequality and the level of economic development, at least in this particular country group. We stress that the existing studies could not include the poorest countries, which makes it difficult to find evidence for the upward sloping part of the Kuznets Curve in the early stages of development. With our predicted regional income data, we are now able to test this pattern. Please see section 4 for further details.

We only comment on some descriptive statistics because it is important to stress that different inequality measures do not usually provide the same country ranking. However, we encourage the reader to study the detailed results for all countries and inequality measures, which are reported in table A.4 in the appendix. Table 2 shows Pearson correlation coefficients for the three variables based on country-level data for the period 2001-2010. The coefficient of variation and the Gini coefficient – both un-weighted inequality measures – are highly correlated across countries with a correlation coefficient of 0.93. The population-weighted coefficient of variation *WCV* has a correlation of 0.54 with the *CV* and 0.67 with the *GINI*. Obviously, the weights are very influential, at least if the focus is on comparing inequality levels. In the following econometric analyses, we concentrate on the *WCV* and use the alternative inequality measures for robustness tests.

**Table 2:** Correlation between inequality measures in the cross-section; 180 countries; mean of period 2001–2010

	CV	GINI	WCV
CV	1.000		
GINI	0.933	1.000	
WCV	0.543	0.667	1.000

## 3.2 SIGMA-CONVERGENCE

Using the inequality measures, we can investigate the changes in regional inequality; i.e., we can test for sigma-convergence. Table A.4 in the appendix provides all the results at the country level. Here, we only discuss aggregated results at the country-group level to save space. Table 3 reports the results and has the same structure as table A.4. All the values are un-weighted means within the respective country group.

In columns (1) and (2), we report period averages for the coefficient of variation in two periods: 1992-2001 and 2002-2011. In columns (3)-(6), we report the results for the Gini coefficient and the weighted coefficient of variation. Columns (7)-(9) report the percentage change of the respective inequality measure between the two periods. Column (10) reports whether all the inequality measures point in the same direction concerning the trend.

**Table 3:** Aggregate results on regional inequality and sigma-convergence

Country Group	CV		Regional Inequality (level)		WCV		Sigma- Convergence Change in %			Unambi- gous trend
	1992- 2001	2002- 2011	1992- 2001	2002- 2011	1992- 2001	2002- 2011	CV GINI WCV			all measures
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
EAP	0.176	0.172	0.086	0.085	0.119	0.119	-2.2	-1.5	0.0	no
ECA	0.102	0.096	0.050	0.047	0.072	0.066	-5.9	-6.9	-8.1	yes
LAC	0.153	0.143	0.077	0.072	0.114	0.106	-6.5	-6.9	-6.8	yes
MENA	0.153	0.147	0.075	0.071	0.107	0.103	-3.7	-4.3	-3.6	yes
NA	0.171	0.169	0.097	0.095	0.074	0.072	-1.3	-1.3	-2.8	yes
SA	0.149	0.146	0.079	0.079	0.097	0.105	-2.4	0.0	8.3	no
SSA	0.243	0.232	0.100	0.099	0.130	0.127	-4.8	-1.5	-2.3	yes

East Asia and the Pacific (EAP), Europe and Central Asia (ECA), Latin America and the Caribbean (LAC), Middle East and North Africa (MENA), North America (NA), South Asia (SA), and Sub-Saharan Africa (SSA).

First, we observe that regional inequality is highest in Sub-Saharan Africa and lowest in Europe and Central Asia. This result is not sensitive to the respective inequality measure. The changes in the inequality measure in time show whether we observe sigma-convergence or divergence. A negative sign indicates convergence, while a positive sign shows an increase in inequality. In East Asia and the Pacific and in South Asia, the results are ambiguous; i.e., they depend on the underlying measurement concept. In all the other regions of the world, we find sigma-convergence in the un-weighted averages.

The un-weighted country group averages discussed above are very crude measures; therefore, the detailed results reported in table A.4 are much more informative. We comment on some highlights, omitting the smallest countries. Considering the percentage change in the *WCV* (column 9), the most rapid convergence could be observed in Kazakhstan, Georgia, Sri Lanka, Tonga, Honduras, Azerbaijan, Lesotho, Ireland, El Salvador, and Poland. The greatest divergence faces Mozambique, Mongolia, Moldova, Mauritania, Republic of Congo, Cambodia, Laos, Ethiopia, Slovakia, and Mali. Among the high-income countries, Ireland, Poland, and Iceland face the most rapid convergence process, while Hungary, Russia, and South Korea diverge. Note that with the exceptions of Hungary and South Korea, all the OECD member countries face regional sigma-convergence. Pooling all countries together, we find – based on the *WCV* – that 126 of 177 countries (71.2%) face sigma-convergence. If we consider only those countries where the different inequality measures point in the same direction (column 10 = “yes”), then we find that 110 of 138 countries (79.7%) face convergence.

The descriptive results for our inequality measures suggest that the majority of countries face sigma-convergence. However, at least one fifth of countries face divergence. These findings may be related to a non-linear effect between inequality and development. Of course, they may also have individual reasons, e.g., asymmetric macroeconomic shocks that affect the different regions of countries differently. In section 4, we will therefore examine the determinants of regional inequality levels and sigma-convergences using cross-country and panel data.

---

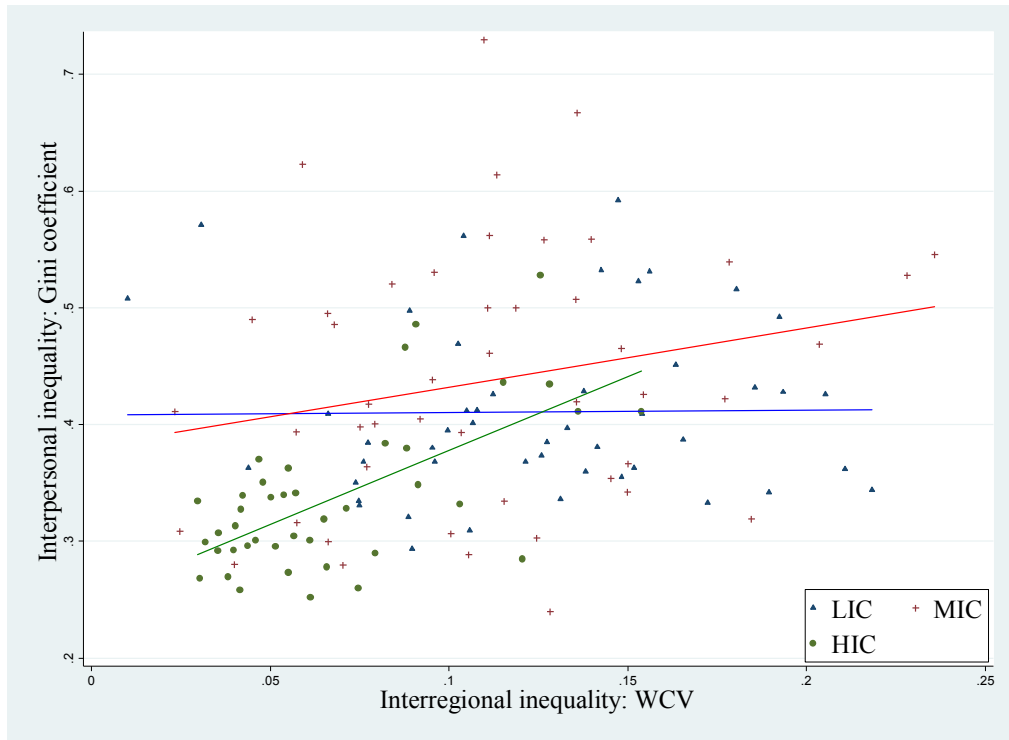
### 3.3 INTERREGIONAL INEQUALITY VERSUS INTERPERSONAL INEQUALITY

In this section, we study the relationship between *interregional* and *interpersonal* inequality. In their book on *Spatial Inequality and Development*, Kanbur and Venables (2005, p. 3) state that “spatial inequality is a dimension of overall inequality, but it has added significance when spatial and regional divisions align with political and ethnic tensions to undermine social and political stability”. Stewart (2000, 2002) distinguishes between vertical inequality and horizontal inequality. Vertical inequality refers to inequality within a group, which could be income inequality among people within one country. This type of inequality is usually measured by the Gini coefficient of the overall income distribution. Horizontal inequality refers to inequality between groups, where the groups can be defined by ethnicity, religion, region, or other factors. Recent papers on violent conflict show that regional inequality – as one particular type of horizontal inequality – has a greater positive effect than vertical inequality [see, e.g., Østby et al. (2009), Deiwiiks et al. (2012), Buhaug et al. (2012) and Lessmann (2015)]. Consequently, the question arises how these different types of inequality interrelate?

Figure 6 shows a scatterplot of the  $WCV$  as a measure of regional inequality (abscissa) and the income Gini coefficient as a measure of personal inequality (ordinate). The income Gini coefficients are taken from the World Bank project “All the Ginis” [see Milanovic (2013)]. This dataset includes combined and standardized Gini data from different sources, with differing quality. Note that we add single missing values from other sources, such as the CIA World Factbook. The figure distinguishes between income groups and includes trendlines with different colors used for the respective groups. Importantly, the relationship between regional inequality and personal inequality is positive across all countries. In the full sample, the Pearson correlation coefficient is 0.45, which is close to previous estimates based on a smaller number of countries [see e.g., Lessmann (2014)]. A linear regression of the type  $GINI_t^{income} = \alpha + \beta \cdot WCV_t^{region} + \varepsilon_t$  yields  $\alpha = 0.330$  and  $\beta = 0.713$ , where both coefficients are statistically significant at the 1% level.

Among the different country groups, we find that the relationship between both types of inequality increases in income. In low-income countries (LIC, blue trend), there is no relationship between personal inequality and regional inequality at all; in middle-income countries (MIC, red trend), the relationship is positive but with large deviations; and in high-income countries (HIC, green trend), the relationship is the largest and with low uncertainty. These results depart from the existing literature because Lessmann (2014) and others do not consider the cross-country data of a larger number of low-income countries. We find that the conclusions concerning the relationship between regional and personal inequality derived from the data of MICs and HICs could not be applied to LICs.

**Figure 6:** Regional inequality versus personal inequality (mean 2001-2010)



## 4 DETERMINANTS OF REGIONAL INEQUALITY AND CONVERGENCE

The previous section has shown that regional inequality matters, at least because it is related to personal inequality. Countries have different levels of regional inequality. The majority of countries converge, while regions in a significant number of countries diverge. In this section, we take one step further by investigating the determinants of regional inequality. We thereby distinguish between two different types of empirical models: (1) cross-sectional regressions, where we build a long period average of our measures of regional inequality; and (2) panel fixed-effects regressions, where we build 5-year period averages. In the cross-sectional regressions, we exploit the between-country variation in the data in the *level* of regional inequality (section 4.1). The panel regressions focus on the within-country variation over time, which allows us to derive conclusions concerning the effect of the determinants on the *changes* of regional inequality, or the sigma-convergence, respectively (section 4.2).

### 4.1 DETERMINANTS OF REGIONAL INEQUALITY: CROSS-SECTIONAL REGRESSIONS

We first study the determinants of the level of regional inequality across countries. For this purpose, we collapse our data to a period average from 2001-2010. We selected this more recent period because the availability of data on potential explanatory variables has improved over time. By employing cross-sectional regressions, we exploit the variation in the data between rather than within countries. Our basic model has the following form:



---


$$I_j = \alpha + \beta \cdot F(\ln(y_{j,t})) + \gamma_k \cdot \sum_{k=1}^m X_{k,j} + \varepsilon_j, \quad [8]$$

where  $I_j$  reflects the level of regional inequality in country  $j$ ,  $F(\cdot)$  is a polynomial function of income  $\ln(y_{j,t})$ ,  $X_{k,j}$  are  $m$  different explanatory variables, and  $\varepsilon_j$  is the error term. Concerning the particular measure of regional inequality, we refer to the population-weighted coefficient of variation of regional income in the body of the paper. The other measures are used for robustness tests and are presented in the appendix.

#### 4.1.1 DETERMINANTS

The literature discusses different determinants of regional inequality. We cluster the different variables into seven different groups:

- Economy
- Geography
- Mobility
- Openness
- Resources
- Institutions
- Ethnicity

The starting point for our analysis is the empirical model used in Lessmann (2014), who focuses on the relationship between inequality and development but also considers other potential determinants of regional inequality. We discuss the related theory and existing evidence topic by topic. Note that a clear separation of issues is not always possible. For example, while natural resources may affect policies, institutions, and trade, location of resources itself is rather a geographic question.

##### *ECONOMY*

Kuznets (1955) proposes an inverted U-shaped relationship between personal inequality and the level of economic development. This concept is applied by Williamson (1965) to the context of regional inequality. The major idea is that the development process in very poor economies usually starts in one part of the country; therefore, these regions become richer, while the others remain poor. The initial shock may be the discovery of natural resources or the implementation (or adaptation) of new technologies and the like. Consequently, regional inequalities rise at low levels of development. At higher levels of economic development, the lagging regions catch up, because, for example, they adopt the new technology or because factors are mobile and induce a convergence process. Barrios and Strobl (2009) have formalized these ideas in one analytical model discovering the famous inverted U, i.e., the Kuznets curve in regional development. Empirical evidence for this relationship is provided by Williamson (1965), Barrios and Strobl (2009), and Lessmann (2014). However, the results of the existing studies are limited to a maximum number of 56 countries with a serious bias toward middle- and high-income economies. We therefore re-examine this issue based on the full populations of countries. The level of economic development is measured by the log of the real GDP p.c. in

---

constant US \$ in power purchasing parities (PPP). We consider a linear specification, a quadric, as well as a cubic relationship between inequality and development. Note that we use only parametric regressions here because semi-parametric models do not outperform our non-linear specifications.

