

Author's Accepted Manuscript

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

Christian Lessmann, André Seidel



PII: S0014-2921(16)30221-5
DOI: <http://dx.doi.org/10.1016/j.euroecorev.2016.11.009>
Reference: EER2930

To appear in: *European Economic Review*

Received date: 4 December 2015

Accepted date: 26 November 2016

Cite this article as: Christian Lessmann and André Seidel, Regional Inequality Convergence, and its Determinants – A View from Outer Space, *European Economic Review*, <http://dx.doi.org/10.1016/j.euroecorev.2016.11.009>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

REGIONAL INEQUALITY, CONVERGENCE, AND ITS DETERMINANTS – A VIEW FROM OUTER SPACE[☆]

Christian Lessmann^{*} and André Seidel¹

^aTU Braunschweig, Ifo Institute Dresden Branch & CESifo

^bTU Dresden, Germany

c.lessmann@tu-braunschweig.de

andre.seidel1@tu-dresden.de

^{*} Corresponding author. Tel.: +49 531 391-2592.

Abstract

This paper provides a new dataset of regional income inequalities within countries based on satellite nighttime light data. First, we empirically study the relationship between luminosity data and regional incomes for those countries for which regional income data are available. Second, we use our estimation results for an out-of-sample prediction of regional incomes based on the luminosity data. These results enable us to investigate regional income differentials in developing countries that lack official income data. Third, we calculate commonly used measures of regional inequality within countries based on predicted incomes. An investigation of changes in the dispersion of regional incomes over time reveals that approximately 67-70% of all countries experience sigma-convergence. Forth, we study different major determinants of within-country changes in inequality, i.e., the determinants of the convergence process. We find evidence for an N-

[☆] We are grateful to the participants in research seminars and conferences at the Friedrich Schiller Universität Jena, the 5th Workshop on Regional Economics at the Ifo Institute Dresden and the annual meeting of the development economics group of the German Economic Association in Heidelberg for their many helpful comments. We also thank Rafael La Porta for providing us regional income data and shape files.

¹ Tel.: +49 351 463-33972.

shaped relationship between development and regional inequality. Resources, mobility, trade openness, aid, federalism and human capital are also very important.

JEL: D30, E01, E23 O11, O15, O57, R10.

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

Accepted manuscript

1. Introduction

In recent decades, the regional distribution of incomes within countries has attracted considerable interest among academics and policy makers. Important research questions include, among others the following: 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, conflict, ethnicity, and geography? Because these questions are obviously important for the economy in particular and society in general, many empirical studies on these issues have been carried out with interesting and instructive results. However, all these studies are limited to a particular country sample with a general bias toward middle- and high-income economies. The major difficulty encountered in conducting this research is the availability of regional income data. Although 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. However, the poorest regions of the world remain blank spots on the map, limiting the informative value of study results to particular country samples. The aim of our study is to fill this gap using satellite nighttime light data to develop 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 as follows: Most economic activities of consumption and production that take place in the evening or at night require light. We can expect that the higher a country's nighttime light intensity, the higher its level of economic activity, i.e., the higher its income. Therefore, 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. Although this issue might be minor for country-level income data (see also Johnson et al. 2013), it is major at the regional level. Regional data are often lacking in developing countries, whose statistical authorities have low capacities and standards. 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° latitude \times 1° longitude), particularly in this group of countries. Consequently, recent studies, such as Besley and Reynal-Querol (2014) or Hodler and Raschky (2014), use these sorts of data to proxy regional income levels. We follow this literature to construct new data on regional income inequality within countries, which we provide at <http://dx.doi.org/10.7802/1339>.

We use the luminosity data to predict regional incomes per capita at a sub-national level for 180 countries from 1992-2012. Our prediction of regional incomes through nighttime light data uses existing regional income data to obtain an estimate of the light-income

relationship at the regional level, which could be used in out-of-sample predictions for those regions for which income data are unavailable. Based on these predictions, we calculated different measures of regional income inequality, for example, the population-weighted Gini coefficient and different generalized entropy indexes, which easily translate to the inequality measures recommended by Atkinson (1970).

Our study's primary innovation is that we analyze the dynamics of regional income inequality and its determinants. Unlike previous studies, we are not limited to a particular country sample. First, we use the data on regional inequality to analyze sigma-convergence within countries. Whereas beta-convergence focuses on the (better) growth performance of initially poor regions within a country, sigma-convergence is the decrease in the dispersion of regional incomes. By comparing the changes in our inequality measures between 1992–2001 and 2002–2012, we find that more than 67% of all countries experience sigma-convergence. However, a significant number of countries—developing countries (e.g., Mozambique and Bangladesh) and industrial economies (e.g., Sweden and Russia) alike—are experiencing increasing inequality.

Next, we use our data to investigate the determinants of regional inequality and convergence. Using panel fixed-effects regressions that focus on within-country variations in the data, we find that in very poor countries—e.g., countries in the center of Africa such as Chad and the Democratic Republic of the Congo—development has an increasing effect on regional inequalities. However, in countries that could be called middle- or high-income countries in their respective country groups, such as Zambia and South Africa, development has a decreasing effect on 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. Furthermore, we find increasing inequalities at very high levels of economic development. Therefore, the relationship is ultimately N-shaped in our data, i.e., it exhibits 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 the impact of several economic and political determinants on regional inequality: natural resources, trade openness, transportation costs, aid payments and ethnic income inequality are positively correlated with regional inequality, whereas the share of arable land, federalism and human capital are negatively correlated with regional inequality. In many respects, our results support earlier studies in the field. However, we now can analyze regions of low-income countries, which helps us generalize these findings. Our empirical results do not suffer from a potential sample selection bias. Here, we emphasize that in this part of the analysis we conduct only OLS regressions; therefore, our results document only statistical correlations, which may not

represent causal relationships. Therefore, we abstain from strong 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 first explain the methodology that we apply to construct regional income proxies from luminosity data. Thereby, we also discuss several important measurement issues that are relevant when working with satellite data in a regional context. We then use the predicted incomes to calculate different measures of regional inequality within countries. In section 3, we analyze differences in regional inequality across countries and its changes over time, i.e., sigma-convergence. In section 4, we regress the inequality measures on selected explanatory variables to study the determinants of the within-country variation in regional inequality. In section 5, we summarize our main findings and conclude.

2. Measuring Regional Inequality with Luminosity Data

To obtain measures of regional inequality for those countries that lack reliable regional income data, we follow a two-step procedure. The first step of our analysis is to predict regional incomes using night light data (section 2.2);² the second step is to calculate inequality measures based on the predicted incomes (section 2.3). In this regard, we deviate from existing studies such as Alesina et al. (2016) and Mveyange (2015), who directly focus on light dispersion measures as a proxy for income dispersion measures. We adopt this approach for two reasons. First, the previous literature finds a log-linear relationship between light density and income. A constant light-income elasticity implies that the relationship between light levels and income levels depends on the light level itself. The same absolute difference in luminosity between two regions, therefore, does not imply the same difference in the incomes of two other regions that are, on average, either richer or poorer. Therefore, inequality measures calculated directly from light density data do not always provide an adequate measure of inequality in regional income. Second and most importantly, the literature indicates that in some cases, light density alone is a poor indicator for income, particularly in regions with a very low output density (Chen and Nordhaus 2011). We can mitigate these types of problems when we use predicted regional incomes instead of pure light data to calculate inequality

² Note that we use income, output, and GDP synonymously because it is impossible to distinguish between consumption and production using the underlying data. We assume that all relevant variables are highly correlated.

measures. The advantage of our approach is that we can include further proxies for regional incomes in our income prediction such as the number of low-coded pixels or the national income. The latter helps us center the predicted regional income data within countries on the national average. As a result, inequality measures based on predicted incomes are much more closely related to inequality measures based on observed income data, whereas pure light dispersion measures show a less robust relationship (section 2.4). Therefore, we are confident that our inequality measures are also reliable for countries that lack regional income data and inequality measures.

2.1. Data

Nighttime Light

The data that are used to measure nighttime light intensities come from meteorological satellites of the U.S. Air Force. Those satellites orbit the earth 14 times per 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). Adjustment of the raw data is necessary to compensate, for example, for local cloud coverage, atmospheric dust, and changes in satellite and sensor technology. The aim of these operations is to measure only manmade 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 kilometers (at the equator). The final data, which are referred to as the “stable lights product”, are available from 1992 onward on an annual basis. One problem with the light data is the censoring of the data at 0 and 63, which we will later discuss in detail.

Sub-national 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 sub-national 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 2.2 for details). Our final dataset is based on 180 countries and 3,166 regions.

Observed Regional Income

In the next section, we investigate the relationship between regional income levels and regional nighttime lights. For this purpose, we use the dataset provided by Gennaioli et al. (2014), which contains information about regional income per capita. The dataset includes 82 countries (1,503 regions) and covers an average time span of 32 years. Note that the panel is highly unbalanced. Income data are based on 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 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). Note that the administrative boundaries used by Gennaioli et al. (2014) differ from the GADM data in a few cases. In the first step of our analysis, the administrative areas are defined as in Gennaioli et al. (2014); in the second step, they are defined by the GADM project.

Population 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 up the results at the regional level; therefore, the minimum population of a region is one. The original data were sampled in 5-year waves; therefore, we interpolate missing values to obtain annual data. Therefore, short-term fluctuations in the population, for example, caused by conflicts and natural disaster-induced migration, are not covered by the data.

Table B.1 in Appendix B provides summary statistics for all the variables generated from the geo-coded data. Note that we refer to only those countries that are ultimately included in our empirical analysis. Table B.2 provides the summary statistics of the relevant variables used from the data provided by Gennaioli et al. (2014). Finally, table B.3 provides further details on the sources of the data used in the analysis.

2.2. Predicting Regional Incomes

2.2.1 Methodology

In the first step of our analysis, we exploit the variations in the data within and across countries and regions to obtain an estimate of the relationship between light and income. We need these results as inputs in our forecasting model of regional income for those

countries for which the regional income data are not available but satellite data as a proxy are available.

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 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 in developing countries 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 at both the national and sub-national levels (grid cells of 1° latitude \times 1° 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 for which standard sources are either unavailable or of poor quality. However, the literature notes several measurement problems that we must consider 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 the case for only a small fraction of pixels. However, it is noteworthy that in a few areas there is no variation in the luminosity data, whereas there are significant differences in socioeconomic variables. For example, the core metropolitan areas of New York City and Mexico City are top-coded, although there are large income differences between the two regions. Not least for this reason, we can expect that the luminosity data are a less appropriate proxy for income in richer countries than in poorer ones. We address this issue in the forthcoming regression analysis by including the number of top coded pixels in a region as an additional light-based income determinant.

Another measurement problem occurs at the other end of the distribution of light intensities. There are many zero-coded pixels in the data, although there might still be some production. The sensitivity of satellite sensors is not high enough (or, more precisely, the average gain given to the signal is too low) and therefore, low light emissions are not captured. This is obvious for deserts or mountainous regions. However, Chen and Nordhaus (2011) note that low-coding is a general problem for

regions with a low output density. Consequently, light density is a poor proxy for production in those regions. We address this issue as implicitly recommended by Chen and Nordhaus (2011) and add further indicators of regional income.³ We consider the number of low-coded pixels within a region, average country size, the number of regions within countries and their interaction.⁴ Note that we focus on a log-log specification and therefore, we follow Hodler and Raschky (2014) and add 0.01 to the average luminosity in those regions for which the value otherwise would be zero.

Finally, we consider satellite-fixed effects because changes in satellite constellations, sensor technology and sensors are aging over time, which affects the sensitivity of the light measurement.⁵

Based on the discussion above, we estimate the following random effects model:

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

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 per capita; $tc_{i,j,t}$ is the number of top-coded pixels (DN = 63) within a region; $lc_{i,j,t}$ is the number of low-coded, dark pixels within a region (DN = 0); r_j is the number of regions within a country; a_j is the overall size of a country in square kilometers; η_k are time-invariant fixed effects for k different lending group regions of the world as defined by the World Bank (North America (NA) is used as reference group);⁶ θ_t are the satellite configuration fixed effects, which change over time (but not always on an annual basis); v_i is a regional random effect with $E[v_i] = 0$; and $\varepsilon_{i,j,t}$ is the

³ In a recent publication, Keola et al. (2015) show that changes in the land used for agricultural production can be a good proxy for income in regions with a low output density. Data on changes in the land used for agricultural are, however, available only from 2000 onward, which would cut our sample in half.

