Kyoto University,
Graduate School of Economics
Discussion Paper Series

Revisiting National Institutions and Subnational Development in Africa with New Nighttime Light Data

Soyoka Okamura, Yotaro Ueno, Toma Yamagoshi, Hisaki Kono

Discussion Paper No. E-23-008

Graduate School of Economics
Kyoto University
Yoshida-Hommachi, Sakyo-ku
Kyoto City, 606-8501, Japan

March, 2024

Revisiting National Institutions and Subnational Development in Africa with New Nighttime Light Data

Soyoka Okamura* Yotaro Ueno† Toma Yamagoshi‡ Hisaki Kono§

January 31, 2024

Abstract

We revisit the empirical investigation of the importance of national institutions for sub-regional economic development using more accurate nighttime light data. In contrast to the original study by Michalopoulos and Papaioannou (2014), we find that national institutions matter even after controlling for ethnic-homeland fixed effects, and even in areas far from the capital. This suggests that the spatial imprecision and blurring of nighttime light data attenuated the association between national institutions and economic activity in their analysis. Nevertheless, our analyses generally corroborate their argument, particularly regarding the role of the limited penetration of national institutions in African countries.

Keywords: Nighttime lights, DMSP, VIIRS, national institutions.

JEL codes: O10, O43, N17, R12.

*Graduate School of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN: soyoka802@gmail.com

†Faculty of Economics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, JAPAN; e-mail: ueno.yotaro.85e@st.kyoto-u.ac.jp

‡Faculty of Economics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, JAPAN; e-mail: yamagoshi.toma.52x@st.kyoto-u.ac.jp

§Corresponding author: Graduate School of Economics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, JAPAN; tel: +81 75 753 3459; e-mail: kono@econ.kyoto-u.ac.jp

1 Introduction

Recent advancements in remote sensing and computer science have significantly expanded the data available for empirical research. These data provide proxies for a wide range of economic outcomes and characteristics on a substantially larger scale and with higher spatial resolution than traditional data (Donaldson and Storeygard, 2016). Examples include nighttime lights (NTL) (Hodler and Raschky, 2014; Henderson et al., 2018), transportation networks (Monte et al., 2018; Heblich et al., 2020; Allen and Arkolakis, 2022), climate and weather (Guiteras et al., 2015; Harari and Ferrara, 2018), environment (Jayachandran, 2009; Ito and Zhang, 2020), and agricultural productivity (Bustos et al., 2016; Costinot et al., 2016). Moreover, the recent integration of machine learning algorithms has added an additional layer of value to satellite data (Jean et al., 2016; Yeh et al., 2020).

Among these, NTL are one of the most widely adopted data as a proxy for economic activity after the seminal work by (Henderson et al., 2012), particularly at the sub-regional or granular levels where public data are absent (Storeygard, 2016; Eberhard-Ruiz and Moradi, 2019). However, recent studies have highlighted flaws in widely used NTL data, specifically the Defense Meteorological Satellite Program (DMSP) data. Issues, such as spatial imprecision and blurring make NTL a poor proxy for economic activities in lower spatial units (Gibson et al., 2021; Gibson and Boe-Gibson, 2021). Consequently, most remote sensing studies have shifted towards newer and more advanced data from the Visible Infrared Imaging Radiometer Suite (VIIRS). VIIRS data offer finer resolution and enhanced spatial precision, thereby addressing the limitations of DMSP data (Chen and Nordhaus, 2015, 2019; Perez-Sindin et al., 2021).

The issues of spatial imprecision and blurring are especially serious in analyses that rely on spatial regression discontinuity (RD) designs or variations across small spatial units to identify the parameters of interests. In such cases, spatial imprecision and blurring can obscure the actual discontinuities at the borders and diminish the differences between adjacent small spatial units. One influential study that employs these research designs is that of Michalopoulos and Papaioannou (2014) (hereafter referred to as MP). MP exploited the drawing of colonial boundaries in Africa, which partitioned more than 200 ethnicities into multiple countries, to estimate the impact of national institutions on regional development as measured by the DMSP NTL. Their estimation strategy was to compare NTL of pixels of ethnic-homelands that were partitioned into countries with different levels of institutions and to examine the existence of discontinuity of NTL at national borders. They found no systematic differences in NTL across borders with different levels of national institutions on average, which cautions against the importance of institutions in economic development (Acemoglu et al., 2001). Moreover, they documented heterogeneity in the correlation between national institutions and regional development, finding that national institutions matter only for regions close to capital cities because of the limited penetration of national institutions in African countries. However, the results derived from the variation across small spatial units in MP's study may be questioned owing to spatial imprecision and blurring in the DMSP data. Consequently, a reevaluation of MP findings using a more precise proxy for economic activities is warranted.

To address this concern, we replicated the MP analyses using VIIRS data. Our replication results generally support the MP argument although some differences exist in the estimation results. Notably, in the pixel level analyses, we found a significantly positive association between national institutions and NTL with a greater magnitude than that of MP. This suggests that spatial imprecision and blurring in the DMSP data attenuated the magnitude of the association between national institutions and NTL. On the contrary, the spatial RD design analyses revealed no evidence of NTL discontinuity at the border. This lack of association is likely because the bordering areas were mostly unindustrialized, and NTL served as a poor proxy for economic activity. Additionally, we addressed a potential concern with MP's RD analysis, which employed global high-order polynomials that would lead to noisy estimates and poor coverage of confidence intervals Gelman and Imbens (2019). We employed local quadratic polynomials as recommended by Gelman and Imbens (2019), yet still did not find a discontinuity in NTL at the border.

Regarding the heterogeneity in the association by distance to the capital cities, when we divided the sample into pixels near and distant from the capital cities by the median as in MP, we found significant positive associations between national institutions and NTL for both regions, and no significant differences in the magnitude of the correlation between the two regions. However, when we further divided the sample into more quantiles, we found a tendency for a diminishing correlation with distance from the capital, supporting MP's argument about the limited penetration of national institutions.

Overall, the results using VIIRS data support MP's argument. The absence of a discontinuity at the border in DMSP and VIIRS data explain the similarity of the results between the two datasets.

The lack of discontinuity in NTL at the border, coupled with the positive correlation between national institutions and NLT, suggests that the positive correlation detected in DMSP was driven by the levels of NTL in pixels distant from the borders. Consequently, spatial imprecision and blurring in small spatial units have a limited impact on the estimation results. This indicates that while DMSP NTL has issues of spatial imprecision and blurring, the reliability of the empirical analyses depends on where the detected correlation came from. If the correlation identified in the empirical analyses is mainly driven by pixels far from the border, where the values of the independent variable vary, the results obtained from DMSP are unlikely to be significantly altered by using VIIRS data.

This study contributes to the existing literature on the reliability of DMSP NTL as a proxy for economic activities in smaller spatial units. While previous studies have primarily focused on the predictive performance of DMSP NTL in relation to local economic activities compared to VIIRS NTL, there is a gap in the literature regarding whether using VIIRS data alters the results obtained from DMSP data. Our study addresses this gap and reveals that the impact depends on the origin of the detected empirical patterns. Specifically, if the identified empirical patterns were primarily derived from pixels located far from the borders where the value of the independent variable varied, the results obtained from DMSP data were unlikely to be significantly influenced when transitioning to VIIRS data.

In the following section, we briefly describe the data set. Section 3 presents our replication results, and Section 4 concludes the study.

2 Data

The MP used the DMSP NTL as a proxy for local economic activities. However, the primary purpose of DMSP satellites was to observe clouds for short-term Air Force weather forecasts, rather than to measure luminescence on earth. Owing to the limitations of the sensor and data-processing capacity, the DMSP NTL lack spatial accuracy and has blurring. Especially, these allocate light to places different from the point of origin, and the pixels are aggregated into 5×5 blocks prior to the data being sent to earth, which further spreads light from the point of origin (Abrahams et al., 2018; Gibson et al., 2021). As a result, DMSP data wrongly attribute light from towns and cities to hinterland areas, which makes DMSP NTL a poor proxy for economic activities in small spatial units (Gibson and Boe-Gibson, 2021).

In contrast, VIIRS Day-Night Band was designed to consistently measure the radiance of light coming from earth for research purposes, achieving a much higher spatial accuracy and temporal comparability than DMSP. Gibson (2021) showed that VIIRS NTL exhibits 80% higher predictive power for real GDP in a cross-section of subnational units in Europe compared with DMSP NTL. Additionally, VIIRS NTL provides a more detailed depiction of spatial inequality.

To replicate MP's results, we constructed NTL from the Version 4 DMSP-OLS Nighttime Lights Time Series data[1] averaged over 2007 and 2008 (DMSP 2007-08). For VIIRS NTL, we used the average masked data of the Version 2 VIIRS nighttime lights (V.2 VNL) annual composites for 2014 (VIIRS 2014), which was the first year in which the VIIRS provided values with stray light adjustment. Following MP, we aggregated NTL at the country-ethnic homeland level and to 12.5km $\times$ 12.5km pixels for the pixel-level analysis.