#### *GEOGRAPHY*

It is reasonable to expect that geographic variables affect regional inequality. First, we acknowledge that the territorial level, at which regional incomes are calculated, varies considerably across countries. In particular, the range of our inequality measures is affected by the number of regions. Differences in the size of sub-national regions may also be relevant because we ultimately compare Alaska with Berlin, which must be considered not only by the population-weights. We follow Lessmann (2014) by controlling for the number of subnational units, country surface (in sq. km), and subnational average unit size (all in logs). Additionally, we add the population density within the entire country. Moreover, some countries have large deserts or mountainous areas. In those cases, the population is highly concentrated in particular regions, which should be taken into account. We therefore add the share of arable land of the total surface as another geographic variable.

#### *MOBILITY*

In the new economic geography, transport costs play an important role in agglomeration and income [see Krugman (1991)]. With infinite transportation costs, factor reallocation is impossible; therefore, the level of regional income inequality is determined by initial factor endowments, labor, knowledge, and other factors. Agglomerations occur at an intermediate level of transportation costs because firms seek to be close to markets. This tendency is usually accompanied by regional income differentials. If transportation costs reach zero, then there are no further benefits from agglomeration; therefore, regional income differentials vanish. We measure factor mobility using two different variables: the road density defined by the length of the street network in relation to the overall surface of a country and the share of paved roads of the total road network. Note that we use a linear and a quadratic term of the road density variable to account for the non-linearity in the effect discussed.

#### *OPENNESS*

The openness of an economy may also affect regional inequality within countries [see Brühlhart (2011) for an overview]. If international trade costs decrease, then those regions of a country with relatively better market access gain more from international trade than those regions with limited access to international markets. Hirte and Lessmann (2014) and other authors show that this effect should depend on conditioning factors, such as internal transport costs. Empirical evidence on this issue clearly supports a positive effect of international trade on regional inequality. Rodríguez-Pose and Gil (2006), Barrios and Strobl (2009), Rodríguez-Pose (2012), Ezcurra and Rodríguez-Pose (2013), and Hirte and Lessmann (2014) investigate this issue using various methods, measures, and data sets. Some of the studies consider non-linearities and/or interaction variables. We use two variables in our model: the trade/GDP ratio, where trade is measured by the sum of exports and imports, and a dummy variable for landlocked countries because this is a geographically determined obstacle to international trade.

---

Another potential determinant of inequality, which is also related to the openness of the economy, is foreign direct investments (FDI). Let us assume a homogeneous country with low levels of regional inequality. An influx of FDI into one region will increase the capital stock in that region, and the marginal product of labor increases as output and consumption (assuming complementarity between capital and labor). Consequently, regional inequalities rise as a response to the FDI influx. The output promoting effect will occur with greenfield investment because the physical capital stock increases directly and also in the case of mergers and acquisitions, which usually involve a transfer of knowledge (intangible capital). The transmission channels have been intensively discussed in the literature on FDI and growth [see, e.g., Borensztein et al. (1997)]. In our context, we are interested in the effect of FDI on regional inequality – not growth. The case of China is a focus of the theoretical and empirical literature on this particular issue. China opened its economy to investors in the early 1990s and has since faced a rapid increase in regional inequality [see, e.g., Fleisher et al. (2010)]. However, after the implementation of several government programs that aimed to channel FDI to the hinterlands, regional inequalities decreased again. We consider this issue in our cross-sectional regressions by including the net FDI inflows as a share of the GDP as a potential determinant of regional inequality.

#### *RESOURCES*

Natural resources are a major determinant of regional incomes. Some resources, such as forests or arable land, are often uniformly distributed within countries, while others, such as minerals and fuels, are highly concentrated within countries. These so-called point resources will affect the industry location and therefore the income dispersion within countries. Ullman (1958) discusses this issue based on the example of the Manufacturing Belt in the U.S.; however, this issue is similar in the Ruhr area in Germany and many other countries. As noted by Williamson (1965), the discovery of natural resources is often a critical juncture in regional development because the resource-rich regions begin to grow faster relative to the rest of the country. Consequently, regional inequalities are expected to rise, at least if the initial income distribution was balanced before. Henderson et al. (2012) show how the accidental discovery of gemstones in southern Madagascar near the town Ilakaka at the end of the 1990s could be traced with the nighttime light data. The light emissions have increased since then significantly due to a rapid increase in the regional population and income. In this example, this development has likely taken a toll on nearby cities, which began to shrink. The net effect of resources on regional inequality therefore depends on the initial income distribution. In a homogeneous country, point resources are likely to increase inequalities; in heterogeneous countries, this effect depends on the initial income distribution. If natural resources help poor regions catch up, then inequalities may also fall, at least if the poor regions do not overtake the richer ones. Note that the net effect of resources on inequality is also ambiguous, if one takes into account a potential resource curse at the regional level [see Sachs and Warner (1999)]. We abstract from these issues by only considering a linear and unconditional effect of resources on inequality; variable interactions are likely to be important but go beyond the scope of our analysis. The variables that we consider are mineral rents as a share of the GDP and oil rents / GDP as reported by the World Development indicators.

#### *TRANSFERS*

Interregional redistribution policies are another potential determinant of regional inequality. In several countries, such as Germany and Canada, formal equalization schemes exist that aim to reduce regional inequalities. One might think

---

also of the European Union, which supports relatively poor regions with different cohesion policies. However, the academic literature is inconclusive on the effectiveness of these programs. Kessler et al. (2011) note that transfers may actually increase regional inequalities because the payments inhibit convergence, promoting migration. Unfortunately, it is impossible to obtain comparable data on within-country equalization payments for a large set of countries, which would undoubtedly be interesting. Instead, we consider social expenditures for unemployment, pensions, and other government programs. These payments may also affect regional inequalities because poor regions within countries usually have more individuals who receive net transfers, while the inhabitants of rich regions on average pay net contributions. Consequently, social expenditures can be seen as an indirect instrument to redistribute between regions. We therefore use the social expenditures as a share of total government revenues as another determinant of regional inequality. The expected result is unclear for two reasons: (1) the theory is ambiguous because transfers may harm convergence as discussed above, and (2) our measure of regional income does not distinguish between output and consumption. Both economic activities are positively linked to nighttime light emissions; however, while transfers may perfectly equalize consumption per capita across regions, there may still exist large differentials in the production of goods and services. A second variable that we consider is a dummy for member countries in the European Union. This variable captures the international redistribution policies that also affect the income distribution within countries. Becker et al. (2010) show, for example, that EU structural funds promote beta-convergence across European regions.

#### *INSTITUTIONS*

Institutions are also likely to affect regional inequalities within countries. Ades and Glaser' (1995) model growth of metropolitan areas as a function of the level of democracy and political stability. In their framework, stable dictatorships tend to centralize resources because they set lower tax rates in the center than in the hinterlands. Hodler and Raschky (2014) provide empirical evidence that autocrats tend to channel resources to their regions of origin, which also affects regional inequalities within countries. We therefore consider the autocracy index as reported by the Polity2 data set as a potential determinant of inequality. The sign of the coefficient is again unclear because numerous conditioning factors, such as political stability and initial income distribution, may be important.

As a more general indicator of institutional quality, we consider the index of bureaucratic quality as reported by the PRS Group's International Country Risk Guide. There is extensive literature on the relationship between institutions and development. Acemoglu et al. (2001, 2002) show in a number of influential studies that current institutions and development are determined by colonial institutions, which in turn depend on historical factors, such as population density and urbanization during the time of colonialization. Today's regional inequality may still be affected by these factors. Extractive institutions were often implemented in densely populated areas to obtain control over the local resources. This practice may have promoted growth in these regions while leaving peripheral regions underdeveloped. Considering the work by Dell (2011), it can even be argued that colonialists may have extracted resources from peripheral regions to promote their rent extraction in the central regions. She provides strong evidence for long-lasting effects of colonial institutions on today's differentials in incomes in Peru and Bolivia. A system of labor obligations used by the colonialists (based on previous Maya rules) affected only parts of the regions under study. The historical income differentials caused by these colonial institutions still have an effect on today's differentials in road density, schooling, health, and consumption. Michalopoulos and Papaioannou (2014) challenge this view at least for the case

---

of African countries. They show that different national institutions have no significant effect on the income of ethnic groups (measured by luminosity data) that are located in different countries after decolonization. Apart from these historical factors, one might also consider that countries with high institutional quality allow for more efficient allocation of resources across regions, which should ultimately decrease regional inequality.

A significant number of studies analyze the relationship between federal institutions and regional inequality. Following McKinnon (1997) and Qian and Weingast (1997), transfers of authority to sub-national levels of government should increase the scope of poor regions for policies that attract mobile capital. These authors refer to the potential for decentralization to preserve markets. In a similar manner, Baldwin and Krugman (2004) argue that a harmonized one-size-fits-all policy in taxation or regulations may harm the development of peripheral regions in the presence of agglomeration rents. However, potential benefits of decentralization may be offset by a ruinous tax competition game as noted by Prud'homme (1995). The empirical literature is rather inconclusive. While early cross-country studies find decentralization to decrease regional inequality [see, e.g., Shankar and Shah (2003)], more recent work finds that this effect depends on the development level [see Rodríguez-Pose and Ezcurra (2010), Lessmann (2012)]. In our analysis, we consider a federal dummy variable, following the approach of Shankar and Shah (2003). We admit that this variable is a rather crude measure of decentralization; however, the data are available for a large set of countries. If we used government finance data to calculate measures of fiscal decentralization, we would shift the focus of the analysis to upper-middle- and high-income countries, which is contrary to our major aim.

#### *ETHNICITY*

A final determinant of regional inequality that we consider in our analysis is the degree of ethnic fractionalization. Easterly and Levine (1997) show that ethnic heterogeneity is an important determinant of political instability, conflict, and poor growth performance. Importantly, ethnic fragmentation usually has a regional dimension because different ethnic groups in a country often live in separate territories. If a large ethnic group dominates a minority, then we should expect this domination to yield income differentials. Data on ethnic fractionalization are provided by Alesina et al. (2003). Note that we provide results only for the degree of ethnic fractionalization. The other indicators on linguistic and religious fractionalization have no significant effect in our regressions.

### 4.1.2 RESULTS

Before we discuss the regression results, we stress that all effects should only be interpreted as correlations rather than causal effects. We are aware of the potential endogeneity bias in the estimates. Unfortunately, all potential sources of endogeneity are relevant: measurement errors, omitted variables, and reverse causation. Measurement errors are inherent in our inequality measure and the different (sometime crude) explanatory variables. Omitted variables are often a problem in cross-sectional regressions with country-level data. In contrast to other studies in the field, we have data on almost all countries in the world, which allows us to consider many variables simultaneously without losing too many degrees of freedom. We are therefore confident that this source of bias is not very significant. Furthermore, we note that in the next section, where we apply panel fixed-effects regressions, unobservable country heterogeneities are covered by the dummy variables. Finally yet importantly, there are many cases of potential reverse causality. One

might think of FDI inflows as a determinant of inequality that may be directly affected by inequalities itself, at least if inequality leads to political instability. This problem can only be solved – if at all – with instrumental variable regressions. We leave these important questions for further research.