⁴ This might also reduce potential biases coming from the fact that administrative boundaries slightly differ between the data provided by Gennaioli et al. (2014) and our GADM-based data (see section 2.1).

⁵ 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 relevant data.

⁶ 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), South Asia (SA), and Sub-Saharan Africa (SSA).

error term. We use all the variables in a logarithmic transformation; therefore, the coefficients can be interpreted as elasticities (except for ϑ as a coefficient of an interaction variable). We cluster standard errors at the country level because the errors are correlated within countries. We additionally estimated standard errors accounting for spatial correlation of an unknown form using the generalized method of moments suggested by Conley (1999). The two approaches yield similar estimates; for further details, see table B.4 in Appendix B.

Our empirical model requires further discussion. In particular, we need to justify 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, for which the respective dummy variables capture the country-specific unobserved heterogeneity between countries. In addition, 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.⁷ 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 either 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 various country groups.

2.2.2 Estimation

The results of different specifications of eq. [1] are reported in table 1. Each column reports the results obtained 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

⁷ Note that we also estimated fixed-effects models for which we proxy the individual intercepts with the “average intercept” of those countries, which were comparable in terms of income and geography. The results are comparable to those obtained from our preferred specification (table 1, column 7).

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)-(7) 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, area, its interaction, and the number of top- and low-coded pixels within a region.

In all 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 0.399, implying that a 10% increase in luminosity is associated with a 3.9% 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 0.153. Note that we have not considered any further potential determinants of regional incomes. In column (4), we add the country-level income, which decreases $\hat{\beta}$ to 0.122.

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 0.232. 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 0.104. In column (7), we add further country-level variables: country size, number of regions, their interaction, and the number of top-coded (low-coded) pixels within a region. The coefficient $\hat{\beta}$ is 0.102, which is very close to the country fixed-effects estimate discussed above. Note that the effect of light on income now also depends on the coefficients δ and ζ , which capture the effects of the number of top- and low-coded pixels. As expected, regional incomes increase in the number of top-coded pixels and decrease in the number of low-coded pixels. The high significance levels and the size of the coefficients suggest that this correction is very important to obtaining meaningful estimates of regional income. If we compare these results with the country-fixed effects specification (column 4), we find that our country-level, time-invariant control variables together with the country-group fixed effects yield results that 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.

Accepted manuscript

Table 1: Regression results of equation [1]

	Dependent variable: $\log(\text{GDPpc}_i)$						
	Pooled, Region FE, Country FE				Country Group FE		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\log(\text{Light}_i)$	0.399*** (0.025)	0.190*** (0.048)	0.153*** (0.030)	0.122*** (0.026)	0.232*** (0.031)	0.104*** (0.019)	0.102*** (0.022)
$\log(\text{GDPpc}_i)$				0.864*** (0.038)		0.901*** (0.025)	0.889*** (0.026)
$\log(\# \text{ TopCodPix}_i)$							0.004*** (0.001)
$\log(\# \text{ LowCodPix}_i)$							-0.007** (0.003)
$\log(\# \text{ Regions}_i)$							0.396*** (0.126)
$\log(\text{Area}_i)$							0.183*** (0.041)
$\log(\# \text{ Regions}_i) \times \log(\text{Area}_i)$							-0.041*** (0.010)
Constant	8.674*** (0.097)	8.760*** (0.105)	9.177*** (0.048)	1.091*** (0.345)	10.354*** (0.089)	0.827*** (0.249)	-1.030** (0.509)
# Observations	5,258	5,258	5,258	5,258	5,258	5,258	5,258
# Regions	1,504	1,504	1,504	1,504	1,504	1,504	1,504
# Countries	81	81	81	81	81	81	81
R-squared within	0.327	0.580	0.580	0.760	0.579	0.760	0.759
R-squared-between	0.350	0.354	0.865	0.865	0.552	0.840	0.852

Region FE	NO	YES	NO	NO	NO	NO	NO
Country FE	NO	NO	YES	YES	NO	NO	NO
Country Group FE	NO	NO	NO	NO	YES	YES	YES
Satellite FE	NO	YES	YES	YES	YES	YES	YES

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

Next, we briefly discuss the regression diagnostics. We can ultimately use the data for 1,504 regions in 81 countries, providing a maximum of 5,258 region-year observations. In that specification, which considers regional and satellite-fixed effects (column 2), we can explain approximately 58% of the variation in regional incomes within regions and more than 35% of the variation between regions. The model, which uses country-fixed effects and country-level income (column 4), explains approximately 76% of the variation within regions and 87% of the variation between them. In the specification reported in column (7), which provides the coefficients for our forecasting model, we are able to explain more than 76% of the variation in regional income within regions and more than 85% of the variation between them. Therefore, we are confident that our regional income estimates based on luminosity data are suitable proxies for regional real income.

We provide detailed tests of our predictions in Appendix A. There, we concentrate on three issues: 1. Does the income level of countries effect the relationship between regional light and regional income? 2. Are our income predictions meaningful also in low output-density regions? 3. Do differences in regional price levels affect our income predictions? Overall, we find that our predictions for high-income countries are less precise than our predictions for low-income regions. The potential bias caused by low output-density regions is appropriately captured by our forecasting model. Differences in regional price levels are important, at least in the U.S., where we could compare our results considering regional price parities. The latter is a very important result. It suggests that our income predictions based on luminosity data are more appropriate for an analysis of real income differentials within countries compared to observed income data, which is usually based on nominal values. Please see Appendix A for the detailed results.

2.2.3 Regional Income Prediction

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 $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\delta}$, $\hat{\zeta}$, $\hat{\kappa}$, $\hat{\lambda}$, $\hat{\vartheta}$, $\hat{\eta}_k$, and $\hat{\theta}$, as reported in column (7) in table 1, enter

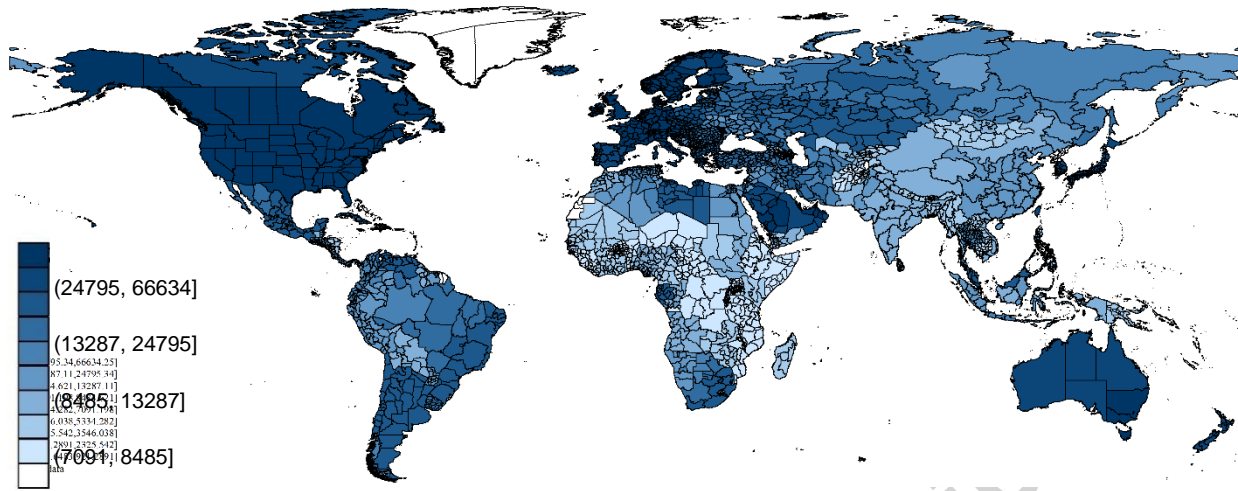
equation [2], yielding the functional relationship that we use to predict regional incomes $\hat{y}_{i,j,t}$:

$$\begin{aligned} \ln(\hat{y}_{i,j,t}) = & -1.030 + 0.102 \cdot \ln(l_{i,j,t}) + 0.889 \cdot \ln(y_{j,t}) + 0.004 \cdot \ln(tc_{i,j,t}) - 0.007 \\ & \cdot \ln(lc_{i,j,t}) \\ & + 0.396 \cdot \ln(r_j) + 0.183 \cdot \ln(a_j) - 0.041 \cdot \ln(r_j) \cdot \ln(a_j) + \eta_k + \theta_{t,i,j,t} \end{aligned}$$

We calculate the luminosity variables $l_{i,j,t}$, $tc_{i,j,t}$, $lc_{i,j,t}$ for all the world's sub-national regions using the 1st-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 $\hat{\beta}$, $l_{i,j,t}$, $\hat{\delta}$, $tc_{i,j,t}$, $\hat{\zeta}$ and $lc_{i,j,t}$. The other parameters affect only the regional income level and not the dispersion within countries.

We emphasize that we change the datasets between equations [1] and [2]. 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 [2]). Our prediction of regional incomes yields a panel dataset of 3,166 regions in 180 countries for the period (1992-2012). The (unweighted) average regional income is 10,189 US\$ (standard deviation 10,776). The richest region is Ad Dawhah in Qatar; the poorest region is Nimba in Liberia. The data cover 99% of the arable global surface and 99% of the gross world product.⁸ Figure 1 illustrates the predicted regional incomes on a world map. The darker a region's color, the higher its regional GDP p.c. All 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 their countries. We could also observe significant income differentials across regions in South America. However, maps that focus on the regional incomes within country groups are more informative. Figure B.1 in Appendix B shows regional income data for Sub-Saharan Africa. Here, we observe much more variation in the data because the scale is more sensitive to income variations in these groups of countries.

⁸ Note that the missing 1% results from the fact that we exclude very small countries (smaller than 50 km²) from the analysis because it makes little sense to analyze regional inequality in those cases.

Figure 1: Predicted regional income (mean 1992-2012)

2.3. Regional Inequality Measures

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 dataset 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 can 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 population-weighted Gini coefficient ($GINIW_j$) and different population-weighted generalized entropy indexes. All the inequality measures account for the different sizes of regions and satisfy the relative income principle (mean independence), population principle, and Pigou-Dalton principle. Omitting time subscripts for reasons of clarity, the inequality measures are given as follows:

$$[1] \quad GINIW_j = \frac{1}{2\hat{y}_j} \sum_i^{n_j} \sum_l^{n_j} \frac{p_i}{P_j} \frac{p_l}{P_j} |\hat{y}_i - \hat{y}_l|, [3,4]$$

$$GE(\alpha)_j = \frac{1}{\alpha^2 - \alpha} \left[\sum_i^{n_j} \frac{p_i}{P_j} \left[\left(\frac{\hat{y}_i}{\hat{y}_j} \right)^\alpha - 1 \right] \right],$$

$$\text{where } \hat{y}_j = \sum_i^{n_j} \frac{p_i}{P_j} \hat{y}_i.$$

We calculate the generalized entropy index for different levels of α , which captures the sensitivity of a specific generalized entropy index to particular parts of the distribution.⁹ For larger positive values of α , the generalized entropy index becomes more sensitive to differences at the top of the distribution; the more negative α is, the more sensitive the generalized entropy index to changes in the distribution that affect the lower tail. We calculate the generalized entropy for $\alpha \in (-1, 0, 1, 2)$. For $\alpha = 0$, the index is equal to the population-weighted mean logarithmic deviation ($MLDW_j$) and for $\alpha = 1$, the index is equal to the population-weighted Theil index ($THEILW_j$). In the cases of $\alpha = 2$, we reshape the index so that it becomes the population-weighted coefficient of variation (CVW_j).¹⁰ All inequality measures are population-weighted using p_i/P_j to give smaller (larger) regions a smaller (larger) weight in the overall inequality measure. Thus, 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 than the country average but are very sparsely populated. Without considering that lower population, Canada appears to be one of the most unequal economies, whereas it is included in the group of countries with the lowest regional inequality if we consider the *GINIW*. Whether to consider population weights is disputed in the literature. The unweighted inequality measures are commonly used in the literature on regional growth within countries (e.g., Barro and Sala-i-Martin), whereas the weighted coefficient of variation has been suggested by Williamson (1965) to compare inequality levels across countries. One difference between the two concepts is important: non-weighted measures are inequalities between spatial units, whereas the weighted measures we use could be interpreted as measures of intergroup inequality in a country, where groups of people are formed by their place of residence. The correlation between unweighted and weighted inequality measures is 0.866 for the Gini coefficient, 0.816 for $GE(-1)$, 0.848 for the mean log deviation, 0.878 for the Theil index and 0.934 for the coefficient of variation. Thus, although population weights are important in some countries, the overall picture on regional inequality does not change substantially.