To focus on the discontinuity across national boundaries and exclude pixels that span these boundaries, we aggregated pixels by country when making 12.5km $\times$ 12.5km pixels. Notably, MP did not specify the pixel aggregation method at the boundaries, and because of the difference in the aggregation methods and settings in the software, we could not perfectly replicate the MP dataset. We restricted the data to pixels within 450 km of the national borders, as in MP. To investigate the impact of national institutions across borders, we used pixels belonging to ethnic homelands divided into multiple countries. Our dataset contained 49,832 pixels, which is 16.7% more than that of MP. To address the potential bias resulting from differences in aggregation methods, we constructed another dataset by aggregating the pixels at once (not by country) and assigning each pixel a country based on its centroid location. The results using this dataset, referred to as the AAO (aggregated at once) data, are reported in the appendix.[2]

For most of the control variables, we used data published by MP. However, owing to the difference in pixel aggregation methods, we used pixel level information on population density from the UN WPP-Adjusted Population Density (v4.11) for 2000. Additionally, we used the Petroleum Dataset v.1.2 (Lujala et al., 2007), an updated version of the dataset used by MP, to ensure consis-

[1] We used the files named `F1?YYYY_v4b_stable_lights.avg_vis.tif`, which remove fires and background noise.

[2] Our data construction and estimation codes are available on GitHub: `https://github.com/hk-git/replicate_MP`

tency and accuracy in our control variables.

The summary statistics of the major variables for the country-ethnic homeland level data and pixel level data are reported in panels (A) and (B) in Table 1, respectively. Compared with MP, the DMSP NTL measure in our dataset had slightly lower mean values with lower standard deviations. Our dataset recorded greater mean values for distance to the capital city, distance to the seacoast, and distance to the border.[3] This indicates that our dataset included more pixels with zero NTL in remote inland areas.

The VIIRS NTL had a significantly lower mean value and much lesser variation than DMSP NLT. This indicates that many of the variations in the DMSP NTL were caused by blurring and overestimation of the spatial extent of the NTL. This caused the DMSP NTL aggregated at the country-ethnic homeland level to have much more variations and greater mean values. Owing to the top-coding in the DMSP data (Gibson et al., 2021), the maximum value of the NTL in VIIRS was much larger.

Comparison between DMSP and VIIRS data

Before proceeding to the replication of MP, it is beneficial to document the discrepancy between DMSP NTL and VIIRS NTL.

Figure 1 illustrates the disparity in NTL between DMSP 2007-08 and VIIRS 2014 near the Angola-Namibia border, a depiction reminiscent of Figure IIB in MP. The DMSP data exhibited brighter NTLs in Namibia (south of the border), whereas the VIIRS data revealed minimal differences between the two countries. Notably, the extent of the illuminated region appeared to be larger in the DMSP data, reflecting the spatial imprecision and blurring inherent in the dataset. This observation raises concerns regarding the suitability of DMSP data for pixel level analyses.

Figure 2 shows the patterns of the measured NTL for different datasets. Panel (A) shows the scatter plots of VIIRS 2014 against DMSP 2007-08. Following MP's analysis, we compute the logarithm of the density of NTL + 0.01 to account for observations where the NTL density is zero.[4] While they are positively correlated with a correlation coefficient of 0.841, there are a non-negligible number of observations where NTL values substantially differ between the two datasets,

[3] The mean values of these variables in MP data are 531.839, 581.363, and 65.938, respectively.

[4] As discussed below, MP used ln(NTL density+0.01) for the country-ethnic homeland level analyses and an indicator for the NTL density exceeding zero for the pixel level analyses.

Table 1: Summary Statistics

(A) Country-Ethnicity level:

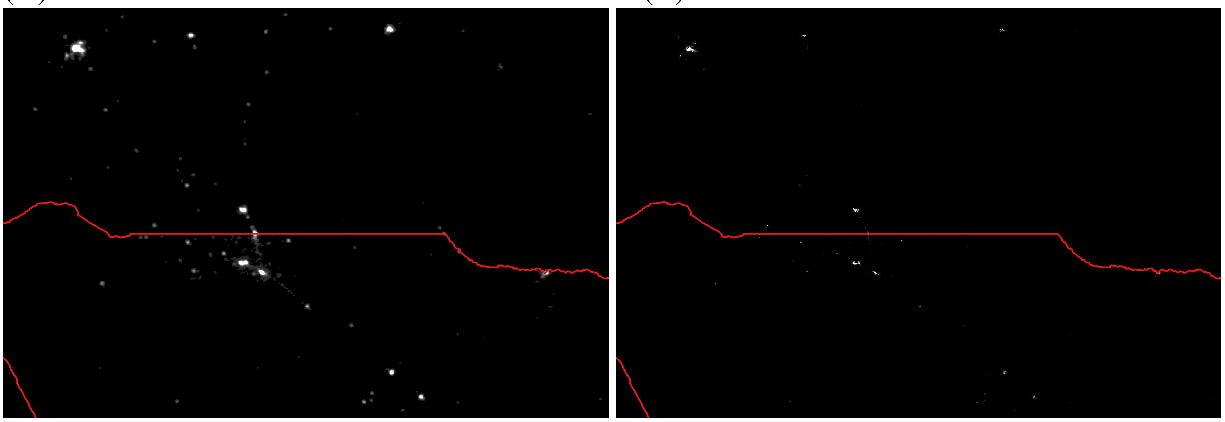
Statistic	N	Mean	St. Dev.	Min	Max
Light density (VIIRS2014)	507	0.040	0.154	0.000	2.369
Light density (DMSP2007)	507	0.239	0.808	0.000	8.853
Light density (DMSP2008)	507	0.230	0.774	0.000	8.583
Ln (0.01+Light density) (VIIRS2014)	507	−3.926	1.000	−4.605	0.867
Ln (0.01+Light density) (DMSP2007)	507	−3.152	1.605	−4.605	2.182
Ln (0.01+Light density) (DMSP2008)	507	−3.165	1.597	−4.605	2.151
Ln (0.01+Population density)	507	2.899	1.610	−4.388	6.334
Distance to the capital city	507	530.979	373.550	11.164	1,892.499
Distance to the sea coast	507	598.585	437.415	0.313	1,789.598
Distance to the border	507	36.986	38.078	0.278	246.214
Rule of law	507	−0.914	0.573	−2.197	0.615
Control of corruption	507	−0.797	0.501	−1.664	0.814

(B) Pixel level:

Statistic	N	Mean	St. Dev.	Min	Max
Light density (VIIRS2014)	49,832	0.048	0.947	0.000	111.298
Light dummy (VIIRS2014)	49,832	0.137	0.344	0	1
Light density (DMSP2007)	49,832	0.235	1.811	0.000	62.825
Light density (DMSP2008)	49,832	0.220	1.754	0.000	61.919
Light dummy (Average of DMSP2007 and DMSP2008)	49,832	0.099	0.299	0	1
Ln (0.01+Light density) (VIIRS2014)	49,832	−4.393	0.775	−4.605	4.712
Ln (0.01+Light density) (DMSP2007)	49,832	−4.220	1.324	−4.605	4.141
Ln (0.01+Light density) (DMSP2008)	49,832	−4.233	1.296	−4.605	4.126
Ln (0.01+Population density)	49,832	1.325	2.178	−4.605	9.454
Distance to the capital city	49,832	622.125	391.184	3.211	1,940.113
Distance to the sea coast	49,832	659.150	444.847	0.129	1,810.693
Distance to the border	49,832	96.087	95.102	0.0004	449.995
Oil deposit dummy	49,832	0.020	0.141	0	1
Rule of law	49,832	−0.876	0.618	−2.197	0.615
Control of corruption	49,832	−0.756	0.552	−1.664	0.814

Figure 1: Nighttime light near the border between Angola and Namibia

(A) DMSP2007-08 (B) VIIRS 2014

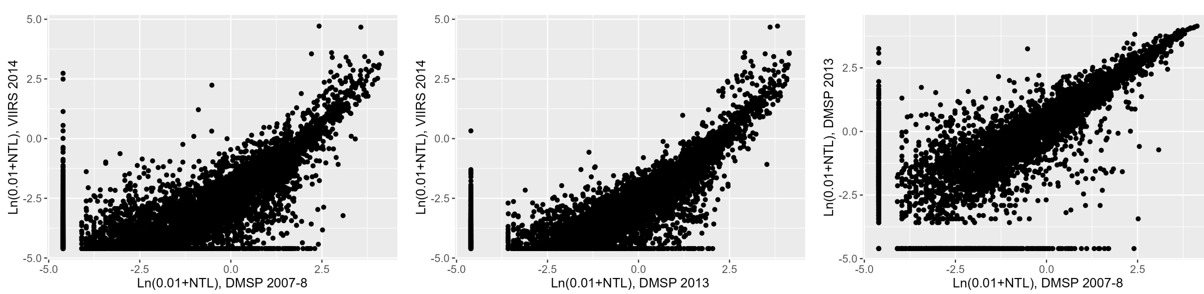


The image on the left displays night lights based on the DMSP stable lights annual composite of 2013, while the image on the right shows the masked average radiance derived from the Version 2 VIIRS night-time lights annual composite of 2014.

reflecting the issue of spatial imprecision and blurring in DMSP data.