The results of OLS regressions using the *WCV* as inequality measure are reported in table 4. Tables A.5 and A.6 in the

**Table 4:** Determinants of regional inequality; OLS regressions, means (2001-2011)

Dependent variable: weighted coefficient of variation (WCV)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(GDP p.c.)	0.598** (0.271)	0.524 (0.353)	0.444 (0.285)	0.650** (0.314)	1.041* (0.589)	0.623 (0.387)	0.552** (0.275)
log(GDP p.c.) <sup>2</sup>	-0.069** (0.031)	-0.062 (0.041)	-0.053 (0.032)	-0.075** (0.035)	-0.119* (0.066)	-0.070 (0.045)	-0.063** (0.031)
log(GDP p.c.) <sup>3</sup>	0.003** (0.001)	0.002 (0.002)	0.002 (0.001)	0.003** (0.001)	0.004* (0.002)	0.003 (0.002)	0.002* (0.001)
log(Area)	0.006** (0.003)	-0.005 (0.003)	0.004 (0.003)	0.005 (0.003)	0.011*** (0.003)	-0.004 (0.004)	0.002 (0.003)
log(# Regions)	-0.004 (0.011)	0.011 (0.011)	0.002 (0.011)	-0.002 (0.013)	-0.020 (0.013)	0.029** (0.014)	0.008 (0.012)
log(Area)/log(# Regions)	-0.012** (0.005)	-0.006 (0.006)	-0.011* (0.005)	-0.008 (0.007)	-0.016*** (0.005)	0.010 (0.008)	-0.006 (0.006)
Population density	-0.103*** (0.036)	-0.060*** (0.014)	-0.188*** (0.051)	-0.075** (0.032)	-0.074*** (0.027)	-0.388*** (0.123)	-0.098*** (0.031)
Arable land (%)	-0.103*** (0.023)	-0.028 (0.029)	-0.104*** (0.023)	-0.098*** (0.026)	-0.095*** (0.026)	-0.112*** (0.026)	-0.104*** (0.023)
Road density		-0.045*** (0.011)					
(Road density) <sup>2</sup>		0.006*** (0.002)					
Paved roads (%)		-0.053*** (0.015)					
Trade/GDP			0.026** (0.011)				
FDI/GDP			-0.102 (0.072)				
Landlocked			-0.013 (0.008)				
Mineral rents / GDP				0.022 (0.092)			
Oil rents / GDP				0.048* (0.027)			
EU 27 member					0.004 (0.008)		
Social epense / revenue					-0.078*** (0.027)		
Bureaucratic quality						-0.006 (0.006)	
Polity2						-0.000 (0.007)	
Federal dummy						0.009 (0.009)	
Ethnic fractionalization							0.023* (0.014)
Constant	-1.528* (0.788)	-1.238 (1.003)	-1.038 (0.827)	-1.679* (0.937)	-2.823 (1.732)	-1.710 (1.103)	-1.446* (0.798)
Observations	178	125	171	132	93	126	171
R-squared	0.388	0.534	0.426	0.461	0.619	0.500	0.391

Robust standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

---

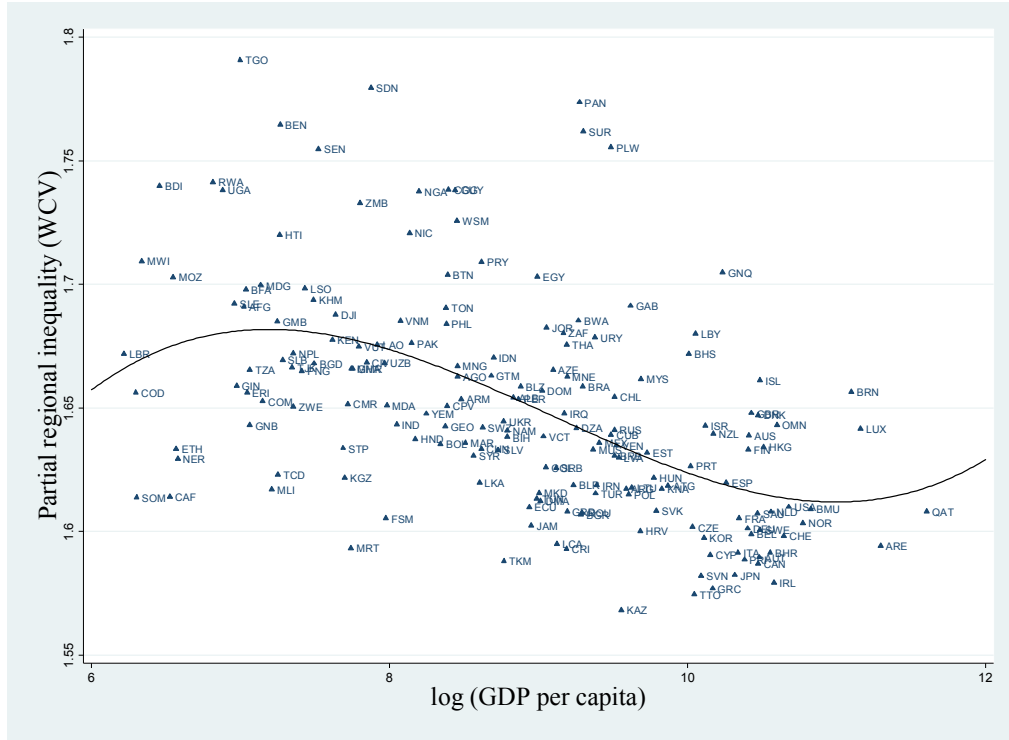
appendix report regression results using the Gini coefficient and the unweighted *CV*, respectively. We run different specifications of equation [8]. The baseline model, which considers the level of economic development and geography, is reported in column (1). We then sequentially add the other variables. Column (2) reports the results for mobility. Column (3) uses the openness measures. Column (4) considers natural resources. Column (5) adds transfers. Column (6) considers institutional quality, and, finally, column (7) focuses on ethnicity.

The baseline specification reported in column (1) shows significant effects for the development level and several geographic variables. The coefficients of the income variable suggest an N-shaped relationship between economic development and regional inequality. Regional inequalities first increase in the development level, peak, and then decrease again but increase at high-income levels. Figure 7 shows a partial scatterplot, including the regression function. For this purpose, we partial out the effects of the control variables from the measure of regional inequality and then compare them with the development level. The threshold values of income can be calculated by setting the first derivatives of the regression function to zero  $\partial I_j / \partial \ln(y_{j,t}) = \beta_1 + 2\beta_2 \ln(y_{j,t}) + 3\beta_3 \ln(y_{j,t})^2 = 0$ . Solving this equation for  $y$  yields 1,320 US\$ as local maximum and 59,659 US\$ as local minimum. Very poor countries face increasing inequalities in our data set; countries such as Afghanistan, Liberia, and Togo belong to that group. Beyond this threshold, regional inequalities fall until a threshold of 59,659 US\$, where inequalities start to increase again. The clear majority of countries are in the downward sloping part of the N-shaped function. This finding supports or results in the sigma-convergence discussed in section 3. The very rich countries with higher inequalities are the United Arab Emirates, Brunei, Luxembourg and Qatar. Note that we refine these results in the panel regressions employed in the next section. Concerning the fragmentation variables, we find that in countries with a low number of subnational regions, regional inequality increases with the countries' sizes. Population density is negatively related to regional inequality. This is the case in the core of Western Europe, which has relatively low regional inequalities. The share of arable land of the total surface is also negatively related to regional inequality, which accords with our expectations.

In column (2), we add the mobility variables. We consider a non-linear relationship between road density and regional inequality. The coefficient of the linear term is negative, while the coefficient of the quadratic term is positive; both are highly significant. These results suggest that regional inequalities are lower in countries with higher road density but with decreasing returns concerning the lowering effect of infrastructure on inequality. The share of paved roads is unambiguously negative; i.e., the greater the share of paved roads, the lower the regional inequalities are. Note that we also considered variables directly measuring mobility, such as the number of vehicles per capita, which do not have any relationship to regional inequality.

In column (3), we consider the openness variables: trade/GDP, FDI/GDP and a landlocked dummy. In line with existing studies, we find a positive and significant effect of trade on regional inequality. Foreign direct investments do not affect inequality in this regression. However, if we interact FDI with income, we find – in line with Lessmann (2013) – different effects across high- and low-income economies. While FDI has no effect on regional inequality in high-income countries, FDI increases regional inequalities in low-income countries. This may be related to the lower factor mobility in low-income countries, which inhibits rapid adjustment to FDI inflows. The landlocked dummy variable is not statistically significant.

**Figure 7:** Adjusted spatial inequality versus log observed GDP per capita at the national level (mean 2001-2010). The points in the graph are partial residuals for spatial inequality; the inequality measure has been adjusted for the effects of the other explanatory variables in the model as specified in column (1) Table 4. The plot represents the N-shaped relationship between economic development and regional inequality, as predicted by the baseline specification in column (1) Table 4.



In column (4), we add the two variables for point resources: mineral rents to GDP and oil rents to GDP. While both variables have a positive coefficient, only oil rents have a significant effect. The larger the share of oil net revenues to GDP, the larger are regional inequalities. Note we also considered other resource variables, such as forest rents and oil exports as a share of total exports; however, only the rent variable ultimately had a significant effect.

In column (5), we focus on transfers – indirect transfers between regions via social expenditures and direct transfers between members of the European Union. While the EU dummy has no significant effect on regional inequality, the social expenditure variable has a highly significant negative effect. Note that the EU dummy is a rather crude measure of international transfers to sub-national regions. Using a direct transfer variable might yield other results.

In column (6), we consider the different variables for institutions: the bureaucratic quality index, the polity2 index, and the federal dummy. None of the variables has a significant effect. The results for the federal dummy are noteworthy because our result differs from those of previous studies, which find a significant negative effect of decentralization on regional inequality. This difference may be a result of the large differences in country samples and also to some extent related to measurement issues. Apart from Shankar and Shah (2003), the existing studies focus directly on measures of fiscal decentralization, which are only available for middle- and high-income countries where regional



---

income is also observable. We have not considered this sort of data because it would shift the focus of our study away from developing countries.

Finally, in column (7), we use the degree of ethnic fractionalization as the explanatory variable. The coefficient has a positive sign and is statistically significant at conventional confidence levels. This result is not surprising. The inequality measure that we use here could be interpreted as a measure of intergroup inequality where the groups are formed by the place of residence of the people. Obviously, different ethnic groups live – at least to some extent – geographically separated from each other. One might think of the Dutch-speaking Flemings and French-speaking Walloons in Belgium or Krahn and Gio in Liberia (see Figures A.3 and A.2 for an illustration). In these cases, we observe larger regional inequalities, implying that there are also significant income differentials between different ethnic groups.

We will briefly comment on the regression diagnostics. Our baseline model explains almost 40% of the variation in regional inequality across countries. Adding further explanatory variables increases the explanatory power up to 60%. The number of observations depends on the particular topic. Only in the case of the transfer variables is the sample under 100 countries.

To summarize, we find that the level of economic development is an important determinant of regional inequality. The relationship is N-shaped; i.e., very poor economies have low regional inequalities, which then increase in income, peak, and then decrease again. At high levels of economic development, regional inequality tends to increase again. Among the other potential determinants of differences in the levels of regional inequality across countries, we find significant effects for roads, trade, oil rents, social expenditures, and ethnicity. The next section evaluates whether these variable also affect the sigma-convergence within countries, i.e., the changes in regional inequality over time.

## 4.2 DETERMINANTS OF SIGMA-CONVERGENCE: PANEL REGRESSIONS

While the previous section focused on the differences in levels of regional inequality across countries, we now consider the changes within countries over time; i.e., we analyze the determinants of sigma-convergence. Doing so is important for several reasons. First, our inequality measures may be biased in levels because we have heterogeneous territorial levels in the different countries. Our control variables and the population-weights in the inequality measure may capture these heterogeneities only imperfectly. Moreover, there may be an omitted variable bias in the estimates if there are unobserved factors that affect both inequality and the explanatory variables simultaneously. Panel regressions with country fixed effects reduce this potential source of bias. By applying this methodology, we focus on the within-country changes in the income dispersion, which we expect to be widely independent of the territorial level.<sup>12</sup>

The basic empirical model in this section has the form

---

<sup>12</sup> See Lessmann (2014) for a comparison of measures of regional inequality using the NUTS1 and the NUTS2 classifications. At least in the EU member countries, the different territorial levels do not affect significantly similar panel regressions.

---


$$I_{j,t} = \alpha_j + \beta \cdot F(\ln(y_{j,t})) + \gamma_k \cdot \sum_{k=1}^m X_{k,j,t} + \theta_t + \varepsilon_{j,t}, \quad [9]$$

where  $\alpha_j$  is country fixed effects, and  $\theta_t$  is time fixed effects. We concentrate on time-variant determinants of regional inequality. Most geographic variables cancel out because they are perfectly collinear to the country fixed effects. Importantly, we use 5-year period averages to eliminate business cycle effects.<sup>13</sup> Over the business cycle, income and inequality are likely to be correlated. If the different regions of a country are heterogeneous with respect to the structure of the economy, than they will react differently on booms or recessions, causing a short-run relationship between income and inequality, which we do not aim to investigate. Germany is an illustrative example because the Western Bundesländer are relatively more open to the world market compared with the Eastern Bundesländer. Therefore, income is more volatile in the West than in the East. Booms and busts therefore cause short-run fluctuations in regional inequality in Germany. See Figure A.4 in the appendix for an illustration of this issue.

The regression results using the *WCV* as inequality measure are reported in table 5.<sup>14</sup> Importantly, we also find evidence for an N-shaped relationship between inequality and development in the panel. In the results reported in column (1), the threshold values, where the relationship between inequality and development changes its sign, are 1,840 US\$ and 82,406 US\$, implying that only very rich countries face increasing inequalities. Concerning the other variables, we find that the population density has no significant effect if we also consider the mobility variables (column 2). This effect is caused by some multi-collinearity. If we eliminate the paved roads variable, then we obtain significant effects of the road density variables. The trade-to-GDP ratio has a highly significant effect (column 3). Natural resources and social expenditures have no significant effects (column 4 and 5). Notably, the Polity2 index is negatively associated with regional inequalities in the panel. We interpret this result to signify that more democratic institutions assist lagging regions (and minorities), which results in a higher speed of convergence.

---

<sup>13</sup> The periods are 1990-1994, 1995-1999, 2000-2004, 2005-2009, and 2010-2013.

<sup>14</sup> Please see tables A.7 and A.8 in the appendix for the alternative inequality measures.