⁹ Note that the generalized entropy index can be transformed into a subclass of the commonly used Atkinson index with $\epsilon = 1 - \alpha$ for $0 \leq \alpha < 1$, where ϵ defines the (relative) inequality aversion. See Allison (1978) or Cowell (2000) for further details on the properties of the different inequality measures used.

¹⁰ For the concrete calculus of the different generalized entropy index, see Table B.5 in Appendix B.

2.4. Discussion of Measurement Approach

Because we are finally interested in finding a proxy for regional income inequality, we next compare the inequality measures based on predicted incomes with inequality measures based on observed income data, which we take from Gennaioli et al. (2014). Two questions are important here: 1. Are our indicators good proxies for regional inequality? 2. Does our approach using predicted incomes perform better than pure light data? Table 2 shows the results. In column (1), we correlate our inequality measures based on predicted income with inequality measures based on observed income. In column (2), we use inequality measures simply based on light density instead of predicted incomes for this comparison.

First, our inequality measures show a relatively high correlation with the same inequality measures based on the observed data. The correlation ranges between 0.40 and 0.52, depending on the particular inequality index. The remaining differences could have different causes. One reason might be that our income predictions remain imperfect because we over- or underestimate incomes in some regions. We discuss the quality of our predictions extensively in Appendix A, suggesting that our predictions do not suffer seriously from any systematic bias caused by the use of nighttime lights.

Table 2: Correlation between observed predicted and light weighted inequality measures in the cross-section mean of period 2001–2012

Inequality measures	Correlation between inequality measures based on observed income and...	
	predicted income	light density
$GINIW_j$	0.50	0.21
$GEW(-1)_j$	0.40	0.11
$MLDW_j$	0.45	0.21
$THEILW_j$	0.50	0.30
CVW_j	0.52	0.29
# Countries	78	78

Population-weighted Gini coefficient ($GINIW_j$), Generalized Entropy class -1 (GEW_j), mean log deviation ($MLDW_j$), Theil index ($THEILW_j$) and the coefficient of variation (CVW_j)

Apart from these technical issues, we would not expect a perfect correlation between inequality measures based on light-predicted incomes and inequality measures based on observed income because observed regional income data usually involve nominal incomes in constant national prices. Instead, the light data reflects a real figure because output and consumption—both of which are covered by the light data—should depend on purchasing power at the regional level. Lessmann (2014) uses the example of the U.S., where regional price parities are available from official statistics at the state level. The comparison of the weighted coefficient of variation of regional income in nominal terms (at “real” national prices) and the same inequality index based on income data in regional price parities suggests that real income inequalities are much lower than nominal values (0.08 vs. 0.14). This is an important result. In particular, regional prices adjust for nominal income differentials in case of non-tradable goods (property, services, etc.). Therefore, standard measures of regional inequality used in other studies have an upward bias, depending on the flexibility of prices within countries. The bias may vary across countries and depend on country size, infrastructure and certain regulations such as country-wide minimum wages.

With these results in mind, we also take the U.S. example and study the effect of regional price parities on the differences in our income prediction in comparison with observed data (see Appendix A). Interestingly, although predicted state incomes based on lights are higher in low-price regions (e.g., Mississippi) than the observed nominal income data, they are lower in high-price regions (e.g., the District of Columbia). These results suggest that light-based predicted incomes better reflect real figures than nominal figures, which is important when evaluating regional inequalities. If this is the case, we should observe that income inequality based on predicted income is smaller than inequality based on nominal income and closer to inequality based on real income. Indeed, comparing the weighted *GINI* coefficient shows that predicted regional income inequality (0.05) is much lower than nominal inequality (0.08) and even very close to real income inequality (0.04). Therefore, we are confident that our inequality measures are meaningful proxies for real income inequalities within countries. As the U.S. example shows, they might be even better than inequality measures based on observed nominal income data.

Second, column (2) shows the correlation of inequality measures directly based on light densities with the inequality measures based on observed income data. In all cases, the correlation is much lower than seen in our forecast model. The correlations are ≤ 0.3 and show only a weak relationship with inequality measures based on official income statistics. This is not surprising given that we argued at the beginning of section 2 that the relationship between light and income is not linear and that light density alone is in some cases a poor indicator for regional income. In light of this problem, it is also not surprising that in the case of the Generalized Entropy class -1 measure, the correlation is with 0.09 very close to zero, which implies that this measure has little explanatory power for income inequalities. The Generalized Entropy class -1 measure is most sensitive to the lower part of the income distribution. This is exactly the part of the income distribution for which the previous literature would suggest that light alone is the most uninformative. As we see, our forecast model incorporates some of these distortions, therefore improving the informative value of measures of regional inequality based on light data.

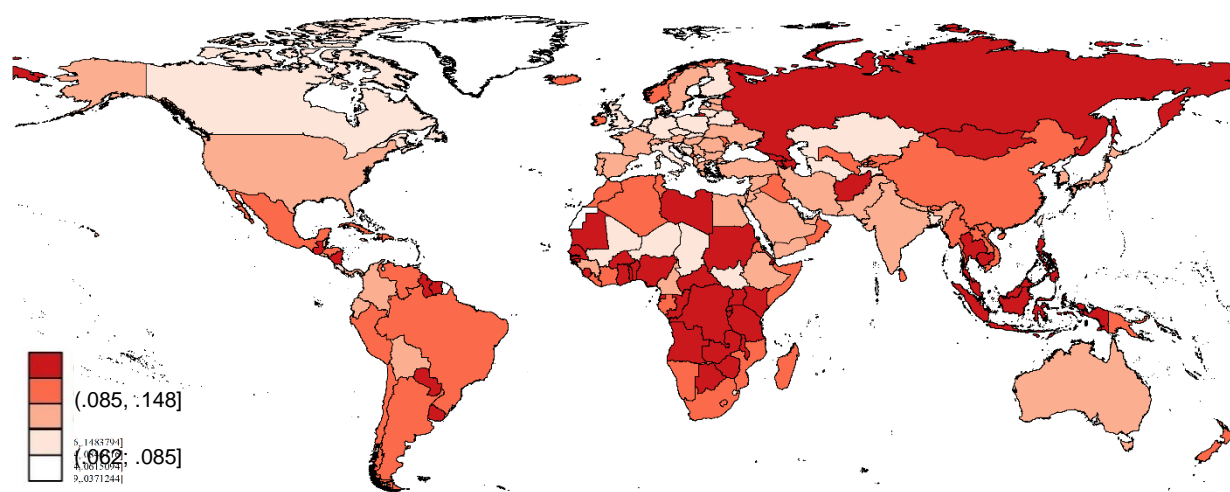
3. A New Dataset on Regional Inequality

In this section, we describe our new dataset on regional inequality, which we derive using the measurement approach developed above. We undertake a descriptive analysis of both the level of regional inequality across countries (section 3.1) and the changes over time (section 3.2). In this, we focus on the relationship between regional inequality and development. The results for single-country cases are provided online in Appendix C. The complete dataset can be downloaded at <http://dx.doi.org/10.7802/1339>.

3.1. Regional Inequality Across Countries

We first discuss the level of sub-national regional inequality across countries. Note that the aim of this section is to give the reader an overview of the extent of regional inequality in various parts of the world. However, the focus of the forthcoming sections will be on changes in regional inequality over time. Figure 2 illustrates the mean of the weighted Gini coefficient (*GINIW*) for the period 1992-2012 (darker colors represent higher levels of inequality).

Figure 2: Regional inequality within countries (*GINIW*, mean (1992-2012))



Industrial countries in North America and the core of Europe have the lowest levels of regional inequality, whereas countries in Latin America, Africa, and East Asia show significantly higher levels of inequality. However, within the different country groups, there are important differences. The results for African countries are particularly interesting because there is no 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 Southwestern and Southeastern Africa, have relatively high inequality. Additionally, the richer countries, e.g., Namibia and South Africa, have relatively low regional inequalities. This observation supports the hypothesis of an inverted U-shaped relationship between regional inequality and the level of economic development, at least in this country group.

Considering different inequality measures does not change the main picture. Table B.6 in the Appendix shows that the correlations between the weighted inequality measures ranges from 0.930 to 0.996. Therefore, we usually obtain very comparable country rankings. However, in single cases such as Algeria and Somalia, the sensitivity of inequality measures to different parts of the income distribution becomes important.

Therefore, we provide several robustness tests for different inequality measures in the forthcoming econometric analysis.

3.2. Changes in Regional Inequality: Convergence or Divergence?

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 is the situation in which poor regions grow faster than rich regions; i.e., the poor regions are catching up. Sigma-convergence involves cases in which the dispersion of income decreases over time. Although one might think that both approaches are equal, in reality they are not. Quah (1993, 1996) and Young et al. (2008) show that beta-convergence is a necessary but not a sufficient condition for sigma-convergence. If, for example, we observe beta-convergence in a particular period, it is possible that some formerly poor regions within a country will overtake previously rich ones, and the dispersion of income will increase rather than decrease.¹¹ Hotelling (1933) and Friedman (1992) note that a real test of a tendency to convergence should concentrate on measures of income dispersion. We intend to follow that suggestion in our paper. However, this approach comes at the cost of losing information about intra-distribution dynamics. Magrini (2004) notes that a constant inequality measure is consistent with very different dynamics within the income distribution, ranging from leapfrogging and crisscrossing to persistent inequality and poverty traps.

The most influential empirical studies on convergence are likely those of Barro and Sala-i-Martin (1991, 1992). Gennaioli et al. (2014), who provide us with regional income data, also analyze beta-convergence in a wider dataset. Those studies find evidence of beta-convergence between regions, with a typical speed of convergence between 1 and 2% per year. However, studies that concentrate on sigma-convergence, i.e., the dispersion of income within countries, arrive at contradictory 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. Originally, Kuznets (1955) stated 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. The major idea is that the development process in very poor economies usually starts in one part of the

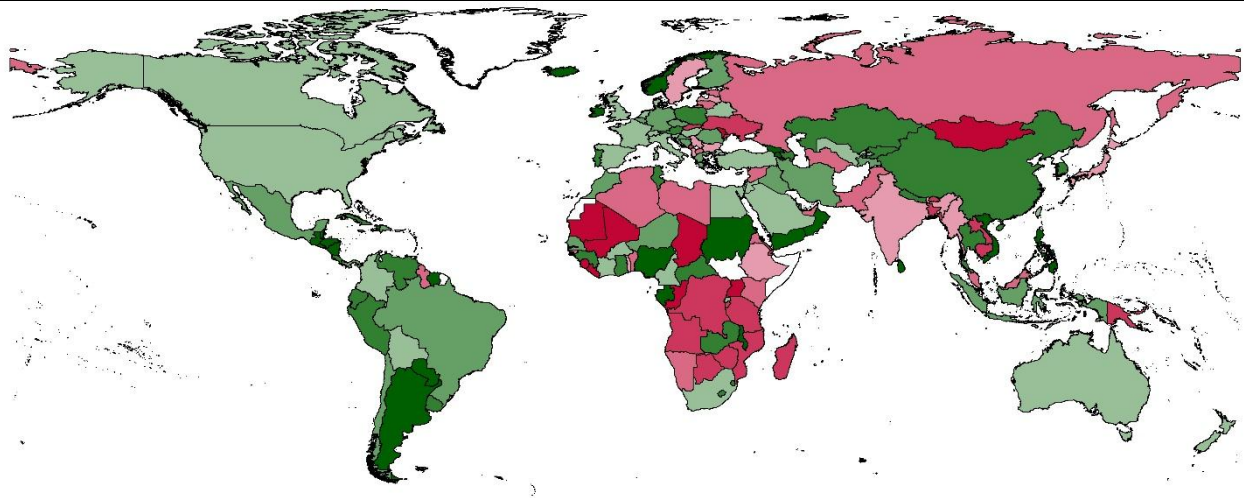
¹¹ Brezis et al. (1993) call these turnovers “leapfrogging”.

country; therefore, these regions become richer while the rest 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, e.g., either they adopt new technology or factors are mobile and induce a convergence process. Barrios and Strobl (2009) provide an analytical framework for this theory.