Figure 2: Comparison between DMSP and VIIRS

(A) DMSP2007-08 vs. VIIRS2014 (B) DMSP2013 vs. VIIRS2014 (C) DMSP2007-08 vs. DMSP2013



It is plausible that the difference between DMSP 2007-08 and VIIRS 2014 results from the difference in measurement timing. To address this concern, we present a scatter plot between VIIRS 2014 and DMSP NTL for 2013, the latest year when DMSP NTL is available(panel (B) of Figure 2). The pattern is similar to that between DMSP 2007-08 and VIIRS 2014, with the similar correlation coefficient (0.866). Notably, NTL in DMSP 2007-08 and DMSP 2013 exhibited significant differences for some observations, consistent with the lack of temporal consistency of DMSP NTL (Gibson et al., 2021).[5]. This indicates that the discrepancy between DMSP NTL and

[5]The correlation coefficient is 0.905, which is greater than the correlation coefficients between DMSP 2007-08 (or DMSP 2013) and NTL

VIIRS NTL mostly reflects measurement errors in DMSP.

For pixel level analyses, MP uses an indicator of NTL density exceeding zero. Table 2 presents the discrepancy in this variable between DMSP and VIIRS data. Among the pixels whose DMSP NTL exceeded zero (DMSP = 1), 15-18$ of the pixels recorded no NTL in VIIRS (VIIRS = 0), mostly resulting from spatial imprecision and blurring. In contrast, there are substantial observations where VIIRS detected nightlights (VIIRS = 1), but DMSP did not (DMSP = 0) by as much as 36-40%. This considerable discrepancy casts doubt on the reliability of the pixel level analyses.

Table 2: Comparison between DMSP and VIIRS (continuous NTL)

(A) DMSP2007-08 vs. VIIRS 2014 (binary)

Whether the pixel is lit	VIIRS = 0	VIIRS = 1
DMSP = 0	42118	2786
DMSP = 1	883	4045

(B) DMSP2014 vs. VIIRS2014 (binary)

Whether the pixel is lit	VIIRS = 0	VIIRS = 1
DMSP = 0	42228	2450
DMSP = 1	773	4381

3 Results

3.1 Baseline specifications

Table 3 shows the estimation results using the country-ethnic homeland level data, which replicate Panel A of Table III in MP. The estimation equation is as follows:

$$y_{ic} = \alpha_0 + \gamma IQL_c + \mathbf{X}_{ic}\beta + \alpha_i + \epsilon_{ic}, \tag{1}$$

where y_{ic} is ln(0.01 + NTL density) of the historical homeland of ethnic group i in country c. The vector of the control variable, $\mathbf{X}_{ic}$ includes the same variables as in MP, namely population density, area, water area indicator, mean elevation, land suitability for agriculture, malaria suitability index, oil deposit indicator, diamond mine indicator, distance to the capital city, distance to the sea coast, and distance to the border. The inclusion of the ethnic homeland fixed effect α_i is the key identification strategy of MP, that is, comparing economic development in homelands of the same ethnicity in adjacent countries with different levels of institutions. The standard errors are doubly clustered along the country and ethnic homeland dimensions.

The national institutions of country i, denoted by IQL_c, are measured by the rule of law and control of corruption obtained from the dataset published by MP. The upper panel shows the results using the rule of law as IQL_c, and the lower panel shows the results using corruption control as IQL_c.

The results obtained from the DMSP data (Columns (1)–(4)) align closely with those in MP, confirming that our replication procedure for data construction works well. While significantly positive associations between national institutions and NTL were observed when we did not control for the ethnic homeland fixed effects, once we controlled for them, the coefficients on the rule of law and control of corruption dropped substantially and became statistically insignificant.

When we use the VIIRS data, the association between national institutions and NTL becomes weaker, irrespective of whether the rule of law or control of corruption is used as a measure of national institutions. This attenuation could be attributed to the fact that the DMSP overestimated the dispersion of economic activity across ethnic homelands due to its blurring.[6] Blurring tends to make areas surrounding economically active regions appear brighter than they are, thereby leading to greater dispersion across regions. However, MP's main argument remains valid with the use of VIIRS data, that is, the significant positive correlation between national institutions and development across ethnic homelands disappears when the same ethnic groups across different countries are compared by including ethnic homeland fixed effects.

water area indicator, mean elevation, land suitability for agriculture, malaria suitability index, oil deposit indicator, diamond mine indicator, distance to the capital city, distance to the sea coast, and distance to the border.

Table 4 presents the estimation results using pixel-level data, where the estimation equation is as follows:

$$y_{pic} = \tilde{\alpha}_0 + \tilde{\gamma} IQL_c + \mathbf{X}_{pic}\tilde{\beta} + \tilde{\alpha}_i + \nu_{pic}. \tag{2}$$

Following MP the dependent variable y_{pic} is a binary indicator that equals 1 if pixel p belonging to ethnic homeland i in country c is lit, and 0 otherwise. The control variables were similar to MP's.[7]

[6]Difference in results between DMSP data and VIIRS data is not due to difference in the years when the NTL was measured. In Panel (A) of Appendix Table 2, we estimated a similar regression using DMSP 2013. The magnitudes of the associations between national institutions and NTL were similar to those using DMSP 2007-08. This finding underscores the importance of precision of NTL data when evaluating the spatial distribution of economic activity and its correlates.

[7]The control variables include the logarithm of the population density, logarithm of the pixel area, and location and geography variables such as the distance to the border, pixel means of the malaria suitability index, elevation, and indicators for diamond mine and petroleum.

Table 3: Ethnicity-country level analysis (binary NTL)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Night time light	DMSP				VIIRS			
Rule of law	0.693	0.249	0.600	0.342	0.430	0.113	0.345	0.160
Double-clustered std. err	(0.188)	(0.195)	(0.184)	(0.210)	(0.128)	(0.123)	(0.128)	(0.134)
Observations	507	507	507	507	507	507	507	507
Adjusted R-squared	0.294	0.644	0.419	0.654	0.252	0.655	0.398	0.658
Controls								
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Night time light	DMSP				VIIRS			
Corrupt	0.870	0.355	0.698	0.421	0.525	0.167	0.406	0.211
Double-clustered std. err	(0.220)	(0.228)	(0.217)	(0.245)	(0.161)	(0.144)	(0.153)	(0.159)
Observations	507	507	507	507	507	507	507	507
Adjusted R-squared	0.306	0.646	0.420	0.654	0.260	0.656	0.400	0.658
Controls								
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the logarithm of population density, logarithm of the partitioned area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, mean elevation, land suitability for agriculture, malaria suitability, and indicators for diamond mine and petroleum.

Standard errors are doubly clustered in the country and ethnic homeland dimensions.

The analysis using the DMSP 2007-08 data closely replicates MP's results. We found significantly positive associations between national institutions and NTL when we do not control for the ethnic homeland fixed effects, but once we control for them, the significant correlation disappears.[8]

However, when we use the VIIRS data, we find significant positive correlations, even after controlling for ethnic homeland fixed effects, although the magnitude of the correlation decreases with the inclusion of ethnic homeland fixed effects. The point estimates also tended to be larger in the VIIRS data than in the DMSP data, reflecting the spatial imprecision and blurring of the DMSP. The blurring effect, causing pixels surrounding illuminated areas to appear brighter, coupled with the concentration of economic activity in central areas far from borders, leads to larger differences when aggregating the DMSP data at the country-ethnic homeland level but smaller differences when comparing pixels across borders.

The AAO data yield similar patterns, albeit with greater estimated coefficients (Appendix Table 3). Given that the AAO data contain fewer pixels at the borders (because we merged the pixels and assigned their country by the location of the centroids), the increase in the coefficient may suggest that the detected association between national institutions and NTL is primarily driven by pixels located away from the borders. We discuss this point later when presenting results based on the regression discontinuity design in section 3.2.

Notably, MP used the binary indicator for the pixel to be lit as the outcome variable for pixel level analysis. As a robustness check, we conducted similar analyses using the level of the measured NTL in three different ways. First, we used log(0.01+NTL) as in the country-ethnic homeland level analysis. However, this measure gives disproportionate weight to changes from 0 to 1. For example, consider two cases of changes in the NTL: (a) from 0 to 1 and (b) from 1 to 2, and denote the outcome measure as y. If we use log(0.01+NTL) as y, the changes in y are 4.615 in (a) and 0.688 in (b). This implies that the change from 0 to 1 is given nearly seven times more weight than the change from 1 to 2. Given that pixel level data contain many zero NTL values, this disproportionate weight of the difference between 0 and 1 is undesirable. It also inflates the influence of measurement errors due to natural fires and other light sources that are not attributable to economic activity. To alleviate this problem, we used log(1+NTL) as the dependent variable.

[8]The results using the 2013 DMSP data are similar, which are reported in Panel (B) of Appendix Table 2.