**Table 5:** Determinants of sigma convergence, panel fixed effects regressions, 5-year averages

	Dependent variable: weighted coefficient of variation (WCV)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(GDP p.c.)	0.574*** (0.078)	0.657*** (0.225)	0.504*** (0.100)	0.434*** (0.139)	0.374 (0.308)	0.570*** (0.072)
log(GDP p.c.) <sup>2</sup>	-0.063*** (0.010)	-0.073*** (0.026)	-0.056*** (0.012)	-0.049*** (0.016)	-0.043 (0.033)	-0.064*** (0.009)
log(GDP p.c.) <sup>3</sup>	0.002*** (0.000)	0.003*** (0.001)	0.002*** (0.000)	0.002*** (0.001)	0.002 (0.001)	0.002*** (0.000)
Population density	-0.308*** (0.100)	-0.167 (0.186)	-0.240*** (0.079)	-0.120 (0.093)	-1.004 (1.051)	-0.201 (0.127)
Arable land (%)	-0.073*** (0.024)	-0.136** (0.061)	-0.057** (0.024)	-0.069* (0.037)	-0.048 (0.033)	-0.063* (0.034)
Road density		0.014 (0.011)				
(Road density) <sup>2</sup>		-0.003 (0.002)				
Paved roads (%)		0.012 (0.009)				
Trade/GDP			0.008*** (0.003)			
FDI/GDP			0.008 (0.012)			
Mineral rents / GDP				0.053 (0.036)		
Oil rents / GDP				0.019 (0.017)		
Social expense / revenue					0.006 (0.024)	
Bureaucratic quality						0.000 (0.001)
Polity2						-0.004* (0.002)
Constant	-1.536*** (0.211)	-1.807*** (0.652)	-1.345*** (0.284)	-1.123*** (0.406)	-0.912 (0.936)	-1.507*** (0.180)
Observations	865	312	814	643	331	485
# Country	178	155	172	134	111	127
R-squared-w	0.203	0.393	0.249	0.374	0.396	0.363
R-squared-b	0.248	0.071	0.242	0.393	0.110	0.432
Country FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## 5 SUMMARY AND CONCLUSIONS

In this paper, we analyze regional inequality within countries based on satellite nighttime light data. We use the luminosity data to predict regional incomes in those countries where this variable is not observed. For this purpose, we first refer to the data set collected by Gennaioli et al. (2014), which contains information on both regional income and average nighttime lights. Taking the empirically estimated relationship between these variables, we predict regional incomes from this particular country sample. This forecast yields a panel data set of regional incomes across the earth covering 180 countries for the period 1992-2012.

Next, we calculate different measures of regional inequality within countries. We find significant differences concerning the levels of regional inequalities between countries. In particular, middle-income economies have above-

---

average levels of inequality. Additionally, very rich countries have a relatively high degree of income dispersion. These findings indicate an N-shaped relationship between inequality and development, which is confirmed by different regressions. Moreover, we find that the majority of countries (71-80%) face decreasing regional inequalities over time, i.e., sigma-convergence. However, our study also reveals that regional convergence is not inevitable because regional income inequalities increase in several countries, such as South Korea or Hungary. We also investigate the relationship between regional inequality and personal inequality, finding an overall correlation of approximately 0.45. However, this relationship only holds for middle- and high-income economies. In low-income countries, regional inequalities are independent of personal inequalities.

Based on the different inequality measures, we further investigate the determinants of regional inequality and sigma-convergence. We thereby distinguish between economy, geography, mobility, openness, natural resources, transfers, institutions, and culture. The cross-sectional regressions reveal that the development level and different geographic variables are particularly decisive. Concerning the other variables, we find that mobility and welfare expenditures reduce the level of regional inequality, while trade openness, oil rents, and ethnic fractionalization increase inequality. We also employ panel regressions using 5-year period averages. Using country fixed effects, we focus on the determinants of the within-country variation in inequality over time, which are ultimately the determinants of sigma-convergence. Here, we find support for the main results of the cross-sectional regressions, particularly those concerning the relationship between regional inequality and development. Additionally, mobility variables and trade openness are significant determinants of convergence.

What have we learnt from this study? (1) Regional inequalities are to some extent persistent. Lessmann (2014) finds an inverted U-shaped relationship between inequality and development. This finding would imply that inequalities decrease over time beyond a certain income threshold. We, however, find robust evidence that regional inequalities increase again at very high levels of development. This finding may be related to structural changes in the economy to a modern service-based economy, where new leaders overtake former rich manufacturing regions. Examples of this scenario may be the Silicon Valley in the U.S. and Bavaria in Germany. If one is concerned with regional inequality, then this finding justifies regional policies that aim at a more equal income distribution within countries. (2) Globalization causes an increase in inequality, which could be a challenge for policy makers, particularly in countries that are ethnically diverse. (3) Effective instruments that help reduce inequality may be related to decreasing transport costs because a better infrastructure is negatively related to inequality. (4) Indirect transfers via the social security system also reduce regional inequalities. Our policy conclusion is therefore to balance the different effects. If increased trade openness gives rise to a conflict between efficiency (economic growth) and redistribution (regional inequality), then transfers, infrastructure investments, and democratic institutions might help offset negative side effects.

Finally, we stress here that all the empirical results are merely interesting correlations. Apart from considering country fixed effects in the panel regressions, we take no further steps to ensure that we estimate causal effects between regional inequality and its potential determinants. In several cases, for example concerning trade openness and social expenditures, an endogeneity bias is likely to occur. However, in this initial study, we abstract from these important technical issues in favor of a broad first glimpse of the nexus between regional inequality and its determinants. Consequently, these relationships should be studied with more appropriate empirical methods in the future. Moreover,

---

we suggest studying various variable interactions, which are also likely to be important. One might ask, for example, whether good institutions are a relevant conditioning factor in the relationship between point resources and regional inequality. By considering interaction terms, one could obtain a much clearer picture of the interplay among all the different variables.

---

## REFERENCES

- Acemoglu, D., S. Johnson and James A. Robinson (2002): Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution, *Quarterly Journal of Economics*, 117(4), 1231-1294.
- Acemoglu, D., S. Johnson and James A. Robinson (2001): The Colonial Origins of Comparative Development: An Empirical Investigation, *American Economic Review*, 91(5), 1369-1401.
- Ades, A. and E. Glaeser (1995): Trade and Circuses: Explaining Urban Giants, *Quarterly Journal of Economics*, 110(1) (1995), 228-258.
- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): Fractionalization, *Journal of Economic Growth*, 8(2), 155-194.
- Amos, O.J. (1988): Unbalanced Regional Growth and Regional Income Inequality in the Latter Stages of Development, *Regional Science and Urban Economics*, 18(4), 549-566.
- Baldwin, R.E. and P. Krugman (2004): Agglomeration, Integration and Tax Harmonisation, *European Economic Review*, 48(1), 1-23.
- Barrios, S. and E. Strobl (2009): The Dynamics of Regional Inequalities, *Regional Science and Urban Economics*, 39(5), 575-591.
- Barro, R.J. and X. Sala-i-Martin (1991): Convergence across States and Regions, *Brookings Papers on Economic Activity*, 22(1), 107-182.
- Barro, R.J. and X. Sala-i-Martin (1992): Convergence, *Journal of Political Economy*, 100(2), 223-251.
- Becker, S.O., P.H. Egger and M. von Ehrlich (2010): Going NUTS: The Effect of EU Structural Funds on Regional Performance, *Journal of Public Economics*, 94(9-10), 578-590.
- Besley, T. and M. Reynal-Querol (2014): The Legacy of Historical Conflict: Evidence from Africa, *American Political Science Review*, 108(2), 319-336.
- Brezis, E.S, P.R. Kruman and D. Tsiddon (1993): Leapfrogging in International Competition: A Theory of Cycles in National Technological Leadership, *American Economic Review*, 83(5), 1211-1219.
- Brühlhart, M. (2011): The Spatial Effects of Trade Openness: A Survey, *Review of World Economics*, 147(1), 59-83.
- Borensztein, E., J. De Gregorio, and J.-W. Lee (1998): How Does Foreign Direct Investment Affect Economic Growth? *Journal of International Economics*, 45(1), 115-135.
- Buhaug, H., K. Gleditsch, H. Holtermann, G. Østby, and A.F. Tollefsen (2012): It's the Local Economy, Stupid! Geographic Wealth Dispersion and Conflict Outbreak Location, *Journal of Conflict Resolution*, 55(5), 814-840.
- Chen, Xi and W.D. Nordhaus (2011): Using Luminosity Data as a Proxy for Economic Statistics, *Proceedings of the National Academy of Science of the USA*, 108(21), 8589-8594.
- Chen, X. and W.D. Nordhaus (2015): A Sharper Image? Estimates of the Precision of Nighttime Lights as a Proxy for Economic Statistics, *Journal of Economic Geography*, 15(1), 217-246.
- Deiwiiks, C. L. Cederman, and K. Gleditsch (2012): Inequality and Conflict in Federations, *Journal of Peace Research*, 49(2), 289-304.
- Easterly, W. and R. Levine (1997): Africa's Growth Tragedy: Policies and Ethnic Divisions, *Quarterly Journal of Economics*, 112(4), 1203-1250.

- 
- Ebener, S., C. Murray, A. Tandon, and C. C. Elvidge (2005): From Wealth to Health: Modelling the Distribution of Income per Capita at the Sub-National Level Using Night-Time Light Imagery. *International Journal of Health Geographics*, 4(1), 5.
- Elvidge, C.D, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis and C. W. Davis (1997): Relation Between Satellite Observed Visible-Near Infrared Emissions, Population, Economic Activity and Electric Power Consumption, *International Journal of Remote Sensing*, 18(6), 1373-1379.
- Ezcurra, R. and A. Rodríguez-Pose (2013): Does Economic Globalization Affect Regional Inequality? A Cross-Country Analysis, *World Development*, 52(C), 92-103.
- Fleisher, B., H. Li and M.Q. Zhao (2010): Human Capital, Economic Growth, and Regional Inequality in China, *Journal of Development Economics*, 92(2), 215-231.
- Friedman, M. (1992): Do Old Fallacies Ever Die? *Journal of Economics Literature*, 30(4), 2129–2132.
- Gennaioli, N., R. La Porta, F. Lopez de Silanes, and A. Shleifer (2014): Growth in Regions, *Journal of Economic Growth*, 19(3), 259-309.
- Gennaioli, N., R. La Porta, F. Lopez de Silanes, and A. Shleifer (2013): Human Capital and Regional Development, *Quarterly Journal of Economics*, 128(1), 105-164.
- Ghosh, T., R. L. Powell, C. D. Elvidge, K. E. Baugh, P. C. Sutton, and S. Anderson (2010): Shedding Light on the Global Distribution of Economic Activity. *The Open Geography Journal*, 3(1), 148–161.
- Henderson, J.V., A. Storeygard, and D. N Weil (2012): Measuring Economic Growth from Outer Space, *American Economic Review*, 102(2), 994-1028.
- Hirte, G. and C. Lessmann (2014): *Trade, Integration, and Interregional Inequality*, CESifo WP 4799.
- Hodler, R. and P.A. Raschky (2014): Regional Favoritism, *Quarterly Journal of Economics*, 129(2), 995-1033.
- Hotelling, H. (1933): Review of the Triumph of Mediocrity in Business, by Horace Secrist, *Journal of the American Statistical Association*, 28(184), 463-465.
- Johnson, S., W. Larson, C. Papageorgiou and A. Subramanian (2013): Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates, *Journal of Monetary Economics*, 60(2), 255-274.
- Kanbur, R. and A. Venables (2005a): Spatial Inequality and Development, in: Kanbur, R. and A. Venables (Eds.), *Spatial Inequality and Development*, 3-11. New York: Oxford University Press.
- Kanbur, R., and A., Venables (2005b): Rising Spatial Disparities and Development, in *United Nations University Policy Brief No. 3*. Helsinki.
- Kessler, A.S., N.A. and C. Lessmann (2011): Interregional Redistribution and Mobility in Federations: A Positive Approach, *Review of Economic Studies*, 78 (4), 1345-1378.
- Krugman, P (1991): Increasing Returns and Economic Geography, *Journal of Political Economy*, 99(3), 483-499.
- Krugman, P. and R.L. Elizondo (1996): Trade Policy and the Third World Metropolis, *Journal of Development Economics*, 49(1), 137-150.
- Kuznets, S. (1955): Economic Growth and Income Inequality, *American Economic Review*, 45(1), 1-28.
- Lessmann, C. (2009): Fiscal Decentralization and Regional Disparity: Evidence from Cross-Section and Panel Data, *Environment and Planning*, 41(A), 2455–2473.
- Lessmann C. (2012): Regional Inequality and Decentralization: An Empirical Analysis, *Environment and Planning*, 44(A), 1363–1388.