Using our new dataset on regional inequality, we can investigate the changes of regional inequality over time; i.e., we can analyze sigma-convergence for all countries in the world. Note that we use the phrases “sigma-convergence” and “changes in regional inequality” synonymously. In this descriptive analysis, we concentrate on the changes between 1992-2001 and 2002-2012. We discuss only the results that we obtain from observing the weighted $GINIW_j$ coefficient. Pooling all countries together, we find that 119 of 177 countries (67.2%) experience sigma-convergence. If we consider only those countries for which the different inequality measures point in the same direction, we then find that 113 of 161 countries (70.2%) experience convergence. The descriptive results for our inequality measures suggest that the majority of countries experience sigma-convergence. However, at least one-quarter of countries experience divergence. These findings may be related to a non-linear effect between inequality and development. Of course, there might also be individual reasons for these findings, e.g., asymmetric macroeconomic shocks that affect different regions of countries differently. In section 4, we will therefore examine the determinants of sigma-convergence using panel data. Figure 3 illustrates the changes in regional inequality, where countries displayed in green experienced convergence and countries displayed in red experienced an increase in inequality. The darker the color, the larger the change.

Figure 3: Changes of regional inequality within countries [GINIW (mean 1992-2001)—GINIW (mean 2002-2012)]





The first observation is that regional inequalities tend to decrease in most high-income countries. The decline is relatively small in absolute terms (light-green) because most of these countries have already comparatively low levels of inequality. Turning to middle-income countries, we also observe a convergence trend. However, there is some heterogeneity within this income group. Take, for example, an upper-middle-income country such as Brazil. Here we observe a convergence process, which is also well documented in the literature, which uses various sources of overserved income data (see, e.g., Silveira-Neto and Azzoni 2006). In contrast, a lower-middle-income country such as India experiences divergence, which is also in line with existing studies (see, e.g., Bhattacharya & Sakthivel 2004 and Ghosh 2008). Turning our attention to low income countries, the results are even more interesting because this is the category for which no comparable cross-country studies exist. Here, we observe the strongest pattern of increasing regional inequality (e.g., Uganda). Additionally, this observation supports the inverted-U hypothesis.

Finally, there seems to be a relationship between regional inequality and internal conflict. Our data indicates that the Ukraine, Libya, and Syria are experiencing increasing inequality. In all three countries, regional inequalities have been on the rise in the period prior to the civil wars. In addition, poor and conflict-torn countries such as Sierra Leone show a clear divergence process. Note that, these observations do not imply a causal relationship. We return to this issue in our concluding remarks.

4. Determinants of Sigma Convergence

The previous section has shown that the majority of countries converge, whereas regions in a significant number of countries diverge. In this section, we take another step by investigating the determinants of regional inequality. To do so, we use panel fixed-

effects regressions in which we build 5-year period averages. The panel regressions focus on the within-country variation over time, allowing us to derive conclusions about the effect of the potential determinants on the changes of regional inequality or the sigma-convergence, respectively. For studies that focus on differences in regional income and inequality between countries, see Henderson et al. (2016) and Alesina et al. (2016). Their frameworks facilitate a deeper analysis of geographic, cultural, historic and other time-invariant effects on income dispersion, which is beyond the focus of our analysis.

4.1. Discussion of Potential Determinants

The literature discusses various determinants of regional inequality. We cluster the different variables into seven groups: Development, Mobility, Openness, Resources, Institutions, Transfers and Education, and Ethnicity. Note that geography is also very likely to affect regional income and income inequality. However, we concentrate on within-country changes over time and therefore, time-invariant geographic factors vanish in the country-fixed effects. The starting point for our analysis is the empirical model used in Lessmann (2014), who not only 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 one topic at a time. Note that a clear separation of issues is not always possible. For example, natural resources might simultaneously affect policies, institutions, and trade. We emphasize that our empirical analysis focuses exclusively on correlations. Of course, some determinants—e.g., the national development level—might also be an outcome of regional inequalities. Our aim in this section is to provide general insights into the empirical relationship between the variables of interest, not to draw causal conclusions. The latter would require instrumental variable regressions or other methods suitable for identification.

Development

As discussed in section 3.2, Williamson (1965) adopts the Kuznets hypothesis in the case of regional development. There is empirical evidence for the inverted U in regional inequality, e.g., Williamson (1965), Barrios and Strobl (2009) and Lessmann (2014). Moreover, Lessmann (2014) finds evidence in a sample of 56 countries for the proposition that regional inequality increases again at very high levels of development. 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 existing studies have a serious bias toward middle- and high-income countries; therefore, the results cannot be generalized. This is particularly an issue if we aim to investigate the upward sloping part of the Kuznets curve at very low income levels. We therefore re-examine this issue based on countries' entire populations. The

level of economic development is measured by the log of the real GDP p.c. in constant U.S. \$ in power-purchasing parities (PPP). We consider different polynomial functions for the relationship between inequality and development.

Resources

Natural resources are a major determinant of regional incomes. Some resources, such as forests or arable land, are often uniformly distributed within countries, whereas 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 than the rest of the country. Consequently, regional inequalities are expected to increase, at least if the initial income distribution was balanced. Henderson et al. (2012) show how the accidental discovery of gemstones in Southern Madagascar near the town of Ilakaka at the end of the 1990s could be traced using the nighttime light data. Since that time, light emissions have significantly increased because of a rapid increase in regional population and income. In this example, 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 decrease, at least if the poor regions do not overtake the richer ones. Note that the net effect of resources on inequality is ambiguous if one considers a potential resource curse at the regional level (see Sachs and Warner 1999). We abstract from these issues by considering only a linear and unconditional effect of resources on inequality; although variable interactions are likely to be important, they exceed the scope of our analysis. The variables that we consider are resource rents as a share of the GDP as reported by the World Development indicators.

Another important resource variable is related to the agricultural suitability of land. Henderson et al. (2016) provide a detailed analysis of terrain ruggedness, soil quality, etc. Because we concentrate on time varying measures only, we use the share of arable land within countries as a further determinant of regional inequality. On average, the share of arable land worldwide has increased from 9.7% of the land mass in 1961 to 10.9% today. There are significant differences between countries. In Argentina, for example, arable land increased from 6.8% to 14.5%, whereas arable land decreased in Chile from 4.9% to 1.8%. The changes are particularly substantial in Africa, where Sierra

Leone increased its arable land from 4.9% to 20.9%, whereas Egypt's amount of arable land stagnated at approximately 2-3%. If we assume that increases in arable land are more likely to take place in a country's poorer, sparsely populated hinterland, we would expect an increase in arable land to decrease regional inequality.

Openness

An economy's openness of an economy can also affect regional inequality within countries [see Brühlhart (2011) for an overview]. If international trade costs decrease, then regions of a country that have relatively better market access gain more from international trade regions that have only 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 datasets. Some of those studies consider non-linearities and/or interaction variables. We investigate the effect of trade in our model using the trade/GDP ratio, in which trade is measured by the sum of exports and imports.

Another potential determinant of inequality, which is also related to the openness of the economy, is that of 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 both output and consumption (assuming complementarity between capital and labor). Consequently, regional inequalities rise as a response to the influx of FDI. The output-promoting effect will occur not only with greenfield investment (because the physical capital stock directly increases) but also in the case of mergers and acquisitions (which usually involve a transfer of knowledge (intangible capital)). 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 issue. China opened its economy to investors in the early 1990s and has since experienced 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. We consider this issue in our regressions by including net FDI inflows as a share of GDP as a potential determinant of regional inequality.

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. Empirical studies on this issue must address the problem that an economy's structure is already given at one point in time and is heavily dependent on both historic patterns (Henderson et al. 2016) and geography (Ramcharan 2009). Studying the spill-over effects of exogenous asymmetric shocks across Sub-Saharan cities, Storeygard (2016) shows that transport costs influence income differentials even after agglomeration has taken place. Positive economic shocks in one city have a stronger influence on a neighboring city when transportation costs are lower. Thus, reduced transportation costs have the potential to reduce regional inequalities because economic shocks tend to influence regions more equally. Because transportation costs are a function of costs per kilometer and distance, we follow the approach of Storeygard (2016) and consider national gasoline prices (cost per distance) and a country's surface in square kilometers (distance). The interaction of the two variables then gives us a measure for transportation costs, which vary in both time and space.

Institutions

Institutions are also likely to affect regional inequalities within countries. Ades and Glaser (1995) model the 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 dataset 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 also consider the index of bureaucratic quality as reported by the PRS Group's International Country Risk Guide.¹² There is extensive literature on the relationship between institutions and development. Acemoglu et al. (2001, 2002) and Dell (2011) show 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 colonization. Today's regional inequality might still be affected by these fundamentals. In addition to these historical factors, one might consider that countries with high institutional quality allow for more efficient resource allocation across regions, which should ultimately decrease regional inequality. Ezcurra and Rodríguez-Pose (2014) find a negative and significant association between government quality and the magnitude of regional disparities, thus supporting this line of reasoning. In contrast, 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) located in different countries after decolonization. This result is also supported by Gennaioli et al. (2013), who find institutions not to be a significant determinant of differences of regional development within countries.

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 policies that attract mobile capital to poor regions. 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 taxation or regulation policy can harm the development of peripheral regions in the presence of agglomeration rents. However, the potential benefits of decentralization can be offset by a ruinous tax competition game, as noted by Prud'homme (1995). The empirical literature is somewhat inconclusive. Whereas early cross-country studies find that decentralization decreases regional inequality (see, e.g., Shankar and Shah 2003 and Lessmann 2009), more recent studies find that this effect depends on the level of development (see Ezcurra and Rodríguez-Pose 2010 and Lessmann 2012). In our analysis, we consider the interaction between the level of development and a federal dummy variable, following the approach of Ezcurra and Rodríguez-Pose (2010) and Lessmann (2012). We admit that a federal dummy is a relatively crude measure of decentralization; however, the data are available

¹² Note that we also used alternative measures for institutional quality such as control of corruption, rule of law, etc. Only in case of the corruption index, we received a significant effect in some of the specifications, which is not robust to the particular inequality measure.

for a large set of countries. Measures of fiscal decentralization based on government finance data would shift the focus of the analysis to upper-middle- and high-income countries, which is contrary to our major aim.

Transfers and Education

There is an ongoing debate in the development economics literature on the effectiveness of development aid programs (Burnside and Dollar 2000, Easterly et al. 2003, Lessmann and Markwardt 2012, 2016). Although the literature usually focuses on aid's effect on national economic growth, aid might also have an important effect on the regional distribution of economic activity within recipient countries. One concern raised in this literature is the inefficient allocation of aid towards specific regions. For example, there is country-level evidence that NGOs tend to cluster, following other NGOs and avoiding difficult institutional environments (Koch, et al. 2009). There is reason to believe that this might also be the case within countries, if, for example development agencies concentrate on capital cities or particular ethnic groups. Therefore, we can expect an increase in inequality with increasing aid. We consider the aid-to-GDP ratio as further determinant of regional inequality.