Table 4: Pixel level analysis (binary outcomes)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Night time light	DMSP				VIIRS			
Rule of law	0.098	0.038	0.076	0.032	0.149	0.066	0.113	0.054
Double-clustered std. err	(0.035)	(0.022)	(0.029)	(0.022)	(0.043)	(0.026)	(0.033)	(0.024)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Adjusted R-squared	0.105	0.314	0.168	0.323	0.169	0.344	0.220	0.351
Controls								
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Night time light	DMSP				VIIRS			
Corrupt	0.126	0.054	0.100	0.039	0.181	0.088	0.141	0.065
Double-clustered std. err	(0.042)	(0.027)	(0.036)	(0.030)	(0.051)	(0.035)	(0.042)	(0.033)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Adjusted R-squared	0.118	0.314	0.176	0.323	0.181	0.345	0.227	0.350
Controls								
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

In this case, the changes in y are 0.693 for (a) and 0.405 for (b), which is much more reasonable. Third, to deal directly with the problem of many zeros, we used the Poisson Pseudo Maximum Likelihood (PPML), which is widely used for international trade analysis involving many zeros (Silva and Tenreyro, 2006; Head and Mayer, 2014).

Table 5 presents the results obtained by using these three specifications. Overall, these three methods yielded qualitatively similar results. In both the DMSP and VIIRS data, we found a significantly positive association between national institutions and NTL when ethnic homeland fixed effects were not controlled for; however, including ethnic homeland fixed effects rendered this association statistically insignificant. These findings suggest the robustness of MP's conclusion to alternative NTL data or econometric specifications.

Note that there areexists significant differences in the coefficient values between the regressions using log(0.01+NTL) and log(1+NTL). Specifically, the coefficients ofn national institutions are more than four4 times larger when using log(0.01+NTL) in the DMSP data, and 10 times larger in the VIIRS data. This disparity arises because log(0.01+NTL) assigns a gives disproportionate weight to changes from 0 to 1, and the VIIRS data, without blurring, contains more zeros and smaller values thancompared to the DMSP data. These results underscore the signicificance of considering the functional forms of outcome variables and their impact on the results, particularly in the presence of measurement errors.

3.2 Regression Discontinuity

The above analysis with ethnic homeland fixed effects exploits variations in NTL within the same ethnic homeland divided into different countries with varying levels of national institutions. Another empirical approach is to focus on border discontinuities. If national institutions matter, one would observe a discrepancy in the level of economic activity at national borders for pairs of countries with different levels of national institutions. Based on this idea, MP conducted the spatial RD analysis using the estimation equation

$$y_{pic} = \ddot{\alpha}_i + \ddot{\gamma} IQL_c^{HIGH} + f(BD_{pic}) + \delta f(BD_{pic}) IQL_c^{HIGH} + \mathbf{X}_{pic}\ddot{\beta} + \ddot{\epsilon}_{pic}, \tag{3}$$

where y_{pic} is the binary indicator for the NTL to be lit in pixel p belonging to the historical homeland of ethnic group i in country c, α_i are the ethnic homeland fixed effects, and IQL_c^{HIGH} is a

Table 5: Pixel level data: Continuous NTL

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$ln(0.01 + NTL)$	DMSP				VIIRS			
Rule of law	0.468	0.199	0.364	0.161	0.268	0.104	0.210	0.087
	(0.181)	(0.110)	(0.149)	(0.121)	(0.099)	(0.063)	(0.082)	(0.069)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Adjusted R-squared	0.116	0.375	0.202	0.386	0.119	0.331	0.194	0.341
Corrupt	0.595	0.278	0.481	0.199	0.334	0.140	0.271	0.097
	(0.215)	(0.146)	(0.186)	(0.169)	(0.119)	(0.084)	(0.103)	(0.098)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Adjusted R-squared	0.129	0.375	0.210	0.386	0.130	0.332	0.200	0.340
$ln(1 + NTL)$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rule of law	0.098	0.047	0.077	0.035	0.023	0.008	0.017	0.004
	(0.042)	(0.028)	(0.036)	(0.032)	(0.009)	(0.009)	(0.009)	(0.010)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Adjusted R-squared	0.083	0.329	0.170	0.340	0.032	0.143	0.074	0.150
Corrupt	0.124	0.064	0.102	0.042	0.028	0.009	0.022	0.000
	(0.050)	(0.040)	(0.044)	(0.046)	(0.011)	(0.012)	(0.010)	(0.015)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Adjusted R-squared	0.091	0.329	0.176	0.340	0.034	0.143	0.075	0.150
PPML	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rule of law	1.193	0.191	0.563	-0.064	0.671	-0.164	-0.249	-0.386
	(0.338)	(0.163)	(0.462)	(0.292)	(0.301)	(0.211)	(0.620)	(0.287)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Corrupt	1.230	0.319	0.641	0.080	0.787	0.344	-0.100	-0.271
	(0.279)	(0.236)	(0.477)	(0.378)	(0.262)	(0.400)	(0.652)	(0.345)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

dummy variable that takes the value of 1 for pixels falling in a country with relatively better institutions. Function $f(BD_{pic})$ represents the polynomial function of the distance from the centroid of pixel c to the national border, BD_{pic}. As the value of IQL_c^{HIGH} switches at the border (i.e., at the pixel where $BD_{pic} = 0$), the coefficient $\ddot{\gamma}$ captures the effect of national institutions at the borders. In the baseline specification, following MP, the third and fourth polynomials are used for $f(BD_{pic})$, allowing the polynomial terms to differ on either side of the border, as captured by δ. To restrict the analysis to observations with sufficient variation in national institutions across borders, we used pairs of countries whose levels of national institutions differ substantially (i.e., the difference exceeded the 75th percentile values), as MP did.

Table 6 presents the results. The point estimates of $\ddot{\gamma}$ tend to be larger in the VIIRS data compared to the DMSP data, which is consistent with the fact that the DMSP data underestimates the discontinuity owing to blurring. However, the associations are rarely significant, supporting MP's argument that national institutions play a limited role at borders because of the limited penetration of national institutions in African countries.

Table 6: Border regression discontinuity (RD) estimates (global polynomial regressions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Data	DMSP						VIIRS					
Band Width	All pixels		100km		50km		All pixels		100km		50km	
RD polynominal	3rd	4th	3rd	4th	3rd	4th	3rd	4th	3rd	4th	3rd	4th
Rule of law	0.040	0.023	0.007	0.081	0.019	-0.007	0.056	0.051	0.025	0.114	0.068	0.027
Double-clustered std. err	(0.040)	(0.023)	(0.024)	(0.061)	(0.036)	(0.032)	(0.045)	(0.029)	(0.027)	(0.062)	(0.044)	(0.037)
Observations	20192	20192	20192	11495	11495	11495	20192	20192	20192	11495	11495	11495
Adjusted R-squared	0.105	0.357	0.357	0.126	0.417	0.418	0.134	0.351	0.353	0.152	0.414	0.416

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Data	DMSP						VIIRS					
Band Width	All pixels		100km		50km		All pixels		100km		50km	
RD polynominal	3rd	4th	3rd	4th	3rd	4th	3rd	4th	3rd	4th	3rd	4th
Corrupt	0.012	0.001	-0.007	-0.004	-0.006	-0.009	0.013	0.003	-0.005	0.000	0.000	-0.009
Double-clustered std. err	(0.019)	(0.017)	(0.014)	(0.015)	(0.015)	(0.020)	(0.021)	(0.019)	(0.018)	(0.020)	(0.020)	(0.026)
Observations	49710	49710	21145	21145	13390	13390	49710	49710	21145	21145	13390	13390
Adjusted R-squared	0.302	0.302	0.300	0.300	0.320	0.320	0.334	0.334	0.341	0.341	0.355	0.355

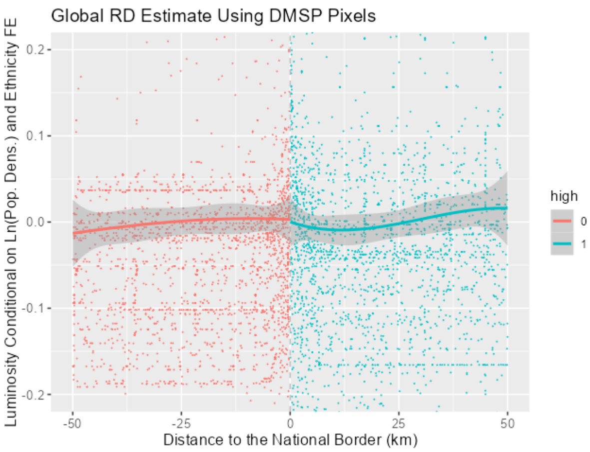
The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the ethnic homeland fixed effects, logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

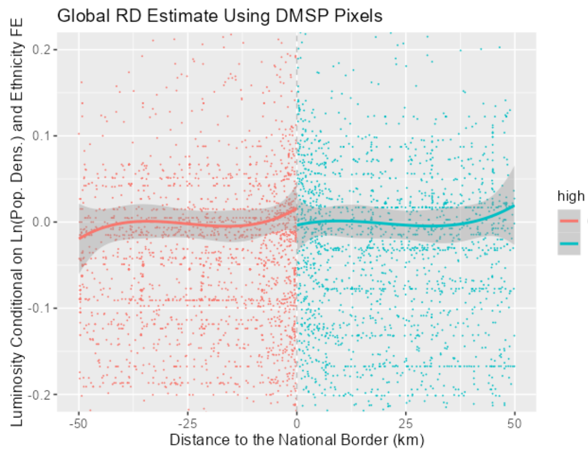
Figure 3 replicates Figure VB of MP, which provides a visual representation of the RD results using the pixels within 50 km of the border and third-order polynomials. Positive distance values (to the right of the figures) are assigned to countries with better national institutions ($IQL_c^{HIGH} = 1$). Consistent with the results in Table 6, there were no discernible differences in NTL at the borders in either dataset.