- 
- Lessmann, C. (2013): Foreign Direct Investment and Regional Inequality: A Panel Data Analysis, *China Economic Review*, 24(C), 129-149.
- Lessmann (2014): Spatial Inequality and Development—Is there an Inverted U Relationship? “*Journal of Development Economics*, 106(C), 35-51.
- Lessmann, C. (2015): Regional Inequality and Internal Conflict, *German Economic Review*, DOI: 10.1111/geer.12073.
- McKinnon, R. (1997): Market-Preserving Fiscal Federalism in the American Monetary Union, in: M. Blejer and T. Ter-Minassian (eds.): *Macroeconomic Dimensions of Public Finance: Essays in Honour of Vito Tanzi*, 73-93, London: Routledge.
- Michalopoulos, S. and E. Papaioannou (2014): National Institutions and Subnational Development in Africa, *Quarterly Journal of Economics*, 129(1), 151-213.
- Milanovic, B. (2014): *All the Gini Data Set*, World Bank Group, <http://data.worldbank.org/data-catalog/all-the-ginis>.
- Mveyange, A. F. (2015): Night Lights and Regional Income Inequality in Africa, University of Southern Denmark, mimeo.
- Østby, G., R. Nordås and J.K. Rød (2009): Regional Inequalities and Civil Conflict in Sub-Saharan Africa, *International Studies Quarterly*, 53(2), 301-324.
- Prud'homme, R. (1995): The Dangers of Decentralization, *World Bank Research Observer*, 10(2), 201-220.
- Qian, Y. and B. Weingast (1997): Federalism as a Commitment to Preserving Market Incentives, *Journal of Economic Perspectives*, 11(4), 83-92.
- Quah, D. (1993): Empirics for Economic Growth and Convergence, *European Economic Review*, 40(6), 1353-1375.
- Quah, D. (1996): Galton's Fallacy and the Test of the Convergence Hypothesis, *Scandinavian Journal of Economics*, 95(4), 427-443.
- Rodríguez-Pose, A. (2012): Trade and Regional Inequality, *Economic Geography*, 88(2), 109-136.
- Rodríguez-Pose, A. and N., Gill. (2006): How Does Trade Affect Regional Disparities? “*World Development*, 34(7), 1201-1222.
- Rodríguez-Pose, A. and R. Ezcurra (2010): Does Decentralization Matter for Regional Disparities? A Cross-Country Analysis, *Journal of Economic Geography*, 10(5), 619-644.
- Sachs, J. D. and A. M. Warner (1999): The Big Rush, Natural Resource Booms and Growth, *Journal of Development Economics*, 59(1), 43-76.
- Shankar, R. and A. Shah (2003): Bridging the Economic Divide within Countries - A Scorecard on the Performance of Regional Policies in Reducing Regional Income Disparities, *World Development*, 32(8), 1421-1441.
- Stewart, F. (2000): Crisis Prevention: Tackling Horizontal Inequality, *Oxford Development Studies*, 28(3), 245-262.
- Stewart, F. (2002): *Horizontal Inequality: A Neglected Dimension of Development*, QEH Working Paper Series 81, University of Oxford.
- Sutton, P. C., C. D. Elvidge, and T. Ghosh (2007): Estimation of Gross Domestic Product at Sub-National Scales Using Nighttime Satellite Imagery, *International Journal of Ecological Economics & Statistics*, 8(S07), 5-21.
- Ullman, E.L. (1958): Regional Development and the Geography of Concentration, *Papers in Regional Science*, 4(1), 179-198.



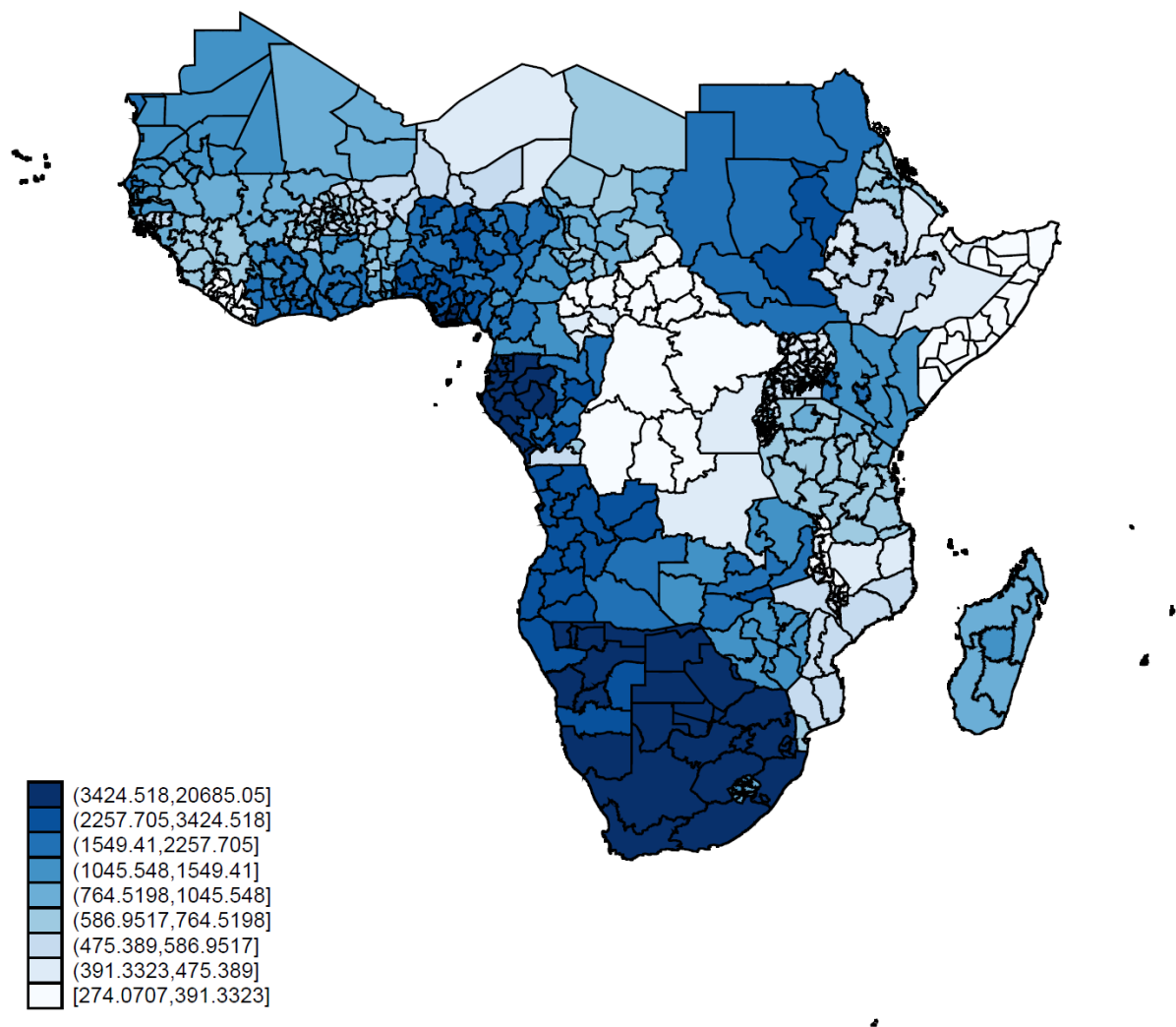
---

Williamson, J.G. (1965): Regional Inequality and the Process of National Development: A Description of Patterns, *Economic Development and Cultural Change*, 13(4), 3-45.

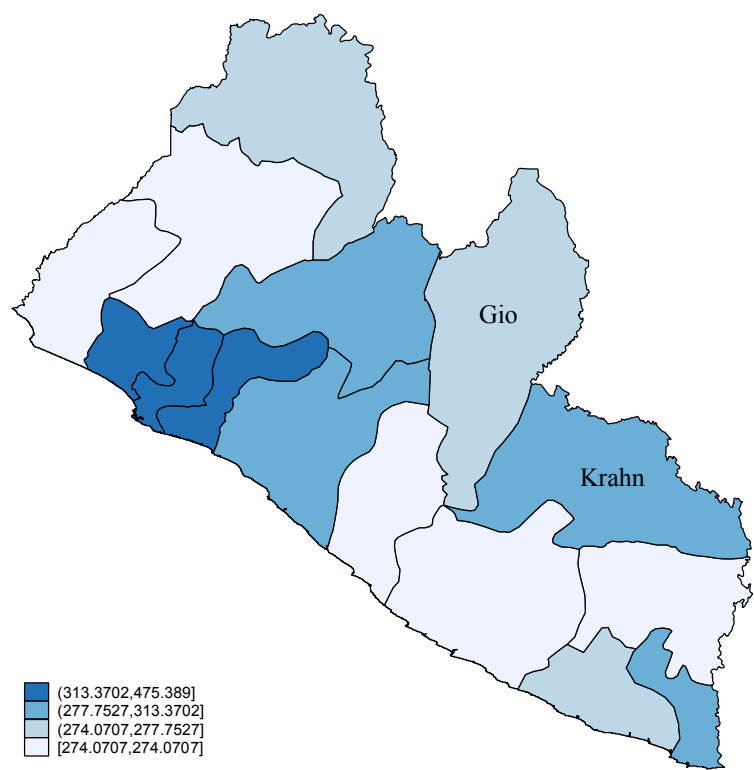
Young, A.T., M.J Higgins and D. Levy (2008): Sigma Convergence versus Beta Convergence, Evidence from U.S. Country Level Data, *Journal of Money, Credit & Banking*, 40(5), 1083-1093.