The major result of Gennaioli et al. (2013) is that human capital is the most important determinant of differences in regional development within countries. Human capital creates positive externalities, which can also spill over regional boundaries. Those authors' empirical framework allows for the use of regional information about differences in human capital within countries. Unfortunately, these data are not available for our extended country sample. We therefore fall back upon the average quality of human capital within countries, as measured by the secondary-school enrollment rate. We expect human capital to decrease regional inequality because it facilitates regional spillovers and convergence, thus promoting internal migration.

Ethnicity

The final determinant of regional inequality that we consider in our analysis is the degree of ethnic heterogeneity. 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 a country's different ethnic groups often live in separate territories. If a large ethnic group dominates a minority, then we should expect this domination to yield income differentials. In line with this expectation, Alesina et al. (2016) find that ethnic income inequality is highly correlated with regional inequality in a cross-country framework. Although we follow their approach, we focus on the within-country variation in ethnic and regional inequality, respectively. This is important because both variables are heavily dependent on geographic issues (Michalopoulos 2012), which are difficult to control in a cross-country

framework. Using country-fixed effects, our panel regressions take the unobserved heterogeneity between countries into account and therefore, we expect to gain additional insights. For this purpose, we collect data on average lights in ethnic homelands based on the boundaries of ethnic groups of the classical Soviet Atlas Narodov Mira (see Weidmann et. al 2010 for further details). We then calculate different weighted and unweighted ethnic income inequality measures, which we consider as further determinants of regional inequality.

4.2. Methodology

Below, we regress our measures of regional inequality within countries on the different potential determinants. All variables are collected at the country level, except for our time-varying measure of ethnic income inequality, which is based on geo-coded data. We use panel-fixed effects regressions; therefore, our study focuses on changes within countries over time. In other words, we analyze the determinants of sigma-convergence. This 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 at the prediction stage and the population weights in the inequality measures may capture these heterogeneities only imperfectly. Moreover, there may be an omitted variable bias in the estimates if there are unobserved factors that simultaneously affect both inequality and the explanatory variables. 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.¹³

The basic empirical model in this section has the form

$$[1] \quad I_{j,t} = \alpha_j + \beta \cdot F(\ln(y_{j,t})) + \gamma_k \cdot \sum_{k=1}^m X_{k,j,t} + \theta_t + \varepsilon_{j,t}, \quad [5]$$

where α_j is country-fixed effects, and θ_t is time-fixed effects. We take different polynomial functions for the relationship between regional inequality $I_{j,t}$ and development level $\ln(y_{j,t})$ into account to investigate a Kuznets curve in regional inequality. $X_{k,j,t}$ represents the other specific determinants of regional inequality; $\varepsilon_{j,t}$ is a

¹³ 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 various territorial levels do not affect significantly similar panel regressions.

random error term. Because errors may be correlated within countries, we always report robust standard errors that are clustered at the country level. Our final dataset considers 180 countries and five periods of 5-year averages, yielding a theoretical limit of 900 observations.¹⁴ Depending on the selection of variables, our panel is unbalanced, with a maximum of 880 observations.

Note that we concentrate on 5-year period averages to eliminate the effects of the business cycle. Over the business cycle, income and inequality are likely to be correlated. If a country's various regions are heterogeneous with respect to the structure of the economy, then they will react differently during 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 than 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 B.3 in Appendix B for an illustration of this issue.

4.3. Results

4.3.1 The Kuznets Curve

We begin our analysis with a closer look at the relationship between regional inequality and development measured by the log of the GDP per capita in PPP. The results of different specifications of equation [5] are reported in table 3. Although the main dependent variable is the weighted Gini coefficient (columns 1-3), we also report the results obtained using other inequality measures (columns 4-7). In contrast to Alesina et al. (2016), we do not find a significant linear relationship between spatial inequality and development in the panel fixed-effects regression. The coefficient is negative, but insignificant (column 1). If we add the quadratic term, both income and squared income are significant at conventional confidence levels, and the sign of the coefficients suggest an inverted U-shaped relationship (column 2). If we consider a cubic function, we find that regional inequalities increase again at very high levels of development (column 3). This implies that the relationship between inequality and development looks N-shaped instead of U-shaped. However, the results do not necessarily imply that many high-income countries are already on the upward sloping part of the Kuznets curve. A cubic function is also consistent with stabilizing inequality levels in rich economies.

¹⁴ The periods are 1990-1994, 1995-1999, 2000-2004, 2005-2009, and 2010-2013.

Our main finding concerning the relationship between development and regional inequality is robust to the particular inequality measure. The results reported in columns (4)-(7) show that the sign and significance of coefficients remain unchanged if we consider inequality measures that are more sensitive to changes in the lower tail of the income distribution. If we calculate the marginal effects of income on inequality, we find that the amplitude of the Kuznets curve is smaller the more sensitive the inequality measure is to changes in the lower tail of the distribution (columns 5-7). Thus, national development seems to influence richer regions more than poorer ones. This implies that the increases in inequality observed at early stages of development originate more from the growth of already-rich regions.

Table 3: Determinants of sigma convergence, panel fixed effects regressions, 5-year averages

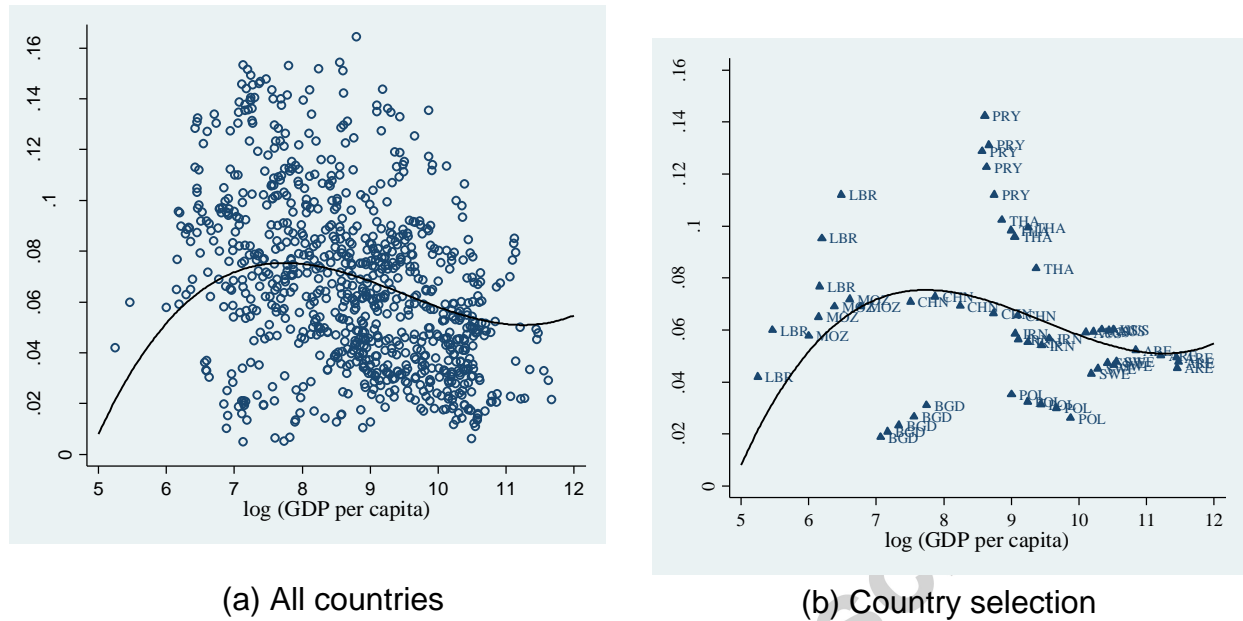
VARIABLES	Dependent variable:						
		<i>GINIW_j</i>		<i>COVW_j</i>	<i>THEILW_j</i>	<i>MLDW_j</i>	<i>GEW(-1)_j</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(GDP p.c.)	-0.003 (0.003)	0.056** (0.023)	0.293** * (0.078)	0.402** * (0.144)	0.066*** (0.023)	0.063** * (0.020)	0.061*** (0.019)
log(GDP p.c.) ²		- 0.003** * (0.001)	- 0.032** * (0.009)	- 0.044** * (0.016)	- 0.008*** (0.003)	- 0.007** * (0.002)	-0.007*** (0.002)
log(GDP p.c.) ³			0.001** * (0.000)	0.002** (0.001)	0.000*** (0.000)	0.000** * (0.000)	0.000*** (0.000)
Constant	0.095** * (0.025)	-0.150 (0.099)	0.799** * (0.225)	- 1.013** (0.418)	- 0.171*** (0.066)	- 0.166** * (0.059)	-0.162*** (0.055)
# Observations	880	880	880	880	880	880	880
# Countries	180	180	180	180	180	180	180
R-squared-w	0.126	0.201	0.244	0.264	0.210	0.205	0.196
R-squared-b	0.166	0.105	0.119	0.182	0.174	0.162	0.148
Country FE	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES

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

We illustrate our findings in figure 4. The left panel (5a) shows a partial scatterplot that includes the regression function. For this purpose, we partial out the effects of the control variables (fixed effects) from the measure of regional inequality and then compare regional inequality to 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 2,288 US\$ as the local maximum and 77,128 US\$ as the local minimum. Very poor countries experience increasing inequalities in our dataset; countries such as Liberia and Mozambique belong to that group. Beyond this threshold, regional inequalities fall to a threshold of 77,128 US\$, where inequalities start to increase moderately.

Importantly, the clear majority of countries are in the downward sloping part of the N-shaped function, which is likely to be the effect found in Alesina et al. (2016). Our results are in line with the descriptive analysis on sigma-convergence provided in section 3. Most countries converge, but for the poorest and the richest countries, we observe divergence across regions.

Figure 4: Regional inequality and development

Adjusted spatial inequality versus log observed GDP per capita at the national level. 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 (3) Table 3. The plot represents the N-shaped relationship between economic development and regional inequality.

In the right panel (5b), we selected single countries because the scatterplot with all observations makes it impossible to investigate single cases. Given that the income in this subsample of countries has grown over time, we could also interpret the abscissa as the time variable. In poor countries such as Liberia, Mozambique and Bangladesh, inequalities increased during our observation period; middle-income countries such as China, Paraguay, and Thailand experience decreasing inequalities; high-income countries such as the U.S. and Germany, experience small decreases of inequality at low levels, whereas rich countries such as Sweden experience (moderately) increasing inequalities. Please note that only few countries are beyond the upper income threshold, therefore the upward sloping part of the Kuznets curve in very rich countries should be interpreted cautiously.

4.3.2 Specific Determinants of Sigma-Convergence

Next, we sequentially add the different potential determinants of regional inequality to our regression model using the classification of factors discussed above. The results

using the *GINIW* as the dependent variable are reported in table 4. Tables B.7-B.10 in the Appendix report the results using the other inequality measures, upon which we comment separately.