Figure 3: Regression discontinuity (3rd order, 50km bandwidth)

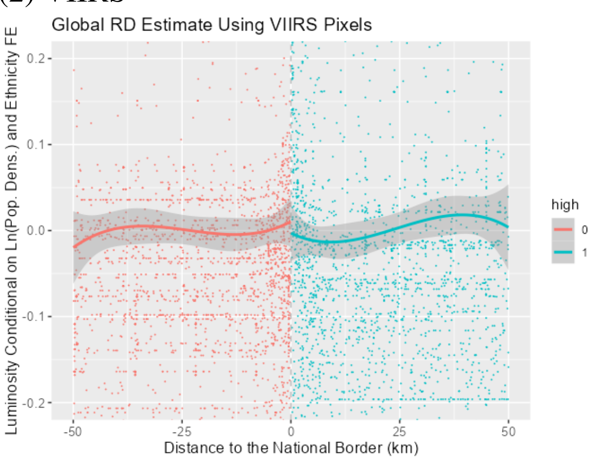
(A) Rule of law | (B) Corruption

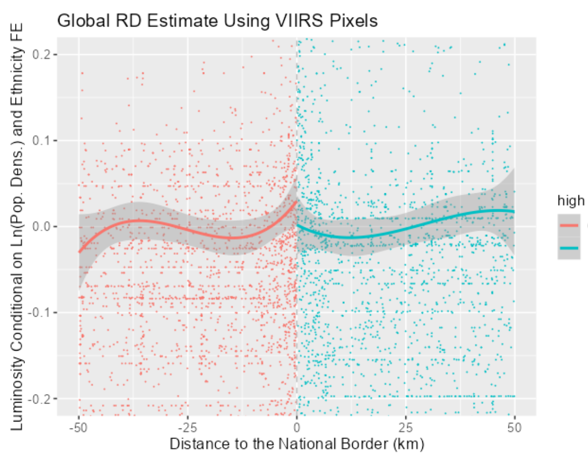
(1) DMSP





(2) VIIRS



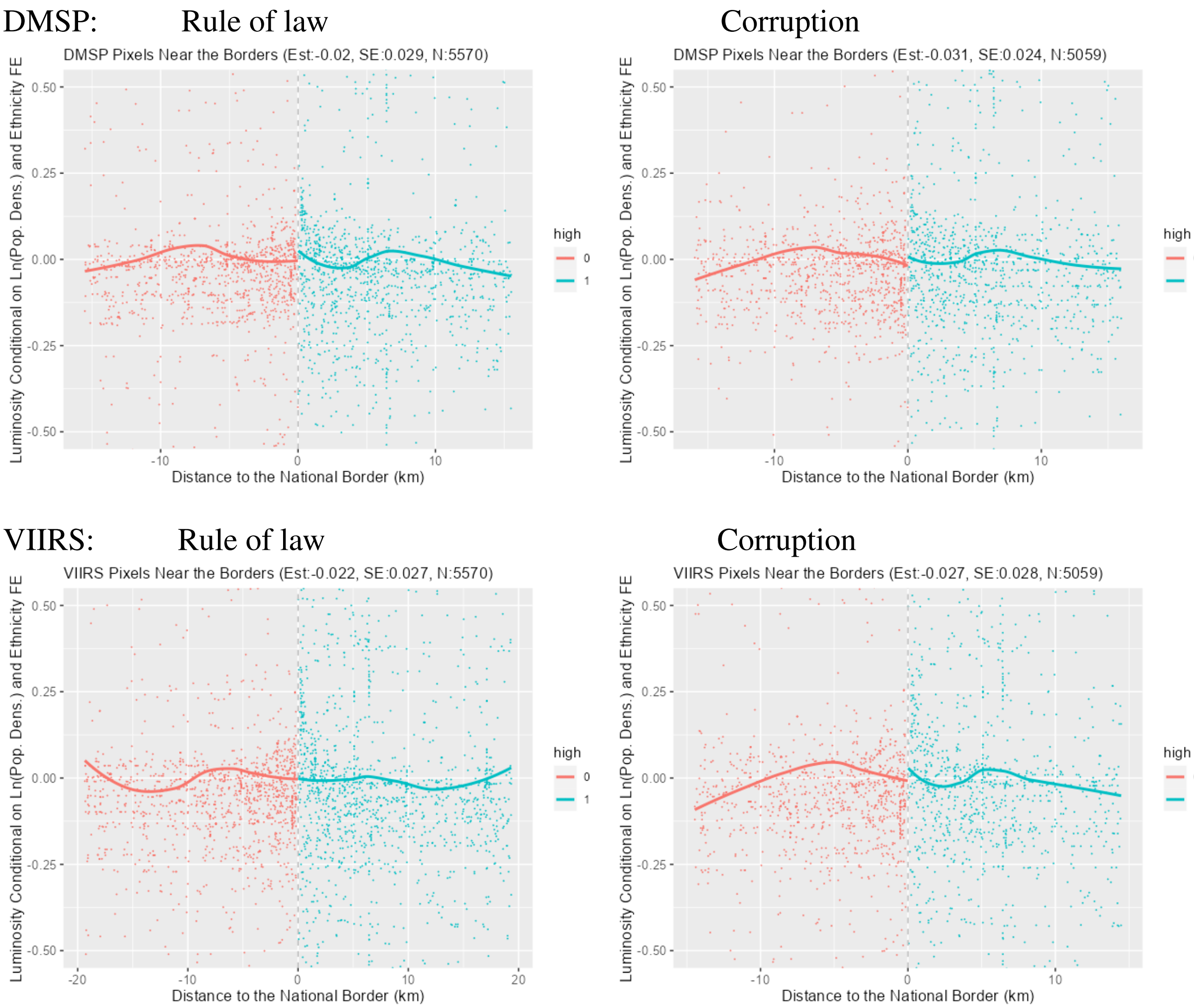


The figures offer a visual representation of the RD analyses with a third-order polynomial function fitted. On the horizontal axis, positive values correspond to the distance from the borders for pixels belonging to the country with a high level of national institutions, whereas negative values represent the distance from the borders for pixels belonging to the country with a low level of national institutions.

However, as highlighted by Gelman and Imbens (2019), controlling for global high-order polynomials in RD analyses should be avoided because it leads to noisy estimates and poor coverage of confidence intervals, and the results can be sensitive to the degree of the polynomial and observations far from the threshold. To address this concern, we adopt the approach recommended by Gelman and Imbens (2019) and implement the RD analyses with local quadratic polynomials.

Figure 4 shows the RD graphs with local quadratic polynomials. The point estimates of $\ddot{\gamma}$ and their standard errors are presented in the graphs.[9] In the DMSP and VIIRS data, we found no significant discontinuity at the borders.[10]

Figure 4: Regression discontinuity: Binary outcome, local quadratic



In summary, the RD analyses revealed no significant discontinuity in NTL at the borders. This

[9]Fitted lines in the graphs were drawn using the local quadratic. Note that the estimated value of $\ddot{\gamma}$ and the discontinuity of the fitted line at the borders may not necessarily coincide because the RD estimates use the kernel, whereas the quadratic lines are fitted without considering the kernel.

[10]Appendix Figure 1 illustrates the analogous RD graphs for the continuous outcome $\ln(1 + NTL)$. As in the case of the binary outcome, the graphs show no significant discontinuity in NTL at the borders.

result aligns with our earlier discussion of the results using the AAO data, supporting the argument that the observed correlations are mainly driven by pixels located farther away from the borders.

3.3 Heterogeneity by the distance to the capital

One of MP's important findings is that the association between national institutions and NTL was significantly positive only in regions near the national capital, which supports the idea that the influence of national institutions weakens as one moves farther away from the capital owing to the limited law enforcement capacity of African nations. We reassess this claim using high-quality VIIRS data.

Table 7 presents the replication results of MP's Table VIII, where the level of national institutions is measured by the rule of law in the top two panels and by the control of corruption in the bottom two panels. Similar to MP, we estimated the association between national institutions and NTL separately for pixels close to and far from the national capitals, using the median distance to the capitals as the cutoff. Distance to a capital is measured in terms of absolute and relative distance, the latter being defined as the pixel level distance to the capital divided by the maximum distance to the capital in each country. [11] For brevity, we referred to pixels with absolute (relative) distances to the capital less than the median as pixels absolutely (relatively) close to the capital and pixels with absolute (relative) distances greater than the median as pixels absolutely (relatively) far from the capital. We present the results using all pixels and those using only pixels within 50km of either side of the border (100km bandwidth pixels). Here, we report the p values instead of standard errors to facilitate the discussion of multiple hypothesis testing.