APPENDIX

Figure A.1: Predicted regional GDP p.c. in Sub-Saharan Africa



**Figure A.2:** Predicted regional GDP p.c. in Liberia



**Figure A.3:** Predicted regional GDP p.c. in Belgium

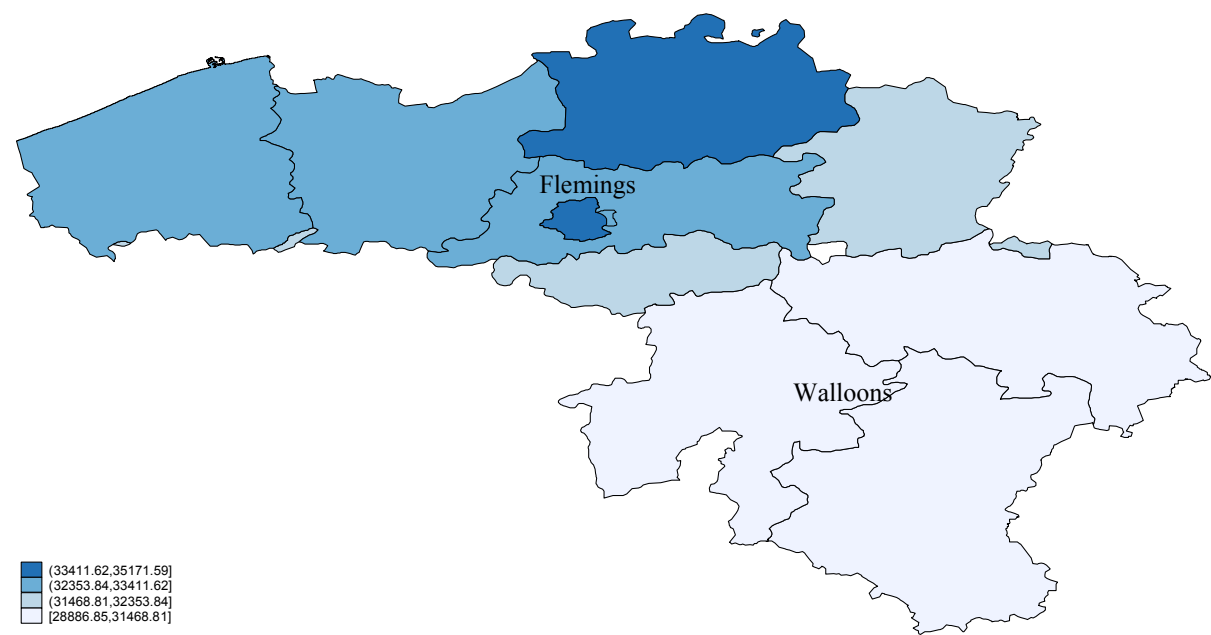


Figure A.4: Regional inequality (WCV) in Germany 2000-2012

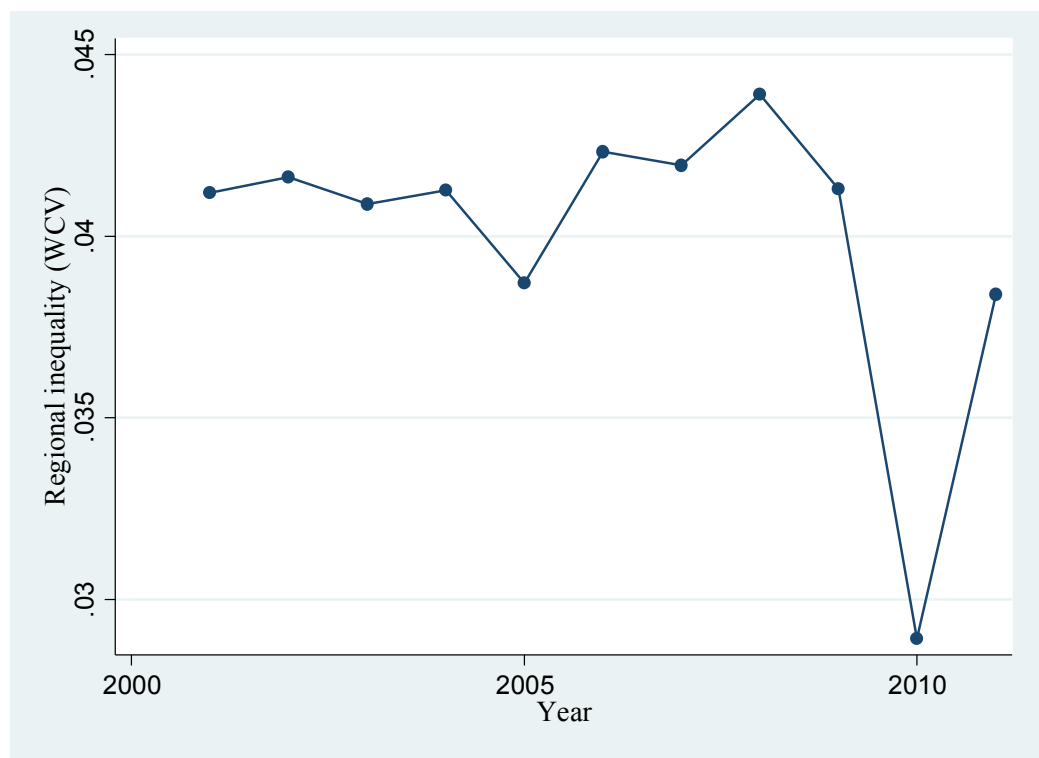
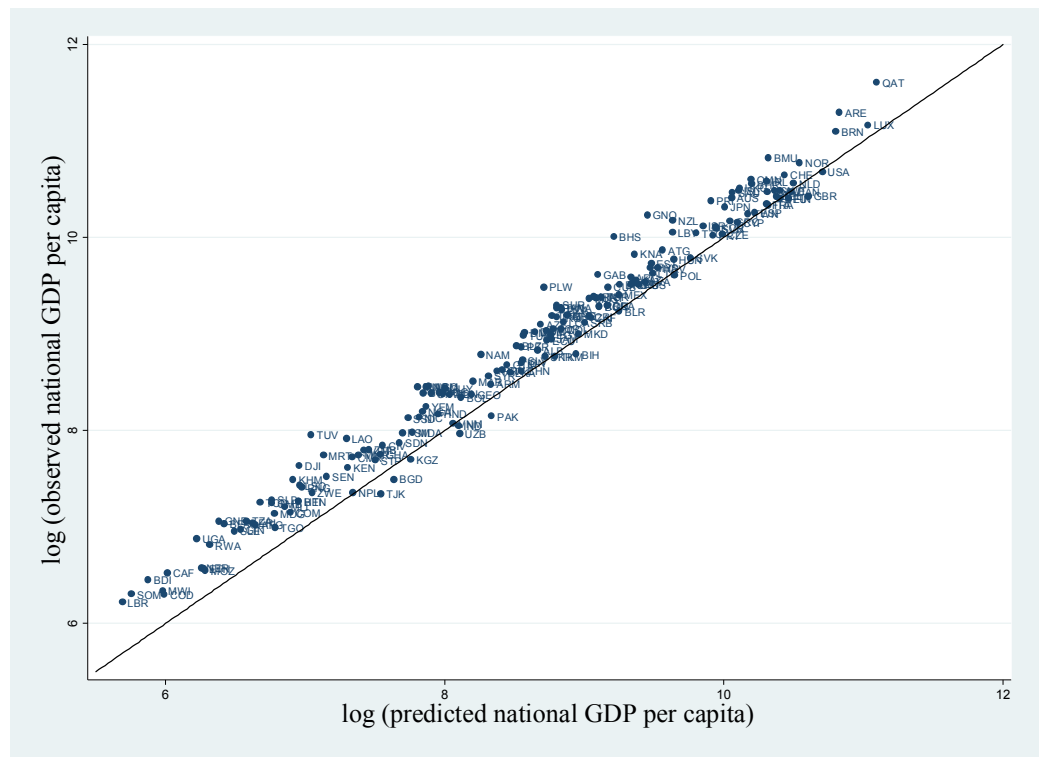


Figure A.5: Predicted vs. observed GDP p.c. at the national level (mean 1997—2007) including bisecting line



**Table A.1:** Descriptive statistics Nighttime Lights Dataset

Variable	Obs	Mean	Std. Dev.	Min	Max
Pix Region	70,581	63,250	278,774	0.740	8.731e+06
Pop Region	70,581	1.674e+06	7.237e+06	1	2.163e+08
# Region	70,581	29.69	23.76	2	89
Light ppix Region	70,581	6.630	11.97	0.01000	63

**Table A.2:** Descriptive statistics Gennaioli, et al. (2014) Dataset

Variable	<i>Obs</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
GDP pc Country	5198	14412.16	12680.81	365.3238	62640.93
GDP pc Region	5198	13290.63	13140.64	188.9685	112293.7

**Table A.3**

Data sources &amp; definitions.

Variable	Sources & definitions
Pix Region	Definition: Area of a region in pixel where 1 pixel is approximately 0.86 square kilometer (at the equator). Source: GADM database of Global Administrative Areas.
Pop Region	Definition: Regional total population. Source: Gridded Population of the World (GPW) v.3 by CIESIN.
# Region	Definition: Number of subnational regions of a country. Source: GADM database of Global Administrative Areas and Gennaioli et al. (2014).
Light ppix Region	Definition: Average nighttime light per pixel, plus 0.01 in those regions, where the average nighttime light per pixel would be zero otherwise. Source: National Oceanic and Atmospheric Administration (NOAA) and Gennaioli et al. (2014).
GDP p.c.	Definition: GDP per capita, PPP (constant 2005 US\$) GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Source: World Bank collection of development indicators and Gennaioli et al. (2014).
Area	Definition: Land area (sq. km). Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases, the definition of inland water bodies includes major rivers and lakes. Source: World Bank collection of development indicators.
Population density	Definition: Population density (100,000 people per sq. km of land area). Population density is midyear population divided by land area in square kilometers. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship--except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases the definition of inland water bodies includes major rivers and lakes. Source: World Bank collection of development indicators.
Arable land	Definition: Arable land (share of land area). Arable land includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded. Source: World Bank collection of development indicators.
Road density	Definition: Road density (km of road per 1 sq. km of land area). Road density is the share of the length of the country's total road network to the country's land area. The road network includes all roads in the country: motorways, highways, main or national roads, secondary or regional roads, and other urban and rural roads. Source: World Bank collection of development indicators.
Paved roads	Definition: Roads, paved (share of total roads). Paved roads are those surfaced with crushed stone (macadam) and hydrocarbon binder or bituminized agents, with concrete, or with cobblestones, as a percentage of all the country's roads, measured in length. Source: World Bank collection of development indicators.
Trade/GDP	Definition: Trade (share of GDP). Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.

---

	Source: World Bank collection of development indicators
FDI/GDP	<p>Definition: Foreign direct investment, net inflows (share of GDP). Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10% or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments. This series shows net inflows (new investment inflows less disinvestment) in the reporting economy from foreign investors.</p> <p>Source: World Bank collection of development indicators.</p>
Mineral rents / GDP	<p>Definition: Mineral rents (share of GDP). Mineral rents are the difference between the value of production for a stock of minerals at world prices and their total costs of production. Minerals included in the calculation are tin, gold, lead, zinc, iron, copper, nickel, silver, bauxite, and phosphate.</p> <p>Source: World Bank collection of development indicators.</p>
Oil rents / GDP	<p>Definition: Oil rents (share of GDP). Oil rents are the difference between the value of crude oil production at world prices and total costs of production.</p> <p>Source: World Bank collection of development indicators.</p>
Social expense / revenue	<p>Definition: Social contributions (share of revenue). Social contributions include social security contributions by employees, employers, and self-employed individuals, and other contributions whose source cannot be determined. They also include actual or imputed contributions to social insurance schemes operated by governments.</p> <p>Source: World Bank collection of development indicators.</p>
Bureaucratic quality	<p>Definition: Bureaucratic quality measured by the subjective index provided by the International Country Risk Guide (ICRG), where high values imply high quality of bureaucratic services.</p> <p>Source: The International Country Risk Guide.</p>
Polity2	<p>Definition: Revised Combined Polity Score (Polity2), rescaled so that it ranges from -1 (strong autocracy) to +1 (strong democracy).</p> <p>Source: The Center for Systemic Peace (CSP) Polity IV Annual Time-Series.</p>
Ethnic fractionalization	<p>Definition: Average value of five different indices of ethnonolinguistic fractionalization. It's value ranges from 0 (low fractionalization) to 1 (high fractionalization).</p> <p>Source: Alesina et al. (2003).</p>

**Table A.4:** Levels and Trends in Regional Inequality within countries

Group	Country	Regional Inequality (level)						Sigma-Convergence			Unambiguous
		CV		GINI		WCV		Change in %			trend
		1992-2001	2002-2011	1992-2001	2002-2011	1992-2001	2002-2011	CV	GINI	WCV	all measures
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
EAP	Australia	0.255	0.257	0.140	0.141	0.103	0.103	0.7	0.8	0.3	yes
EAP	Cambodia	0.254	0.251	0.094	0.108	0.125	0.149	-1.0	14.7	18.7	no
EAP	China	0.173	0.162	0.098	0.091	0.122	0.110	-6.5	-7.2	-9.9	yes
EAP	Hong Kong	0.042	0.041	0.014	0.014	0.033	0.032	-3.5	-4.7	-2.8	yes
EAP	Indonesia	0.242	0.228	0.120	0.113	0.147	0.145	-5.6	-5.3	-1.9	yes
EAP	Japan	0.067	0.066	0.037	0.037	0.055	0.055	-1.2	-1.4	0.3	no
EAP	Laos	0.194	0.204	0.076	0.085	0.128	0.151	5.0	11.6	17.9	yes
EAP	Malaysia	0.155	0.157	0.080	0.080	0.125	0.128	1.5	-0.1	2.4	no
EAP	Micronesia	0.026	0.047	0.013	0.025	0.023	0.049	80.6	84.5	116.1	yes
EAP	Mongolia	0.288	0.283	0.106	0.111	0.139	0.151	-1.8	4.7	8.4	no
EAP	Myanmar	0.166	0.164	0.089	0.086	0.123	0.123	-1.2	-4.1	0.4	no
EAP	New Zealand	0.171	0.164	0.090	0.086	0.121	0.113	-3.8	-3.7	-6.8	yes
EAP	Palau	0.232	0.223	0.130	0.116	0.229	0.235	-3.9	-10.3	2.4	no
EAP	Papua New Guinea	0.119	0.126	0.067	0.070	0.135	0.143	5.3	5.3	6.3	yes
EAP	Philippines	0.221	0.203	0.114	0.102	0.174	0.154	-8.0	-10.5	-11.5	yes
EAP	Samoa	0.154	0.169	0.086	0.096	0.183	0.193	10.2	12.2	5.0	yes
EAP	Solomon Islands	0.149	0.144	0.043	0.039	0.132	0.136	-3.5	-7.8	2.7	no
EAP	South Korea	0.078	0.073	0.041	0.038	0.048	0.050	-6.2	-6.1	2.8	no
EAP	Taiwan	0.311	0.309	0.160	0.158	0.019	0.017	-0.7	-1.3	-10.7	yes
EAP	Thailand	0.153	0.138	0.081	0.073	0.145	0.133	-9.9	-9.5	-8.3	yes
EAP	Tonga	0.243	0.224	0.123	0.112	0.159	0.121	-7.9	-9.1	-23.9	yes
EAP	Vanuatu	0.212	0.208	0.105	0.107	0.121	0.130	-2.0	2.5	6.9	no
EAP	Vietnam	0.146	0.121	0.080	0.067	0.149	0.120	-17.3	-15.6	-19.3	yes
ECA	Albania	0.112	0.108	0.058	0.055	0.105	0.100	-3.5	-5.5	-4.7	yes
ECA	Armenia	0.221	0.199	0.116	0.103	0.123	0.103	-10.3	-10.8	-16.3	yes
ECA	Austria	0.096	0.086	0.038	0.035	0.039	0.035	-10.0	-9.6	-11.4	yes
ECA	Azerbaijan	0.187	0.156	0.099	0.080	0.157	0.124	-16.4	-19.1	-21.4	yes
ECA	Belarus	0.037	0.034	0.018	0.016	0.040	0.038	-7.4	-11.7	-4.9	yes
ECA	Belgium	0.052	0.052	0.028	0.028	0.041	0.040	0.3	0.7	-3.1	no