In column (1), we investigate the resource variables—i.e., resource rents and arable land. Both variables show the expected signs. Whereas resource rents increase regional inequality within countries, lower pressure from land decreases inequality. In column (2), we add the openness variables trade/GDP and FDI/GDP. The trade variable is positive and significant in support of the results of Barrios and Stobl (2009) and Rodriguez-Pose (2012), among others. The coefficient of the FDI variable is not significant at conventional confidence levels. However, if we interact FDI with income, we find—in line with Lessmann (2013)—different effects across high- and low-income economies. Whereas FDI has no effect on regional inequality in high-income countries, FDI increases regional inequalities in low-income countries.¹⁵ This may be related to the lower factor mobility in low-income countries, which inhibits rapid adjustment to FDI inflows. In column (3), we concentrate directly on factor mobility, using gasoline prices and country size as proxies for transportation costs. The interaction variable is statistically significant, implying that increasing transport costs increase regional inequality in large countries. This result is in line with Storeygard (2016), who investigates the effect of transport costs on the income differential between African coastal cities with ports and the hinterland. In column (4), we consider the institutional variables, finding no significant effect for bureaucratic quality and the Polity2 index, but a significant effect for an interaction variable of income and a dummy for countries with a federal constitution. In line with Ezcurra and Rodriguez-Pose (2010) and Lessmann (2012), our results indicate that rich federations tend to have lower regional inequalities. In column (5), we focus on transfers and education. Transfers in the form of development aid tend to increase the concentration of economic activity within countries. This result can be interpreted in two ways. First, aid increases development and therefore, aid-receiving countries move up the upward slope of the Kuznets curve. However, we control for the development level in a very flexible way and therefore, this interpretation does not seem very convincing. Second, aid concentrates on single regions of a country, thereby increasing regional inequality. This can hamper the general effectiveness of aid in poor regions, which may also be inhabited by a different ethnicity, engaged in rent-seeking activities or experiencing internal resource conflicts. Moreover, we find a significant negative coefficient for the education variable. This result is very much in line with Gennaioli et al. (2013), who find human capital to be the most

¹⁵ The results are available from the authors upon request.

important determinant of regional growth. Human capital is decisive for regional knowledge spillovers; therefore, poor regions receive the opportunity to catch up with richer ones. Finally, we add the ethnic income inequality index in column (6). We find a positive relationship between regional inequality and ethnic inequality. The result is in line with Alesina et al. (2016). Importantly, we can show this

Accepted manuscript

Table 4: Determinants of sigma convergence, panel fixed effects regressions, 5-year averages

	Dependent variable: weighted coefficient of variation (WGINI)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(GDP p.c.)	0.350*** (0.075)	0.205** (0.081)	0.171** (0.078)	0.321*** (0.046)	0.226** (0.110)	0.149** (0.059)
log(GDP p.c.) ²	-0.038*** (0.009)	-0.022** (0.009)	-0.019** (0.009)	-0.035*** (0.006)	-0.023* (0.014)	-0.015** (0.007)
log(GDP p.c.) ³	0.001*** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001*** (0.000)	0.001 (0.001)	0.000* (0.000)
Resources rents / GDP	0.018*** (0.007)					
Arable land (%)	-0.053*** (0.014)					
Trade/GDP		0.005*** (0.002)				
FDI/GDP		0.009 (0.007)				
Price gasoline			0.001 (0.002)			
Price gasoline X Area			0.006** (0.003)			
Bureaucratic quality				-0.001 (0.001)		

Polity2				-0.002		
				(0.002)		
log(GDP p.c.) X Federal				-0.331***		
				(0.119)		
Aid/GDP					0.015**	
					(0.007)	
School enrolment secondary					-0.014*	
					(0.007)	
Ethnic inequality						0.071***
						(0.016)
Constant	-0.962***	-0.543**	-0.438*	-0.725***	-0.621**	-0.419**
	(0.216)	(0.233)	(0.231)	(0.114)	(0.297)	(0.168)
# Observations	857	817	676	485	591	845
# Country	176	172	162	126	148	173
R-squared-w	0.287	0.258	0.166	0.329	0.251	0.363
R-squared-b	0.187	0.104	0.179	0.004	0.135	0.146
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. Due to space limitations, column 4 only reports the coefficients of interest. Note that all necessary variables are included in the regression such as the interaction of the federal dummy and log(GDP p.c.)² etc.

effect in a panel context in which geography is taken into account through the country-fixed effects. Therefore, we are confident that this result is not driven by any country-specific, time-invariant factors. The interpretation of the effect is straightforward. The inequality measure that we use here could be interpreted as a measure of intergroup inequality where the groups are formed by people's places of residence. Obviously, different ethnic groups live—at least to some extent—geographically separated from each other. One might think of the Dutch-speaking Flemings and the French-speaking Walloons in Belgium or the Krahn and the Gio in Liberia. In these cases, we observe larger regional inequalities, implying significant income differentials between different ethnic groups. For a more detailed analysis of the relationship between ethnic inequality and spatial inequality, see Alesina et al. (2016).

Finally, let us comment the results we obtain if we consider other inequality measures (table B.7—B.10 in the Appendix). The results on the specific determinants are in almost all cases robust to the underlying inequality measure. However, if we consider the Kuznets curve, we find that the N-shaped relationship is robust only for the weighted Gini coefficient. Using other inequality measures, the effect disappears depending on the particular other determinants of regional inequality we add to the regression. Considering the openness variables, transport costs, and transfers and schooling, we do not find a robust N-shaped relationship. The signs of the coefficients of the development indicators do not change, but we miss conventional confidence levels. This may have different reasons. First, the specific determinants are not available for all countries around the world. Our country sample changes from one regression to another, which is likely to affect our results, in particular if the poorest countries drop out of the sample. Of course, also multicollinearity may result in overestimated standard errors. Nevertheless, we have to admit that we cannot ensure that the Kuznets curve relationship is depending on the development level itself. It could just as well be the case that the Kuznets curve is driven by specific determinants such as the openness variables that are highly correlated with economic development. However, the Kuznets hypothesis is rather a positive description for the relationship between income and (regional) inequality. We cannot necessarily conclude that all countries will follow this path disregarding the other important factors that correlate with economic development.

5. Summary and Conclusions

In this paper, we analyze regional inequality within countries based on satellite nighttime light data. We use luminosity data to predict regional incomes in countries in which this variable is not observed. For this purpose, we refer to existing data on income and nighttime lights. In contrast to previous studies, we explicitly consider both top- and low-coding issues in the light data by including additional light-based variables in the empirical model. Taking the estimated relationship between these variables, we predict

regional incomes for all countries. This forecast yields a panel dataset of regional incomes worldwide covering 180 countries for 1992-2012. Our study's main contribution is its calculation of different measures of regional inequality within countries based on their predicted incomes. We investigate the level of regional inequality across countries and its changes over time, which corresponds to the concept of sigma convergence. Finally, we study the impact of the national development level on regional inequality and other potential determinants.

This study has seven main findings. (1) Light is a suitable indicator for predicting regional incomes. This result is in line with previous studies. However, adding light-based variables such as the number of top- and low-coded pixels within a region significantly improves the estimations. (2) Inequality measures based on predicted incomes are closely related to observed regional income inequality, whereas pure light-based inequality measures—as commonly used in the literature—show much weaker correlations with observed inequality. This result directly relates to the first result because we consider some potential measurement errors in the light data in our forecast model. In our preferred specification, the correlation between regional inequality based on predicted income and observed income is 0.50, whereas the correlation is only 0.21 if we compare inequality in light density with observed income inequality. (3) We identify differences in regional price levels as an important factor that explains parts of the remaining differences between inequality measures based on predicted income and observed income. Using the U.S. example, we show that the prediction errors of regional incomes depend on regional price parities. From this, we draw the conclusion that light-based predicted incomes better reflect real figures than nominal ones. (4) We find significant differences in regional inequality across countries, which are robust to different inequality measures. Very poor countries such as those in the Sahel zone show low regional inequality, middle-income countries such as those in Latin America have relatively high inequalities, and high-income countries have lower inequalities. (5) Investigating changes in inequality over time, we find that approximately 70% of all countries experience sigma convergence. However, some poor and lower-middle-income countries, along with some rich economies, show divergence. This points to an inverted-U shaped or even N-shaped relationship between regional inequality and development. (6) Using panel data regressions with country-fixed effects, we find evidence of an N-shaped relationship between inequality and development. However, the threshold value for the upward-sloping part at the right-hand-side of the income distribution is relatively high (~ 80.000 US\$) therefore, this result is based only on a few observations. A robust result is that very poor countries show increasing inequalities, whereas most middle- and high-income countries show decreasing or stabilizing inequalities. (7) Using several potential determinants of regional inequality as further covariates in our panel regressions, we arrive at the following results: natural resources,

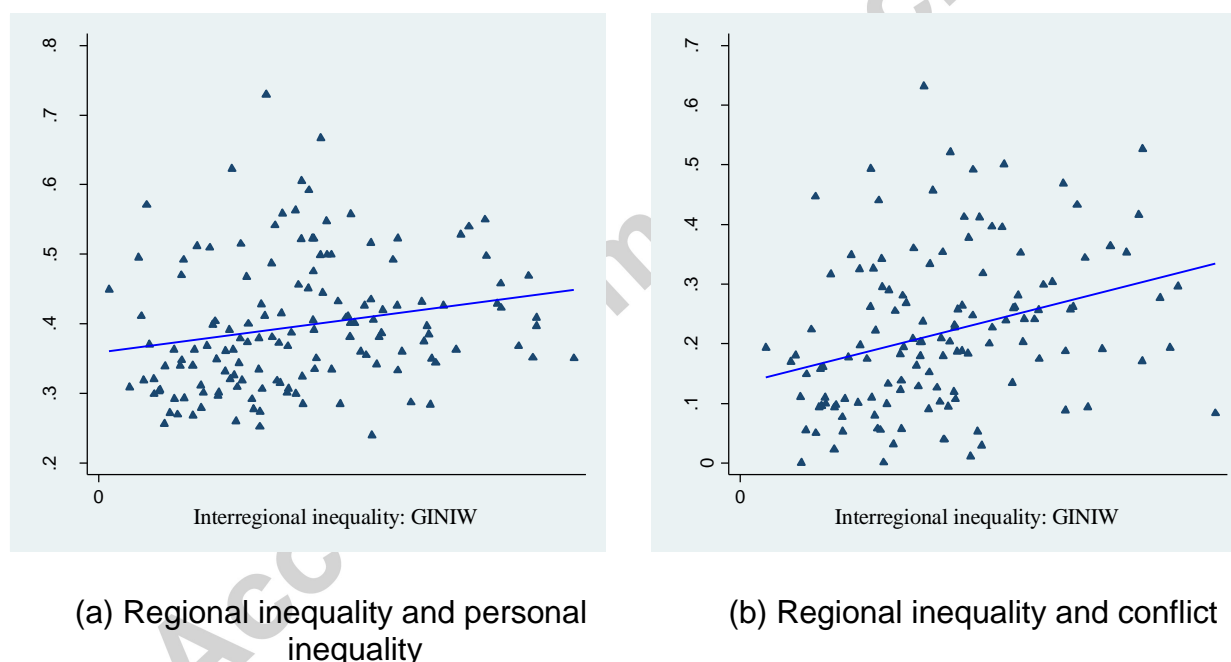
trade openness, transportation costs, foreign assistance, and ethnic income inequality are positively correlated with regional inequality, whereas the shares of arable land, federalism and human capital are negatively correlated with regional inequality.

Our analysis has certain limitations. First, light is only an imperfect proxy for regional income or even regional living standards. There remain unsolved measurement issues, which could bias our estimates. A crucial point for future research is to determine how to obtain more variation in the light data, particularly in low-coded (dark) and top-coded (maximum of 63 DN) regions. A recent study by Bluhm and Krause (2016) makes a first attempt to construct variations in luminosity values in top-coded regions based on a Pareto distribution just like the literature on personal income inequality does, where the income data of the richest part of the population is also not observed. Moreover, our results may depend on the underlying territorial classification. We concentrate on first-level administrative boundaries and use population weights to address regional heterogeneity. A different approach would be to use gridded data, which can yield different results. Finally, we emphasize that all 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 the causal effects between regional inequality and its potential determinants. In several cases, for example, situations involving trade openness and development aid, an endogeneity bias is likely to occur. In this initial study, however, we abstract from these important technical issues in favor of a broad first glimpse of the nexus between regional inequality and its potential determinants. 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 variables. We encourage future research on these determinants and are happy to provide our data.