However, our results were not as pronounced as those of MP. MP found significant positive associations between the rule of law and NLT in pixels relatively close to the capital, and between the control of corruption and the NTL in pixels relatively or absolutely close to the capital, regardless of whether all pixels or the 100km bandwidth pixels were used. In our DMSP dataset (columns

[11] MP used the median distance to the capital across all pixels as the cutoff, reported as 367.4km, although it is not explicitly mentioned where this number originated. Based on the summary statistics, the median distance to the capital was 439km. While we attempted to replicate this number using published pixel level data, we were unable to do so using any sample selection method. Therefore, we used the median distance for the entire sample. As we included ethnic homeland fixed effects, we excluded observations of ethnic homelands that belonged to only one country in this subsample.

Table 7: Heterogeneity: National institutions and regional development within partitioned ethnicities close to and far from the capital cities (binary NTL)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DMSP		VIIRS		DMSP		VIIRS	
	Absolute distance to the capital				Relative distance to the capital			
All	Close	Far	Close	Far	Close	Far	Close	Far
Rule of law	0.058	0.046	0.086	0.082	0.094	0.019	0.096	0.041
p-value	(0.161)	(0.110)	(0.008)	(0.004)	(0.030)	(0.269)	(0.021)	(0.079)
FDR q-value	[0.138]	[0.571]	[0.017]	[0.017]	[0.131]	[0.571]	[0.028]	[0.042]
Observations	24916	24916	24916	24916	24904	24928	24904	24928
Adjusted R-squared	0.389	0.124	0.412	0.164	0.363	0.261	0.402	0.277
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.729		0.745		0.107		0.255	
100km bandwidth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rule of law	0.076	0.015	0.125	0.058	0.034	0.012	0.056	0.052
p-value	(0.165)	(0.529)	(0.004)	(0.032)	(0.352)	(0.520)	(0.118)	(0.016)
FDR q-value	[0.138]	[0.710]	[0.014]	[0.034]	[0.241]	[0.710]	[0.054]	[0.025]
Observations	10630	10631	10630	10631	10618	10643	10618	10643
Adjusted R-squared	0.384	0.181	0.429	0.230	0.383	0.244	0.418	0.297
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.229		0.126		0.587		0.924	
All	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corruption	0.078	0.066	0.099	0.118	0.120	0.032	0.111	0.065
p-value	(0.170)	(0.042)	(0.038)	(0.004)	(0.031)	(0.180)	(0.035)	(0.048)
FDR q-value	[0.138]	[0.511]	[0.033]	[0.017]	[0.131]	[0.571]	[0.033]	[0.034]
Observations	24916	24916	24916	24916	24904	24928	24904	24928
Adjusted R-squared	0.389	0.125	0.412	0.165	0.362	0.261	0.401	0.278
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.630		0.803		0.145		0.456	
100km bandwidth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corruption	0.124	0.030	0.158	0.077	0.066	0.026	0.070	0.071
p-value	(0.043)	(0.237)	(0.002)	(0.028)	(0.131)	(0.311)	(0.106)	(0.013)
FDR q-value	[0.131]	[0.571]	[0.013]	[0.034]	[0.138]	[0.571]	[0.054]	[0.025]
Observations	10630	10631	10630	10631	10618	10643	10618	10643
Adjusted R-squared	0.386	0.182	0.429	0.231	0.384	0.245	0.418	0.298
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.074		0.080		0.424		0.985	

The table reports the estimated coefficients of the regression, with p values based on the standard errors clustered at the country and ethnic homeland in parentheses. The control variables include the ethnic homeland fixed effects, logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

(1), (2), (5), and (6)), the results align with those of MP when we used all pixels; however, when we used 100km bandwidth pixels, we found a significant association for the control of corruption in pixels close to the capital.

When we used the VIIRS dataset, we found significant positive associations between national institutions and NTL for pixels relatively or absolutely close to the capital in most cases; however, we found similar significant associations for pixels far from the capital. For example, using all observations, the estimated coefficient of the rule of law is 0.086 for pixels absolutely close to the capital and 0.082 for pixels absolutely far from the capital, both of which are statistically significant at the 5% level. The difference between these two coefficients was insignificant ($p = 0.745$). Overall, we found significantly positive associations between the rule of law and NTL in three out of four specifications for pixels close to the capital, and found significantly positive associations in three out of four specifications for pixels far from the capital. The same pattern holds for corruption control. We found significantly positive associations in three out of four specifications for pixels close to the capital and significantly positive associations in all specifications for pixels far from the capital. Furthermore, the magnitude of the coefficient is comparable between pixels close to the capital and pixels far from the capital in many cases, and we reject the null hypothesis of the equality of the coefficients in only one case (control of corruption for 100km bandwidth pixels).

As tested many hypotheses, some significant results may be due to false positives. To account for multiple hypothesis testing, we applied the BKY procedure Benjamini et al. (2006) to control for the false discovery rate (FDR). While Benjamini et al. (2006) derived this procedure by assuming independence among the null hypotheses, their simulation exercises show that it performs well for positively dependent p values, as is the case here. The family of hypotheses for multiple hypothesis testing consisted of close/far subsamples for each dataset. For example, the eight hypotheses corresponding to columns (1) and (5), that is, hypothesis testing for the subsample of pixels close to the capital in the DMSP data, are treated as a family of hypotheses. In brackets, we report the FDR q values corresponding to the smallest level q at which the hypothesis was rejected (Anderson, 2008).

For the DMSP dataset, we found that none of the eight hypotheses exhibited q values below 0.1 for either pixels close to the capital or pixels far from the capital. However, when we used the VIIRS dataset, the FDR q values were smaller than 0.054 for all cases for pixels closer to the capital and pixels far from the capital. This casts some doubt on MP's results of the positive association between national institutions and economic development for areas close to the capital due to the limited law enforcement capacity of African nations.

To further explore heterogeneity by distance, we divided the sample into quartiles, quintiles, and deciles by distance from the capital and run the same regression. Note that large absolute distance values were observed in a few large countries. Consequently, if we divide the sample based on absolute distance, the subsample of pixels far from the capital includes pixels from a limited number of countries, resulting in small variations in national institutions. Figure 5 visually shows this pattern, drawing scatter plots of absolute or relative distance and national institutions. Notably, at large absolute distances, only a few distinct values are observed for national institutions. By contrast, because the relative distance is computed as the pixel level distance to capital divided by the maximum distance to capital in each country, it varies from 0 to 1 for all countries, and the number of countries for each quantile is well balanced. For this reason, we mainly report the results using relative distance. [12]

Figure 6 plots the estimated coefficients of the measures of the national institutions and their confidence intervals for each quartile (left panels), quintile (middle panels), and decile (right panels). The upper two panels show the results for the rule of law, and the lower two panels show the results for the control of corruption. The red dots and lines indicate point estimates and confidence intervals obtained from the DMSP data, respectively whereas the blue dots and lines indicate point estimates and confidence intervals obtained from the VIIRS data, respectively.

The patterns of the estimation results are quite similar between DMSP and VIIRS data, regardless of whether all pixels or 100km bandwidth pixels are used. National institutions matter only for pixels close to the capital. While positive associations were detected for the 3rd quintile or the 6th decile, the association between national institutions and NTL was insignificant for pixels other than those close to the capital.

Note that the point estimates tend to be slightly higher when using VIIRS data[13]

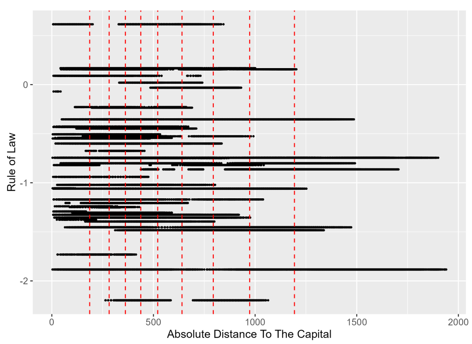
[12]Appendix Figure 2 shows a scatter plot of the relative and absolute distances. As the relative distance is computed as the pixel level distance to the capital divided by the maximum distance to the capital in each country, the relationship between the relative distance and absolute distance within the same country should be linear. This implies that points on a given linear line in Appendix Figure 2 represent pixels within the same country, and different lines correspond to different countries. The estimation results based on the absolute distance are reported in Appendix Figure 6.

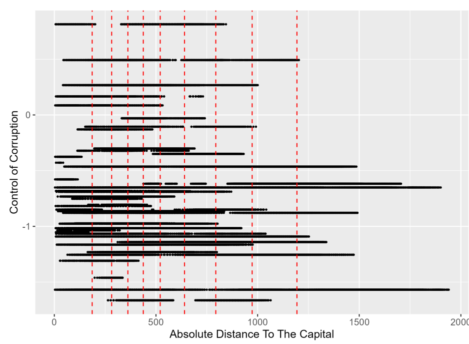
[13]Since we separately estimated the regression for each quantile and the coefficients on the controls varied across these regressions, averaging the coefficients in the figure will not coincide with the coefficient reported in Table 7.