---

ECA	Bosnia and Herzeg.	0.057	0.089	0.025	0.036	0.030	0.024	57.4	44.4	-18.8	no
ECA	Brunei	0.202	0.191	0.097	0.090	0.097	0.093	-5.5	-6.7	-4.1	yes
ECA	Bulgaria	0.061	0.059	0.031	0.028	0.059	0.057	-3.7	-8.3	-3.8	yes
ECA	Croatia	0.083	0.070	0.042	0.036	0.075	0.059	-15.4	-13.7	-20.3	yes
ECA	Cyprus	0.041	0.032	0.022	0.017	0.037	0.031	-21.5	-21.0	-17.2	yes
ECA	Czech Republic	0.059	0.057	0.026	0.024	0.034	0.031	-3.5	-7.5	-9.7	yes
ECA	Denmark	0.066	0.065	0.034	0.033	0.061	0.061	-2.3	-4.0	-0.8	yes
ECA	Estonia	0.111	0.116	0.061	0.064	0.091	0.089	4.6	4.1	-2.6	no
ECA	Finland	0.111	0.102	0.055	0.050	0.070	0.064	-8.7	-9.9	-7.7	yes
ECA	France	0.059	0.055	0.031	0.029	0.055	0.051	-6.2	-5.7	-7.8	yes
ECA	Georgia	0.230	0.198	0.098	0.084	0.130	0.100	-13.8	-15.1	-23.1	yes
ECA	Germany	0.084	0.079	0.044	0.042	0.044	0.040	-5.9	-5.3	-8.4	yes
ECA	Greece	0.148	0.141	0.061	0.056	0.031	0.029	-4.7	-7.2	-6.6	yes
ECA	Hungary	0.076	0.077	0.031	0.032	0.045	0.047	0.9	3.2	4.5	yes
ECA	Iceland	0.250	0.201	0.105	0.086	0.148	0.118	-19.5	-18.3	-19.9	yes
ECA	Ireland	0.082	0.067	0.039	0.031	0.054	0.042	-18.4	-19.9	-21.8	yes
ECA	Italy	0.047	0.046	0.026	0.025	0.044	0.042	-2.6	-3.6	-4.0	yes
ECA	Kazakhstan	0.064	0.038	0.034	0.021	0.051	0.035	-39.5	-37.7	-32.0	yes
ECA	Kyrgyzstan	0.200	0.192	0.105	0.096	0.088	0.073	-3.8	-8.1	-16.5	yes
ECA	Latvia	0.074	0.074	0.036	0.035	0.055	0.054	-0.9	-4.0	-2.0	yes
ECA	Lithuania	0.053	0.052	0.029	0.029	0.052	0.052	-1.1	0.5	0.2	no
ECA	Luxembourg	0.057	0.056	0.029	0.028	0.059	0.056	-2.3	-2.3	-4.4	yes
ECA	Macedonia	0.059	0.060	0.030	0.031	0.056	0.058	1.3	2.9	3.5	yes
ECA	Moldova	0.089	0.109	0.043	0.052	0.063	0.074	22.1	20.4	16.2	yes
ECA	Montenegro	0.135	0.144	0.074	0.078	0.115	0.124	6.6	6.7	7.9	yes
ECA	Netherlands	0.068	0.067	0.035	0.035	0.044	0.044	-1.7	-1.1	-1.7	yes
ECA	Norway	0.142	0.113	0.074	0.058	0.096	0.078	-20.0	-21.1	-18.6	yes
ECA	Poland	0.058	0.047	0.030	0.025	0.058	0.046	-17.6	-18.5	-19.7	yes
ECA	Portugal	0.097	0.084	0.054	0.047	0.099	0.088	-14.2	-13.7	-11.3	yes
ECA	Romania	0.066	0.057	0.035	0.029	0.062	0.052	-14.2	-17.1	-15.9	yes
ECA	Russia	0.180	0.179	0.097	0.094	0.132	0.136	-0.8	-2.9	2.6	no
ECA	Serbia	0.069	0.068	0.038	0.038	0.067	0.066	-1.8	-1.8	-2.2	yes
ECA	Slovakia	0.037	0.048	0.019	0.025	0.031	0.039	32.4	30.5	25.8	yes
ECA	Slovenia	0.047	0.043	0.026	0.024	0.044	0.040	-7.3	-6.8	-8.0	yes
ECA	Spain	0.098	0.090	0.050	0.046	0.074	0.071	-8.5	-7.1	-4.5	yes

---

ECA	Sweden	0.084	0.084	0.041	0.041	0.076	0.074	-0.5	-1.1	-2.7	yes
ECA	Switzerland	0.085	0.075	0.048	0.042	0.070	0.065	-11.2	-11.3	-7.8	yes
ECA	Tajikistan	0.212	0.201	0.097	0.091	0.094	0.088	-4.9	-6.2	-7.0	yes
ECA	Turkey	0.087	0.075	0.049	0.041	0.088	0.076	-13.6	-15.6	-13.2	yes
ECA	Turkmenistan	0.032	0.026	0.016	0.014	0.029	0.024	-18.8	-15.5	-17.3	yes
ECA	Ukraine	0.087	0.098	0.045	0.048	0.068	0.070	11.9	7.6	2.7	yes
ECA	United Kingdom	0.076	0.073	0.032	0.031	0.049	0.047	-4.2	-2.4	-4.5	yes
ECA	Uruguay	0.187	0.182	0.082	0.080	0.151	0.147	-3.1	-2.7	-2.7	yes
ECA	Uzbekistan	0.164	0.166	0.086	0.086	0.134	0.131	0.9	-0.5	-2.3	no
LAC	Antigua and Barbuda	0.232	0.233	0.117	0.116	0.072	0.068	0.2	-0.2	-6.0	no
LAC	Argentina	0.175	0.163	0.077	0.072	0.098	0.090	-7.2	-6.6	-8.2	yes
LAC	Bahamas	0.287	0.280	0.131	0.119	0.158	0.147	-2.4	-9.4	-6.5	yes
LAC	Barbados	0.066	0.061	0.038	0.034	0.065	0.059	-8.6	-8.2	-9.8	yes
LAC	Belize	0.142	0.128	0.056	0.046	0.115	0.108	-9.7	-18.9	-6.0	yes
LAC	Bermuda	0.041	0.046	0.023	0.026	0.036	0.041	13.6	13.8	15.3	yes
LAC	Bolivia	0.137	0.122	0.072	0.066	0.106	0.095	-11.0	-8.5	-10.5	yes
LAC	Brazil	0.188	0.177	0.107	0.100	0.144	0.139	-6.2	-6.4	-3.6	yes
LAC	Chile	0.175	0.180	0.087	0.091	0.129	0.125	2.8	5.4	-3.4	no
LAC	Colombia	0.231	0.224	0.129	0.124	0.117	0.111	-3.1	-3.8	-5.0	yes
LAC	Costa Rica	0.081	0.064	0.035	0.028	0.052	0.044	-21.0	-18.7	-14.3	yes
LAC	Cuba	0.127	0.115	0.055	0.051	0.086	0.076	-9.6	-6.5	-11.1	yes
LAC	Dominica	0.094	0.076	0.049	0.039	0.083	0.070	-19.1	-21.6	-15.4	yes
LAC	Dominican Republic	0.165	0.148	0.087	0.077	0.129	0.117	-10.0	-11.9	-9.4	yes
LAC	Ecuador	0.158	0.139	0.083	0.073	0.092	0.083	-12.1	-11.5	-9.8	yes
LAC	El Salvador	0.100	0.079	0.056	0.044	0.087	0.068	-21.1	-21.6	-21.7	yes
LAC	Grenada	0.071	0.057	0.035	0.028	0.071	0.059	-19.6	-20.6	-16.9	yes
LAC	Guatemala	0.136	0.119	0.075	0.064	0.139	0.126	-12.8	-14.3	-9.3	yes
LAC	Guyana	0.211	0.227	0.115	0.124	0.193	0.204	7.2	8.0	5.3	yes
LAC	Haiti	0.137	0.127	0.072	0.067	0.160	0.144	-7.1	-5.7	-9.7	yes
LAC	Honduras	0.164	0.141	0.091	0.075	0.132	0.102	-13.7	-17.7	-22.9	yes
LAC	Jamaica	0.096	0.093	0.046	0.045	0.059	0.058	-3.4	-1.1	-0.6	yes
LAC	Mexico	0.129	0.122	0.073	0.069	0.114	0.111	-5.6	-5.7	-2.9	yes
LAC	Nicaragua	0.211	0.192	0.115	0.105	0.194	0.180	-8.6	-8.4	-7.4	yes
LAC	Panama	0.242	0.243	0.127	0.130	0.240	0.236	0.6	2.3	-1.6	no
LAC	Paraguay	0.254	0.236	0.134	0.126	0.184	0.177	-6.9	-6.3	-4.1	yes

---

LAC	Peru	0.139	0.129	0.079	0.073	0.142	0.135	-7.2	-6.9	-4.9	yes
LAC	Puerto Rico	0.088	0.089	0.037	0.038	0.057	0.058	1.5	2.6	1.9	yes
LAC	St. Kitts and Nevis	0.063	0.063	0.035	0.034	0.066	0.062	-0.6	-1.4	-5.2	yes
LAC	St. Lucia	0.069	0.053	0.034	0.027	0.062	0.053	-22.4	-20.7	-15.6	yes
LAC	St. Vincent and G.	0.099	0.090	0.051	0.046	0.090	0.082	-9.1	-9.4	-8.9	yes
LAC	Suriname	0.398	0.385	0.155	0.149	0.227	0.230	-3.3	-3.8	1.1	no
LAC	Trinidad/+Tobago	0.102	0.079	0.055	0.043	0.050	0.034	-23.0	-22.3	-31.1	yes
LAC	Venezuela	0.183	0.175	0.102	0.097	0.121	0.115	-4.2	-4.7	-5.0	yes
MENA	Algeria	0.175	0.172	0.096	0.095	0.133	0.131	-1.9	-0.6	-1.1	yes
MENA	Bahrain	0.034	0.025	0.013	0.010	0.036	0.026	-24.2	-23.9	-26.8	yes
MENA	Djibouti	0.331	0.367	0.124	0.135	0.123	0.135	11.1	9.2	9.4	yes
MENA	Egypt	0.186	0.178	0.096	0.091	0.189	0.184	-4.4	-5.5	-2.6	yes
MENA	Iran	0.101	0.090	0.054	0.048	0.102	0.095	-10.7	-11.8	-6.8	yes
MENA	Iraq	0.160	0.146	0.083	0.075	0.121	0.114	-8.7	-10.0	-5.3	yes
MENA	Israel	0.108	0.100	0.049	0.045	0.086	0.081	-7.7	-8.8	-5.5	yes
MENA	Jordan	0.137	0.127	0.078	0.072	0.160	0.149	-7.5	-7.8	-7.0	yes
MENA	Libya	0.215	0.216	0.121	0.121	0.169	0.168	0.2	0.0	-0.8	no
MENA	Morocco	0.174	0.164	0.088	0.085	0.095	0.092	-5.6	-3.0	-2.7	yes
MENA	Oman	0.118	0.112	0.066	0.064	0.103	0.104	-4.8	-3.5	0.1	no
MENA	Qatar	0.101	0.092	0.058	0.052	0.075	0.069	-8.9	-9.4	-8.8	yes
MENA	Saudi Arabia	0.119	0.116	0.067	0.065	0.082	0.079	-2.6	-2.3	-3.3	yes
MENA	Syria	0.136	0.130	0.060	0.059	0.078	0.077	-4.6	-1.7	-1.0	yes
MENA	Tunisia	0.166	0.149	0.077	0.072	0.079	0.075	-9.9	-6.8	-5.7	yes
MENA	U.A. Emirates	0.134	0.135	0.033	0.031	0.052	0.054	1.0	-5.6	3.8	no
MENA	Yemen	0.206	0.185	0.105	0.093	0.140	0.125	-9.9	-11.4	-10.6	yes
NA	Canada	0.213	0.211	0.121	0.120	0.060	0.057	-1.0	-1.2	-4.6	yes
NA	United States	0.129	0.127	0.072	0.071	0.089	0.088	-1.8	-1.6	-1.5	yes
SA	Afghanistan		0.170		0.088		0.166				
SA	Bangladesh	0.049	0.059	0.023	0.029	0.033	0.045	20.3	25.9	38.6	yes
SA	Bhutan	0.163	0.180	0.079	0.102	0.158	0.173	10.3	28.8	9.6	yes
SA	India	0.169	0.161	0.095	0.089	0.077	0.074	-5.0	-5.6	-3.7	yes
SA	Nepal	0.130	0.112	0.069	0.060	0.105	0.093	-13.2	-13.0	-12.2	yes
SA	Pakistan	0.215	0.213	0.119	0.119	0.107	0.106	-1.1	-0.3	-0.6	yes
SA	Sri Lanka	0.169	0.125	0.089	0.066	0.102	0.078	-26.2	-26.0	-23.1	yes
SSA	Angola	0.292	0.264	0.107	0.111	0.153	0.137	-9.5	3.5	-10.4	no