Why should we care about regional inequality? Our study focuses on the measurement of regional inequality, convergence and its determinants. Finally, we aim to provide a brief outlook on the potential consequences of regional inequality. In their book *Spatial Inequality and Development*, Kanbur and Venables (2005, p. 3) state, “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”. This suggests a relationship between regional inequality and personal income inequality, ethnic inequality, and political stability. What do our data say about this issue? We have analyzed the relationship between regional and ethnic inequality in the panel framework. However, there are also interesting correlations with the other two variables. Figure 5, panel (a), illustrates a scatterplot for regional inequality (*GINIW*) and personal inequality measured by the Gini index; panel (b) illustrates a scatterplot for

regional inequality and the risk of internal conflict measured by the ICRG index.¹⁶ The illustrations suggest that both variables—personal inequality and internal conflict—are positively associated with regional inequality (see also Lessmann 2014, Deiwi et al. 2012, Buhaug et al. 2012 and Lessmann 2016). Therefore, researchers and decision makers should be aware of regional imbalances within countries, which might not merely be a byproduct of an undesired development; instead, they might be either an independent issue or at least a channel through which we can influence the main economic outcomes. There are numerous examples of high- and low-income countries in which unbalanced regional development yields income inequality, fueling ethnic tensions and the forces of segregation. Examples in the developed world include Catalonia (Spain) and Eastern Ukraine.

Figure 5: Regional inequality versus personal inequality and the risk of internal conflict (mean 2001-2012)



¹⁶ The International Country Risk Guide provided by the Political Risk Services Group ranks countries according to three dimensions: terror, civil war and civil unrest. We rescale the original index so that higher values indicate higher risk. The income Gini coefficients are taken from the World Bank project "All the Ginis" (see Milanovic 2014).

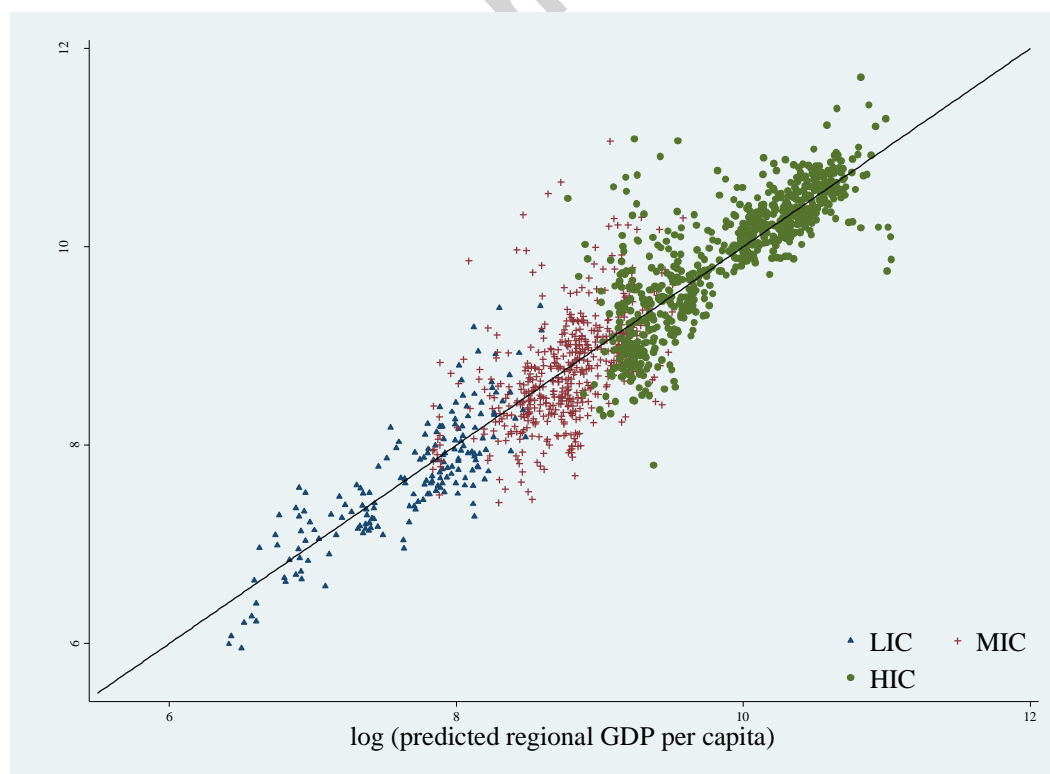
Appendix A. Robustness of Regional Income Prediction

In this section, we test the robustness of our regional income prediction, which is the basis for our inequality measures. We use two different approaches: first, we compare observed and predicted regional income and test whether we find any of the known biases that are associated with using light intensity as income proxy; second, we test whether national income levels calculated from the predicted regional incomes fit to observed national income.

Observed Regional Income vs Predicted Regional Income

Figure A.1 illustrates the results of our prediction. The abscissa reports the regional income predictions as calculated from equation [2]; 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.

Figure A.1: Predicted regional GDP per capita, measured regional GDP per capita (mean 2001—2010), and national income level



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 observed incomes. The Pearson correlation coefficient is 0.87, which deviates only slightly from the ideal value of 1. The quality of our estimates, however, seems to depend on the income level itself, because the deviation of observations from the bisecting line is particularly large in middle-income countries.¹⁷ 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 and low-income countries; however, there are several outliers in middle-income countries. Obviously, our model underestimates regional income in some cases; these are often exceptionally rich regions in otherwise poor countries. This is a common problem of light density as a proxy for income. It seems that we are not fully able to remedy this problem by adding further explanatory variables such as the number of top-coded pixels. However, at the end of this section, we will argue that at least in some cases this deviation originates in mistakes in the observed income data and not in the prediction.

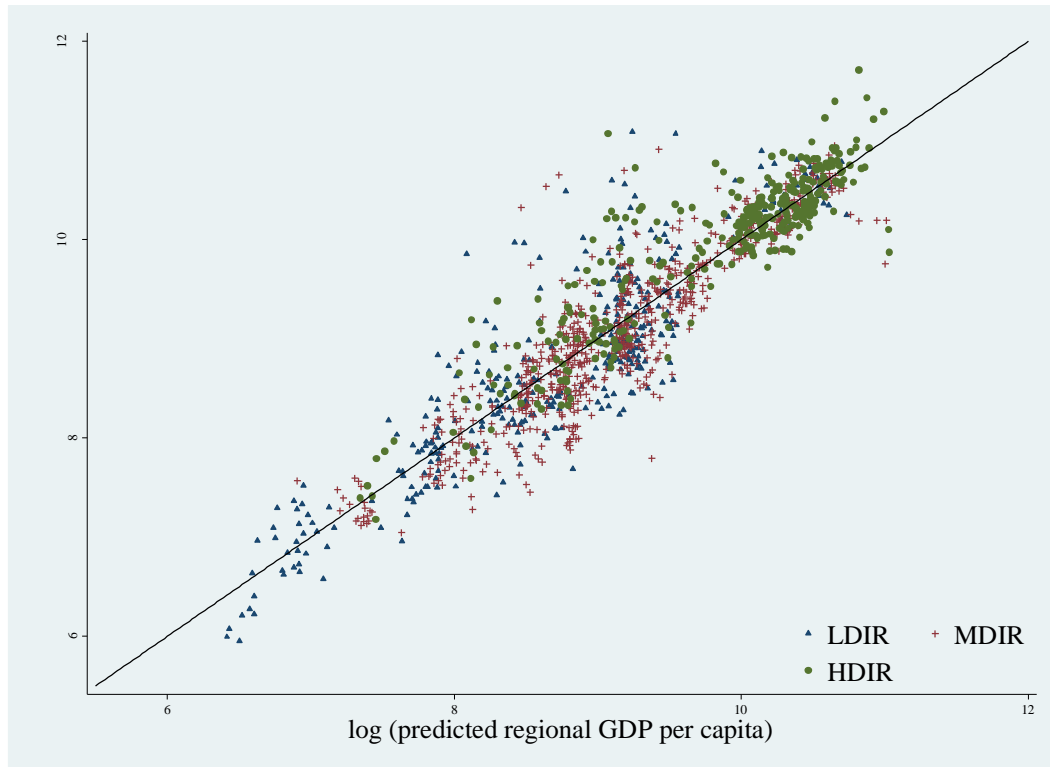
Output density bias

A further concern expressed in the literature is that light density alone is a poor income proxy in low output density regions [see for example Chen and Nordhaus (2011)]. As Chen and Nordhaus (2011) indirectly recommend, we therefore use further output indicators in our prediction besides light density to mitigate this problem. In order to check whether output-density is still a source of bias, we now distinguish low output density regions (*LIDR*), middle output density regions (*MIDR*) and high output density regions (*HIDR*). Since a low-coding and top-coding bias is most likely for extreme values, we choose the 25% percentile as definition for *LIDR*, and 75% for *HIDR*. Figure A.2 shows the same data as figure A.1. adding the information on output density of regions instead of income. Following the arguments mentioned above, we should observe larger prediction errors on particular for the low output-density regions (blue triangles). However, there is no systematic impact of output density on the difference between observed and predicted income – in particular for this type of regions. We also applied a simple test on this issue. For this purpose, we calculate the mean squared

¹⁷ The definition of low-, middle- and high-income countries follows the current World Bank definition for 2015, where LIC: $y_j < 4.086 \text{ US\$pc}$; MIC: $4.086 \text{ US\$pc} \leq y_j < 12.615 \text{ US\$pc}$; HIC: $y_j \geq 12.615 \text{ US\$pc}$

prediction error, which is a common measure to assess the overall performance of output prediction models [Lehmann and Wohlrabe (2015)]. The correlation between *LIDR* and the mean squared prediction error is just -0.05. This supports our hypothesis, that the inclusion of additional explanatory variables in our prediction model (in particular the number of low-coded pixels per region) helps to reduce this sort of bias.

Figure A.2: Predicted regional GDP per capita and measured regional GDP per capita (mean 2001—2010) and regional output density.



Real income vs. nominal income

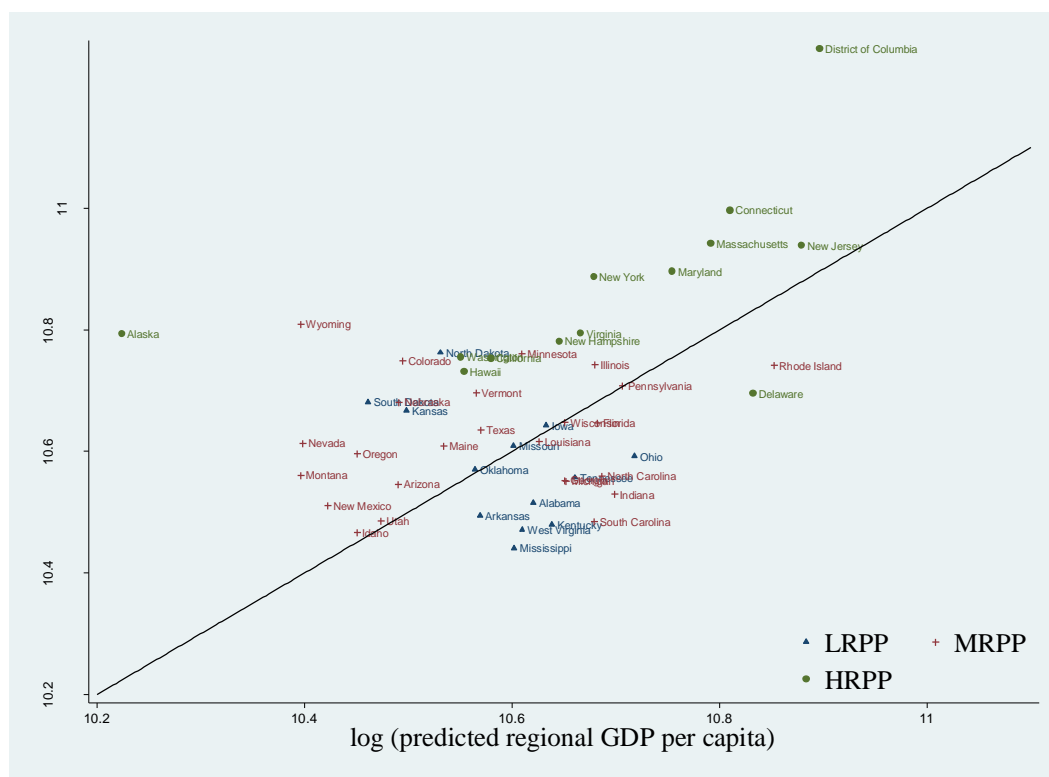
A so far little discussed issue in the evaluation of light density as income proxy is whether light is a proxy for real or nominal income. National statistics usually do not report regional prices. Hence, regional income measured in constant PPP dollars refers to national prices instead of regional prices. The regional income data provided by Gennaioli et al. (2014) is usually based on nominal regional income. Nominal regional GDP is converted into (current purchasing power) US\$ values by multiplying national GDP in PPP terms by the share of each region in national GDP, and then the authors use the regional population to compute per capita GDP in each region [Gennaioli et al. (2014)]. This, however, implicitly assumes that price levels are equal across regions. If this is not the case, the difference between predicted and observed income we observe in our

comparison may be a result of differences in nominal and real figures. The observed incomes are nominal, while light based income should reflect real figures considering differences in regional price levels within countries.