Figure 5: National institutions and distance measures

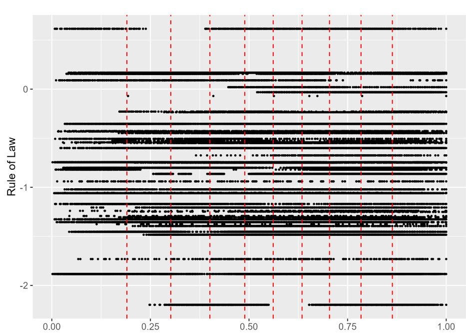
(A) Rule of law (B) Control of corruption

(1) Absolute distance





(2) Relative distance



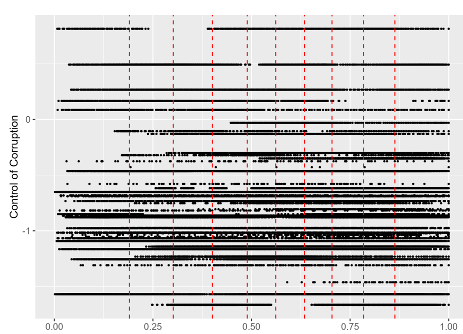
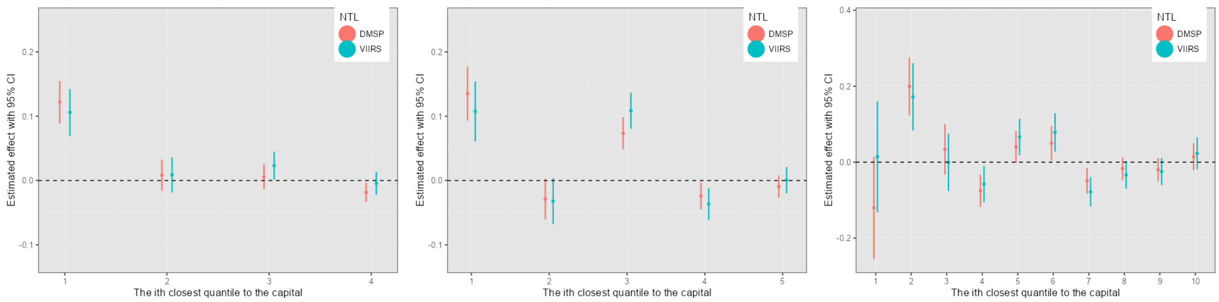


Figure 6: Heterogeneity by relative distance to the capital cities (binary NTL): Finer divisions

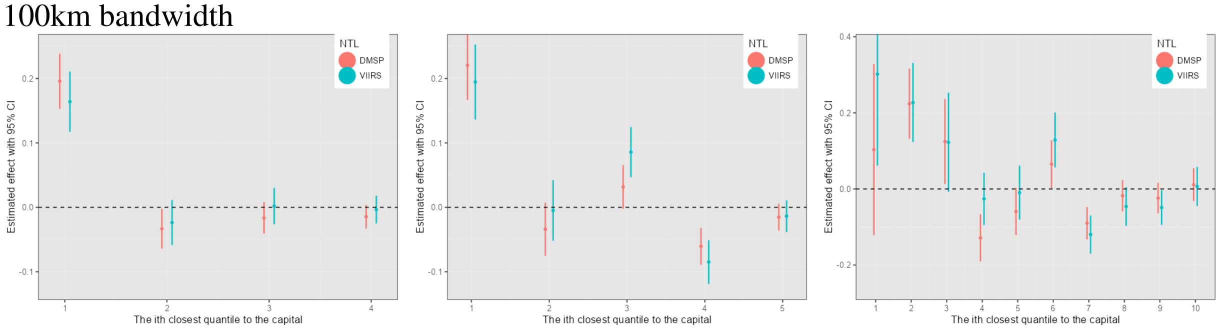
(A) Quartiles (B) Quintiles (C) Deciles

(1) Rule of law

All

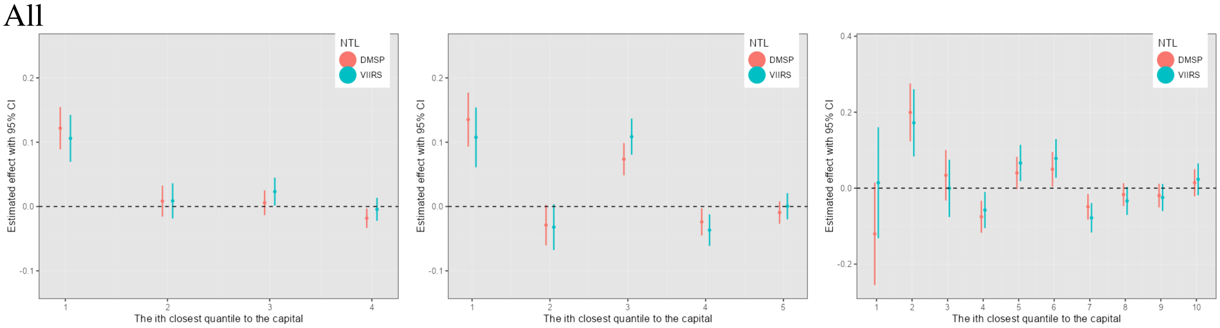


100km bandwidth

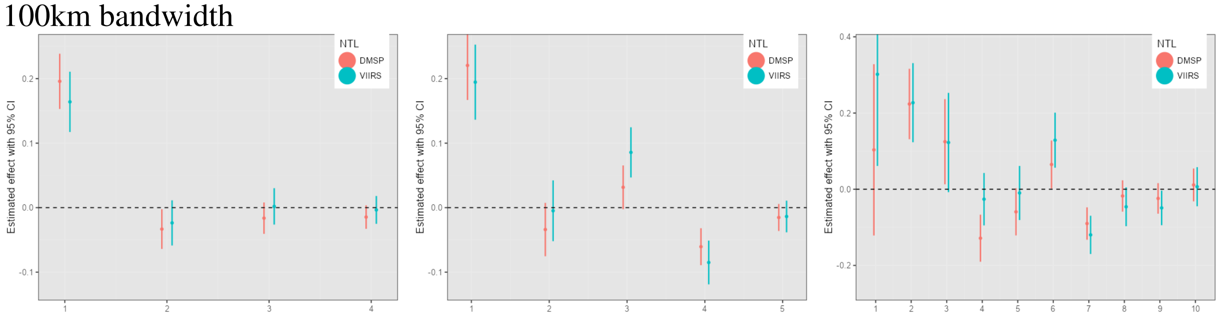


(2) Control of corruption

All



100km bandwidth



The bar indicates the 95% confidence intervals.

As an additional robustness check, we conducted separate RD analyses for pixels close to and far from the capital. Figure 7 illustrates the RD plots with the estimated coefficients for national institutions. The upper two panels show the results for the DMSP dataset and the lower two panels show the results for the VIIRS dataset. The left panels show the results for pixels close to the capital city, and the right panels show the results for pixels far from the capital city. In both datasets, we found no significant difference at the borders, regardless of the distance to the capital. The use of the continuous outcome $\ln(1 + NTL)$ does not change the results.[14] These results align with our earlier discussion that bordering areas are mostly rural or unpopulated, and that the associations between national institutions and NTL are driven by pixels distant from the border. This explains the small differences in the estimation results between DMSP and VIIRS data.

4 Conclusion

We revisited the empirical exploration of the importance of national institutions for sub-regional economic development in African countries by utilizing more recent and precise nighttime light data. While we found slightly different empirical patterns than in the original study by MP, the overall results support their core argument that national institutions matter for economic development, mainly for regions close to the national capital, reflecting the limited enforcement capacity of central governments in extending their policies to remote regions. While the spatial imprecision and blurring of the DMSP attenuated the magnitude of the association between national institutions and NTL, the fact that most NTL were detected far from the borders mitigated the sensitivity of MP's results to these issues.

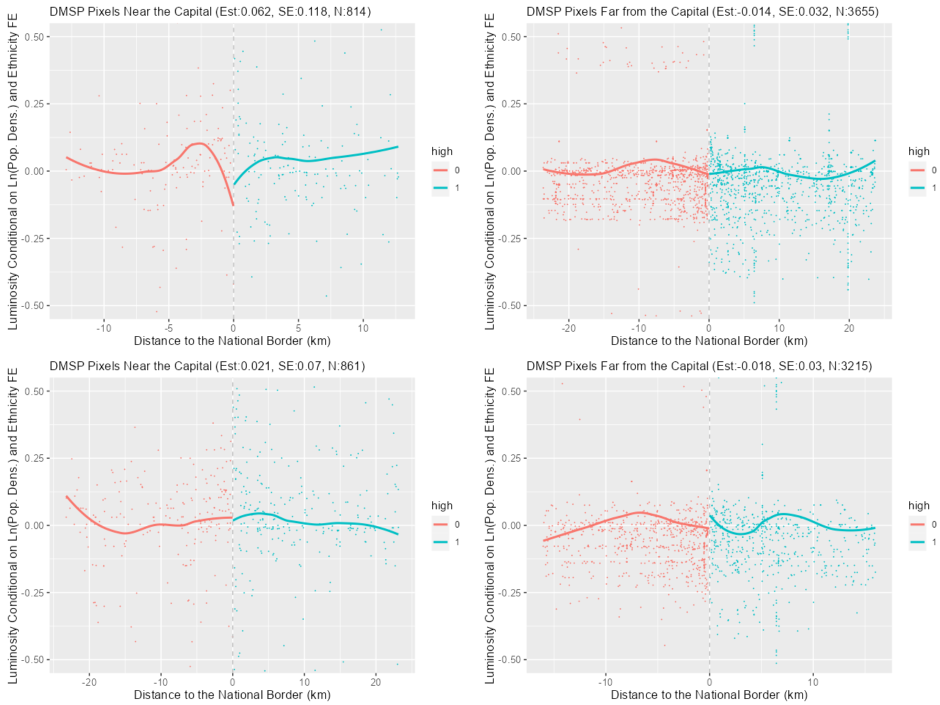
It should be noted that NTL may not be a good proxy for economic activity in rural areas and may not be suitable for studying the impact of national institutions in areas far from urban centers. To uncover a more accurate relationship between national institutions and sub-regional development, higher-quality data that can be generated by the application of machine learning techniques across a diverse dataset, encompassing NTL, daytime satellite images, transport connectivity, agricultural productivity, land use, and survey data is required. Enhanced data quality is imperative for conducting credible examinations of economic activity at the granular level.