---

SSA	Benin	0.323	0.311	0.154	0.155	0.199	0.212	-3.6	0.7	6.8	no
SSA	Botswana	0.238	0.239	0.126	0.132	0.144	0.150	0.7	5.3	4.1	yes
SSA	Burkina Faso	0.148	0.151	0.064	0.072	0.169	0.164	1.7	12.3	-3.2	no
SSA	Burundi	0.273	0.270	0.105	0.108	0.172	0.172	-1.1	2.0	-0.2	no
SSA	Cameroon	0.124	0.119	0.068	0.065	0.107	0.104	-4.2	-4.9	-3.0	yes
SSA	Cape Verde	0.147	0.108	0.080	0.059	0.132	0.109	-26.5	-26.5	-17.3	yes
SSA	Central African Rep.	0.300	0.287	0.078	0.074	0.098	0.088	-4.2	-5.6	-10.2	yes
SSA	Chad	0.136	0.166	0.037	0.058	0.063	0.105	22.3	54.1	66.1	yes
SSA	Comoros	0.113	0.088	0.049	0.039	0.038	0.030	-21.5	-20.9	-21.1	yes
SSA	Côte d'Ivoire	0.122	0.139	0.064	0.078	0.146	0.140	13.6	21.4	-3.6	no
SSA	Dem. Rep. of Congo	0.281	0.277	0.105	0.109	0.113	0.118	-1.5	3.5	4.5	no
SSA	Equatorial Guinea	0.417	0.268	0.156	0.115	0.188	0.154	-35.7	-26.2	-18.3	yes
SSA	Eritrea	0.294	0.283	0.108	0.114	0.097	0.099	-3.6	5.8	1.9	no
SSA	Ethiopia	0.445	0.408	0.160	0.164	0.077	0.091	-8.5	2.7	19.0	no
SSA	Gabon	0.202	0.187	0.109	0.100	0.160	0.151	-7.4	-8.4	-5.6	yes
SSA	Gambia	0.447	0.408	0.163	0.158	0.102	0.102	-8.8	-2.9	-0.4	yes
SSA	Ghana	0.184	0.173	0.099	0.094	0.116	0.111	-5.8	-5.5	-4.1	yes
SSA	Guinea	0.359	0.360	0.126	0.130	0.112	0.107	0.3	3.0	-4.8	no
SSA	Guinea-Bissau	0.381	0.402	0.126	0.125	0.097	0.100	5.6	-0.2	3.1	no
SSA	Kenya	0.358	0.344	0.144	0.141	0.139	0.127	-4.0	-2.4	-8.6	yes
SSA	Lesotho	0.199	0.165	0.094	0.083	0.186	0.147	-17.3	-11.9	-20.9	yes
SSA	Liberia	0.151	0.207	0.050	0.088	0.097	0.146	37.6	76.4	49.6	yes
SSA	Madagascar	0.116	0.120	0.057	0.059	0.138	0.142	3.6	4.9	3.1	yes
SSA	Malawi	0.165	0.143	0.089	0.074	0.166	0.153	-13.3	-17.3	-8.1	yes
SSA	Mali	0.436	0.416	0.137	0.149	0.063	0.078	-4.5	9.0	24.4	no
SSA	Mauritania	0.280	0.293	0.079	0.089	0.060	0.068	4.7	12.4	13.3	yes
SSA	Mauritius	0.234	0.238	0.115	0.115	0.056	0.052	1.5	0.1	-7.8	no
SSA	Mozambique	0.142	0.145	0.069	0.072	0.148	0.162	2.5	4.5	9.5	yes
SSA	Namibia	0.122	0.130	0.067	0.072	0.116	0.113	6.3	8.6	-2.5	no
SSA	Niger	0.410	0.401	0.134	0.135	0.076	0.076	-2.3	0.6	0.7	no
SSA	Nigeria	0.221	0.204	0.121	0.112	0.202	0.181	-7.8	-7.7	-10.1	yes
SSA	Republic of Congo	0.197	0.235	0.083	0.111	0.176	0.206	19.3	33.0	17.1	yes
SSA	Rwanda	0.179	0.185	0.092	0.093	0.156	0.157	3.4	1.2	0.8	yes
SSA	Sao Tome and P.	0.071	0.064	0.027	0.025	0.011	0.010	-9.7	-8.0	-9.8	yes
SSA	Senegal	0.298	0.284	0.134	0.130	0.209	0.204	-4.7	-2.5	-2.3	yes

---

SSA	Sierra Leone	0.364	0.392	0.123	0.133	0.097	0.106	7.8	7.6	9.1	yes
SSA	Somalia		0.211		0.077		0.091				
SSA	South Africa	0.195	0.197	0.102	0.102	0.139	0.135	0.9	-0.1	-2.7	no
SSA	South Sudan		0.056		0.028		0.059				
SSA	Sudan	0.244	0.218	0.125	0.111	0.240	0.214	-10.6	-11.4	-10.6	yes
SSA	Swaziland	0.081	0.071	0.043	0.037	0.070	0.066	-12.5	-13.1	-6.4	yes
SSA	Tanzania	0.283	0.283	0.117	0.120	0.134	0.138	0.0	2.1	3.0	yes
SSA	Togo	0.178	0.178	0.086	0.081	0.185	0.190	0.3	-5.8	2.4	no
SSA	Uganda	0.240	0.239	0.097	0.104	0.190	0.194	-0.5	7.9	2.1	no
SSA	Zambia	0.195	0.185	0.103	0.097	0.198	0.191	-5.3	-6.4	-3.8	yes
SSA	Zimbabwe	0.365	0.375	0.129	0.134	0.102	0.106	2.6	3.6	3.5	yes

---

**Table A.5:** Determinants of regional inequality; OLS regressions, means (2001-2011), CV

	Dependent variable: coefficient of variation (CV)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(GDP p.c.)	0.350 (0.541)	0.601 (0.778)	0.267 (0.581)	0.340 (0.707)	1.243 (1.030)	0.810 (0.774)	0.248 (0.570)
log(GDP p.c.) <sup>2</sup>	-0.052 (0.062)	-0.085 (0.090)	-0.043 (0.066)	-0.052 (0.078)	-0.149 (0.116)	-0.103 (0.088)	-0.038 (0.065)
log(GDP p.c.) <sup>3</sup>	0.002 (0.002)	0.004 (0.003)	0.002 (0.002)	0.002 (0.003)	0.006 (0.004)	0.004 (0.003)	0.002 (0.002)
log(Area)	0.013*** (0.004)	0.004 (0.007)	0.012** (0.005)	0.014*** (0.005)	0.019*** (0.004)	0.003 (0.008)	0.010 (0.006)
log(# Regions)	-0.032* (0.019)	-0.030 (0.026)	-0.034 (0.023)	-0.018 (0.018)	-0.056** (0.023)	-0.004 (0.029)	-0.027 (0.025)
log(Area)/log(# Regions)	-0.018* (0.009)	-0.016 (0.014)	-0.019* (0.010)	-0.014 (0.009)	-0.025*** (0.008)	0.003 (0.017)	-0.015 (0.012)
Population density	-0.126** (0.058)	-0.078* (0.041)	-0.119 (0.089)	-0.095* (0.051)	-0.085** (0.036)	-0.794*** (0.285)	-0.111** (0.054)
Arable land (%)	-0.166*** (0.037)	-0.103* (0.053)	-0.175*** (0.039)	-0.177*** (0.032)	-0.036 (0.061)	-0.176*** (0.039)	-0.154*** (0.039)
Road density		-0.040* (0.022)					
(Road density) <sup>2</sup>		0.005 (0.004)					
Paved roads (%)		-0.064** (0.025)					
Trade/GDP			-0.001 (0.019)				
FDI/GDP			-0.058 (0.100)				
Landlocked			-0.005 (0.015)				
Mineral rents / GDP				-0.169 (0.249)			
Oil rents / GDP				0.031 (0.054)			
EU 27 member					-0.014 (0.015)		
Social expense / revenue					-0.127*** (0.047)		
Bureaucratic quality						0.001 (0.011)	
Polity2						0.003 (0.015)	
Federal dummy						0.033** (0.014)	
Ethnic fractionalization							0.050* (0.026)
Constant	-0.400 (1.566)	-0.906 (2.241)	-0.122 (1.682)	-0.396 (2.133)	-3.113 (3.021)	-1.771 (2.250)	-0.186 (1.645)
Observations	178	125	171	132	93	126	171
R-squared	0.412	0.527	0.416	0.508	0.498	0.529	0.416

Robust standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table A.6:** Determinants of regional inequality; OLS regressions, means (2001-2011), Gini

	Dependent variable: gini coefficient of variation (GINI)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(GDP p.c.)	0.334 (0.207)	0.437 (0.280)	0.293 (0.220)	0.387 (0.280)	0.952*** (0.325)	0.409 (0.279)	0.289 (0.215)
log(GDP p.c.) <sup>2</sup>	-0.040* (0.024)	-0.054* (0.033)	-0.036 (0.025)	-0.047 (0.031)	-0.109*** (0.038)	-0.048 (0.032)	-0.034 (0.025)
log(GDP p.c.) <sup>3</sup>	0.002* (0.001)	0.002* (0.001)	0.001 (0.001)	0.002 (0.001)	0.004*** (0.001)	0.002 (0.001)	0.001 (0.001)
log(Area)	0.007*** (0.002)	0.003 (0.003)	0.006*** (0.002)	0.008*** (0.003)	0.009*** (0.002)	0.004 (0.003)	0.005** (0.003)
log(# Regions)	-0.014* (0.008)	-0.014 (0.010)	-0.014 (0.009)	-0.009 (0.009)	-0.024** (0.010)	-0.002 (0.010)	-0.012 (0.010)
log(Area)/log(# Regions)	-0.009*** (0.003)	-0.009* (0.005)	-0.010** (0.004)	-0.008* (0.004)	-0.012*** (0.004)	-0.001 (0.006)	-0.008* (0.005)
Population density	-0.063*** (0.022)	-0.038** (0.018)	-0.065* (0.038)	-0.046** (0.021)	-0.054*** (0.015)	-0.253** (0.107)	-0.057*** (0.021)
Arable land (%)	-0.067*** (0.016)	-0.027 (0.021)	-0.072*** (0.017)	-0.085*** (0.015)	-0.019 (0.025)	-0.082*** (0.015)	-0.063*** (0.017)
Road density		-0.022** (0.010)					
(Road density) <sup>2</sup>		0.003* (0.002)					
Paved roads (%)		-0.033*** (0.011)					
Trade/GDP			0.000 (0.008)				
FDI/GDP			-0.039 (0.047)				
Landlocked			-0.004 (0.006)				
Mineral rents / GDP				-0.087 (0.080)			
Oil rents / GDP				0.009 (0.024)			
EU 27 member					-0.009 (0.007)		
Social epense / revenue					-0.065*** (0.021)		
Bureaucratic quality						0.000 (0.004)	
Polity2						0.002 (0.006)	
Federal dummy						0.017*** (0.006)	
Ethnic fractionalization							0.020* (0.012)
Constant	-0.782 (0.593)	-0.981 (0.804)	-0.635 (0.629)	-0.932 (0.834)	-2.614*** (0.926)	-1.039 (0.791)	-0.683 (0.618)
Observations	178	125	171	132	93	126	171
R-squared	0.373	0.513	0.379	0.505	0.542	0.564	0.370

Robust standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05,  
\* p<0.1

**Table A.7:** Determinants of sigma convergence, panel fixed effects regressions, 5-year averages, CV

	Dependent variable: coefficient of variation (CV)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(GDP p.c.)	0.560*** (0.183)	0.448 (0.383)	0.339* (0.194)	0.048 (0.220)	0.368 (0.387)	0.604*** (0.163)
log(GDP p.c.) <sup>2</sup>	-0.062*** (0.021)	-0.049 (0.043)	-0.038* (0.022)	-0.005 (0.025)	-0.045 (0.042)	-0.069*** (0.019)
log(GDP p.c.) <sup>3</sup>	0.002** (0.001)	0.002 (0.002)	0.001 (0.001)	0.000 (0.001)	0.002 (0.002)	0.003*** (0.001)
Population density	-0.355** (0.146)	-0.141 (0.228)	-0.278** (0.124)	-0.115 (0.128)	-2.043 (1.386)	-0.219 (0.158)
Arable land (%)	-0.199*** (0.053)	-0.180*** (0.063)	-0.176*** (0.054)	-0.149** (0.061)	-0.191*** (0.062)	-0.163*** (0.057)
Road density		-0.003 (0.016)				
(Road density) <sup>2</sup>		-0.001 (0.003)				
Paved roads (%)		0.020** (0.010)				
Trade/GDP			0.014** (0.006)			
FDI/GDP			0.019 (0.016)			
Mineral rents / GDP				0.056 (0.037)		
Oil rents / GDP				-0.002 (0.022)		
Social expense / revenue					-0.031 (0.023)	
Bureaucratic quality						-0.000 (0.001)
Polity2						0.005 (0.005)
Constant	-1.384*** (0.519)	-1.158 (1.132)	-0.715 (0.567)	0.115 (0.644)	-0.726 (1.177)	-1.485*** (0.468)
Observations	865	312	814	643	331	485
# Country	178	155	172	134	111	127
R-squared-w	0.275	0.089	0.327	0.396	0.528	0.356
R-squared-b	0.304	0.059	0.325	0.411	0.101	0.397
Country FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses \*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1



**Table A.8:** Determinants of sigma convergence, panel fixed effects regressions, 5-year averages, Gini

	Dependent variable: Gini coefficient (GINI)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(GDP p.c.)	0.384*** (0.072)	0.355* (0.189)	0.322*** (0.091)	0.184** (0.075)	0.218 (0.267)	0.376*** (0.049)
log(GDP p.c.) <sup>2</sup>	-0.042*** (0.009)	-0.038* (0.021)	-0.035*** (0.010)	-0.020** (0.009)	-0.025 (0.029)	-0.042*** (0.006)
log(GDP p.c.) <sup>3</sup>	0.001*** (0.000)	0.001 (0.001)	0.001*** (0.000)	0.001** (0.000)	0.001 (0.001)	0.001*** (0.000)
Population density	-0.164** (0.071)	-0.053 (0.116)	-0.147** (0.069)	-0.055 (0.054)	-1.388* (0.707)	-0.152 (0.113)
Arable land (%)	-0.061*** (0.017)	-0.079** (0.032)	-0.055*** (0.017)	-0.047* (0.025)	-0.036 (0.023)	-0.053** (0.023)
Road density		-0.006 (0.008)				
(Road density) <sup>2</sup>		-0.000 (0.001)				
Paved roads (%)		0.008 (0.005)				
Trade/GDP			0.005** (0.002)			
FDI/GDP			0.008 (0.008)			
Mineral rents / GDP				0.037 (0.024)		
Oil rents / GDP				0.011 (0.013)		
Social expense / revenue					0.004 (0.013)	
Bureaucratic quality						-0.000 (0.001)
Polity2						-0.001 (0.002)
Constant	-1.042*** (0.198)	-0.996* (0.558)	-0.857*** (0.260)	-0.440** (0.213)	-0.526 (0.806)	-0.996*** (0.128)
Observations	865	312	814	643	331	485
# Country	178	155	172	134	111	127
R-squared-w	0.223	0.260	0.261	0.319	0.400	0.316
R-squared-b	0.259	0.120	0.252	0.411	0.064	0.449
Country FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses \*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1