In the cases of the United States, the regional price parity is actually available for recent years. Therefore, we can compare the predicted and observed regional incomes in relation to differences in regional price levels. In figure A.3 the abscissa reports the regional income predictions as calculated from equation [2]; the ordinate shows reported (nominal) regional incomes for US states in 2010. We also add the information on the regional price parities distinguishing between the categories low regional price parities (*LRPP*), middle regional price parities (*MRPP*), and high regional price parities (*HRPP*).¹⁸ Interestingly, the differences between observed and predicted regional income depends on regional prices in a way one could expect. In regions with low prices (blue triangles), we overestimate the observed income, while it is just the other way round in regions with high prices (green dots), where we underestimate the observed income. In the middle category (red crosses), there is not systematic differences. This supports the hypothesis that light emissions are a better proxy for real income rather than nominal income. This is a serious advantage of this sort of data compared to the observed one, since the regional price level are reported only for a handful of countries around the globe. By using light data, one could solve a part of the measurement error in studies solely based on observable national account data. Moreover, this issue might explain a part of the difference between the inequality measures based on predicted income and regional income. The light-based data seems to be an even better proxy for real income differentials within countries. For a discussion on this point, see section 2.3.

¹⁸ Borders are defined by the upper bound of the 25% and 75% percentiles of the output density distribution. $LRPP \in [86.7, 90.2]$, $MRPP \in [90.2, 102.8]$ and $HRPP \in [102.8, 118.2]$.

Figure A.3: Predicted regional GDP per capita and measured regional GDP per capita USA 2010 and regional price parity.

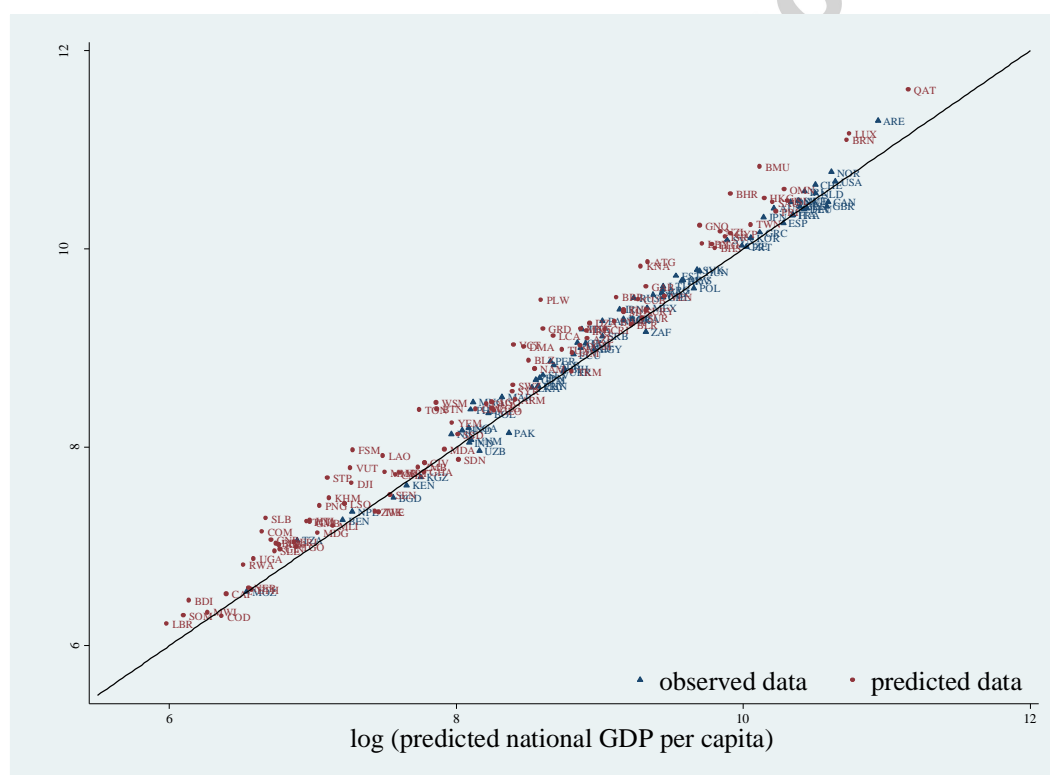


Out of Sample Prediction

The light—income relationship gives us the parameters for our prediction of regional incomes for all countries. 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 income data provided by Gennaioli et al. (2014) consider many countries at different levels of economic development. However, there is a selection bias in the data toward countries with higher income levels. This might bias our prediction if the light—income relationships within low-income countries is completely different. However, it is not possible to test this issue directly, as we do not have data regional incomes in very poor countries. Instead, we aggregate the constructed regional incomes to country-level incomes and compare these results with observed country-level incomes. Figure A.4 provides a scatterplot of predicted incomes at the country level (abscissa) and measured incomes (ordinate). We distinguish between those countries where we have observed regional income data (triangles) and those countries that are out of sample (dots). 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 from the observed data. This is particularly true for maritime states, such as Tuvalu, Palau, and the Bahamas. Here, we significantly underestimate incomes. This may be related to the night light data. Those pixels, which cover a coastal area, have a smaller landmass within a pixel, since also the water surface is considered in the calculation of light densities. This effect biases our light measure downwards, and, consequently, our predicted incomes. However, there is also a bias in the other direction. To compensate for glaring, the NOAA sets light measured over water to zero. This might be a strong data correction given that glaring also occurs on land, where it is not corrected. Therefore, we face different measurement errors in small and densely populated islands.

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



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., S. Michalopoulos and E. Papaioannou (2016): Ethnic Inequality, *Journal of Political Economy*, 124(2), 428-288.
- Allison (1978): Measures of Inequality, *American Sociological Review*, 43(6), 865-880.
- 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.
- Atkinson, A. B. (1970): On the Measurement of Inequality, *Journal of Economic Theory*, 2(3), 244-263.
- 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.
- Barro, R.J. and X. Sala-i-Martin (1995): *Economic Growth*. Boston, MA: McGraw-Hill.
- Besley, T. and M. Reynal-Querol (2014): The Legacy of Historical Conflict: Evidence from Africa, *American Political Science Review*, 108(2), 319-336.
- Bhattacharya, B. B., & Sakthivel, S. (2004): Regional Growth and Disparity in India: Comparison of pre-and post-Reform Decades, *Economic and Political Weekly*, 39(10), 1071-1077.
- Bluhm, R. and M. Krause (2016): Top Lights – Bright Spots and their Contribution to Economic Development, mimeo, University of Hannover.

-
- 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.
- Burnside, C., & Dollar, D. (2000): Aid, Policies, and Growth, *American Economic Review*, 90(4), 847-868.
- 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.
- Conley, T. G. (1999): GMM Estimation with Cross Sectional Dependence, *Journal of Econometrics*, 92(1), 1-45.
- Cowell, F. A. (2000): Measurement of Inequality, *Handbook of Income Distribution*, 1, 87-166.
- Deiwijs, C. L. Cederman, and K. Gleditsch (2012): Inequality and Conflict in Federations, *Journal of Peace Research*, 49(2), 289-304.
- Dell, M. (2010): The Persistent Effects of Peru's Mining Mita, *Econometrica*, 78(6), 1863-1903.
- Easterly, W. and R. Levine (1997): Africa's Growth Tragedy: Policies and Ethnic Divisions, *Quarterly Journal of Economics*, 112(4), 1203-1250.
- Easterly, W. (2003): Can Foreign Aid Buy Growth? *Journal of Economic Perspectives*, 17(3), 23–48.
- Ebener, S., C. Murray, A. Tandon, and C.D. 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 (2010): Does Decentralization Matter for Regional Disparities? A Cross-Country Analysis, *Journal of Economic Geography*, 10(5), 619-644.
- Ezcurra, R. and A. Rodríguez-Pose (2013): Does Economic Globalization Affect Regional Inequality? A Cross-Country Analysis, *World Development*, 52(C), 92-103.
- Ezcurra, R., and. A. Rodríguez-Pose (2014): Government Quality and Spatial Inequality: A Cross-country Analysis. *Environment and Planning A*, 46(7), 1732-1753.
- 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, M. (2008): Economic Reforms, Growth and Regional Divergence in India, *Journal of Applied Economic Research*, 2(3), 265-285.
- 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, *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.
- Henderson, J. V., Squires, T. L., Storeygard, A., & Weil, D. N. (2016): *The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade* (No. w22145), National Bureau of Economic Research.
- 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 (2005): Spatial Inequality and Development, in: Kanbur, R. and A. Venables (Eds.), *Spatial Inequality and Development*, 3-11. New York: Oxford University Press.
- Keola, S., M. Andersson and O. Hall (2015): Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth. *World Development*, 66, 322-334.
- Koch, D. J., Dreher, A., Nunnenkamp, P., & Thiele, R. (2009): Keeping a Low Profile: What Determines the Allocation of Aid by Non-governmental Organizations? *World Development*, 37(5), 902-918.
- Krugman, P (1991): Increasing Returns and Economic Geography, *Journal of Political Economy*, 99(3), 483-499.
- Kuznets, S. (1955): Economic Growth and Income Inequality, *American Economic Review*, 45(1), 1-28.
- Lehmann, Robert, and K. Wohlrabe (2015): Forecasting GDP at the Regional Level with Many Predictors, *German Economic Review*, 16(2), 226-254.
- Lessmann, C. (2009): Fiscal Decentralization and Regional Disparity: Evidence from Cross-Section and Panel Data, *Environment and Planning A*, 41, 2455–2473.
- Lessmann C. (2012): Regional Inequality and Decentralization: An Empirical Analysis, *Environment and Planning A*, 44, 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. (2016): Regional Inequality and Internal Conflict, *German Economic Review*, 17(2), 157-191.
- Lessmann, C., and Markwardt, G. (2012): Aid, Growth and Devolution. *World Development*, 40(9), 1723-1749.

- Lessmann, C., and Markwardt, G. (2016): Aid, Growth, and Devolution: Considering Aid Modality and Different Types of Decentralization, *World Development*, 84, 188-130.
- Magrini, S. (2004): Regional (Di) Convergence, *Handbook of regional and urban economics*, 4, 2741-2796.
- 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. (2012): The Origins of Ethnolinguistic Diversity, *American Economic Review* 102(4), 1508-1539.
- 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.
- Prud'h, omme, 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.
- Ramcharan, R. (2009): Why an Economic Core: Domestic Transport Costs, *Journal of Economic Geography*, 9(4), 559-581.
- Rodríguez-Pose, A. (2012): Trade and Regional Inequality, *Economic Geography*, 88(2), 109-136.
- Rodriguez-Pose, A. and N., Gill. (2006): How Does Trade Affect Regional Disparities? *World Development*, 34(7), 1201-1222.
- Sachs, J. D. and A. M. Warner (1999): The Big Push, Natural Resource Booms and Growth, *Journal of Development Economics*, 59(1), 43-76.

-
- Storeygard, Adam (2016): Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa, *Review of Economic Studies*, 83, 1263-1295.
- 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.
- Silveira-Neto, R. and Azzoni, C. R. (2006): Location and Regional Income Disparity Dynamics: The Brazilian Case. *Papers in Regional Science*, 85(4), 599-613.
- 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.
- Weidmann, N. B., Roslashd, J. K., & Cederman, L. E. (2010): Representing ethnic groups in space: A new dataset, *Journal of Peace Research*, 47(4), 491-499.
- 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.