[14] See Appendix Table 5 and Appendix Figure 6. Using the AAO data did not change the results, as shown in Appendix Table 6.

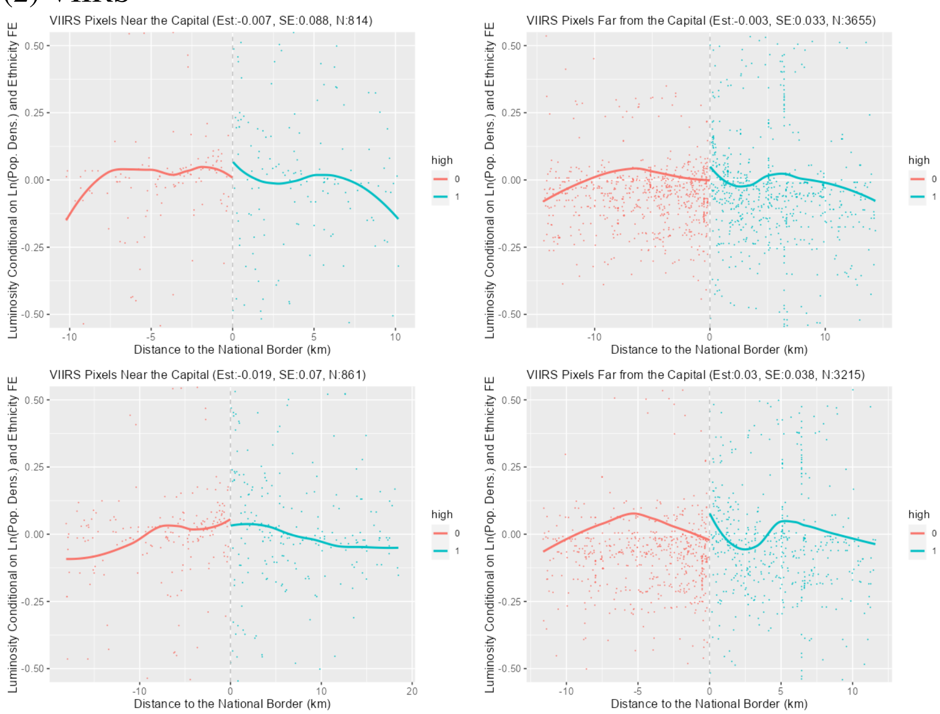
Figure 7: Regression discontinuity graphs by relative distance to the capital cities (binary NTL)

(A) Close to capital (B) Far from capital

(1) DMSP



(2) VIIRS



References

Abrahams, Alexei, Christopher Oram, and Nancy Lozano-Gracia, "Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination," *Remote Sensing of Environment*, 2018, *210*, 242–258.

Acemoglu, Daron, Simon Johnson, and James A Robinson, "The colonial origins of comparative development: An empirical investigation," *American economic review*, 2001, *91* (5), 1369–1401.

Allen, Treb and Costas Arkolakis, "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 2022, *89* (6), 2911–2957.

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484), 1481–1495.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 2006, *93* (3), 491–507.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural productivity and structural transformation: Evidence from Brazil," *American Economic Review*, 2016, *106* (6), 1320–1365.

Chen, Xi and William D Nordhaus, "VIIRS nighttime lights in the estimation of cross-sectional and time-series GDP," *Remote Sensing*, 2019, *11* (9), 1057.

_ **and William Nordhaus**, "A test of the new VIIRS lights data set: Population and economic output in Africa," *Remote Sensing*, 2015, *7* (4), 4937–4947.

Costinot, Arnaud, Dave Donaldson, and Cory Smith, "Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world," *Journal of Political Economy*, 2016, *124* (1), 205–248.

Donaldson, Dave and Adam Storeygard, "The view from above: Applications of satellite data in economics," *Journal of Economic Perspectives*, 2016, *30* (4), 171–198.

Eberhard-Ruiz, Andreas and Alexander Moradi, "Regional market integration in East Africa: Local but no regional effects?," *Journal of Development Economics*, 2019, *140*, 255–268.

Gelman, Andrew and Guido Imbens, "Why high-order polynomials should not be used in regression discontinuity designs," *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.

Gibson, John, "Better night lights data, for longer," *Oxford Bulletin of Economics and Statistics*, 2021, *83* (3), 770–791.

_ and Geua Boe-Gibson, "Nighttime lights and county-level economic activity in the United States: 2001 to 2019," *Remote Sensing*, 2021, *13* (14), 2741.

_ , Susan Olivia, Geua Boe-Gibson, and Chao Li, "Which night lights data should we use in economics, and where?," *Journal of Development Economics*, 2021, *149*, 102602.

Guiteras, Raymond, Amir Jina, and A Mushfiq Mobarak, "Satellites, self-reports, and submersion: Exposure to floods in Bangladesh," *American Economic Review*, 2015, *105* (5), 232–236.

Harari, Mariaflavia and Eliana La Ferrara, "Conflict, climate, and cells: A disaggregated analysis," *Review of Economics and Statistics*, 2018, *100* (4), 594–608.

Head, Keith and Thierry Mayer, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm, "The making of the modern metropolis: Evidence from London," *The Quarterly Journal of Economics*, 2020, *135* (4), 2059–2133.

Henderson, J Vernon, Adam Storeygard, and David N Weil, "Measuring economic growth from outer space," *American economic review*, 2012, *102* (2), 994–1028.

_ , Tim Squires, Adam Storeygard, and David Weil, "The global distribution of economic activity: Nature, history, and the role of trade," *The Quarterly Journal of Economics*, 2018, *133* (1), 357–406.

Hodler, Roland and Paul A Raschky, "Regional favoritism," *The Quarterly Journal of Economics*, 2014, *129* (2), 995–1033.

Ito, Koichiro and Shuang Zhang, "Willingness to pay for clean air: Evidence from air purifier markets in China," *Journal of Political Economy*, 2020, *128* (5), 1627–1672.

Jayachandran, Seema, "Air quality and early-life mortality: Evidence from Indonesia ' s wildfires," *Journal of Human resources*, 2009, *44* (4), 916–954.

Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, 2016, *353* (6301), 790–794.

Lujala, Paivi, Jan Ketil Rod, and Nadja Thieme, "Fighting over oil: Introducing a new dataset," *Conflict Management and Peace Science*, 2007, *24* (3), 239–256.

Michalopoulos, Stelios and Elias Papaioannou, "National institutions and subnational development in Africa," *The Quarterly journal of economics*, 2014, *129* (1), 151–213.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, migration, and local employment elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Perez-Sindin, Xaquin S, Tzu-Hsin Karen Chen, and Alexander V Prishchepov, "Are nighttime lights a good proxy of economic activity in rural areas in middle and low-income countries? Examining the empirical evidence from Colombia," *Remote Sensing Applications: Society and Environment*, 2021, *24*, 100647.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *Review of Economics and statistics*, 2006, *88* (4), 641–658.

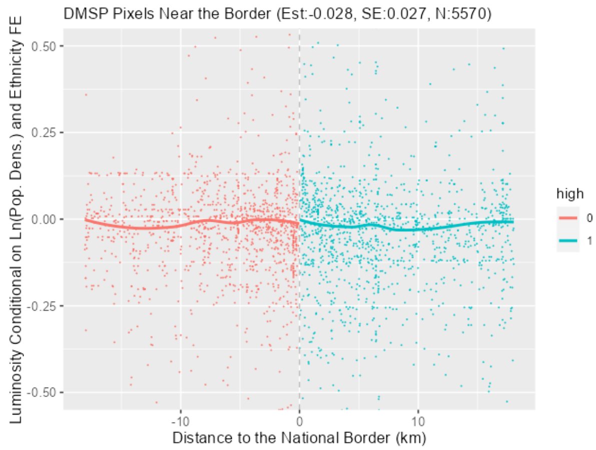
Storeygard, Adam, "Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa," *The Review of economic studies*, 2016, *83* (3), 1263–1295.

Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke, "Using publicly available satellite imagery and deep learning to understand economic well-being in Africa," *Nature communications*, 2020, *11* (1), 2583.

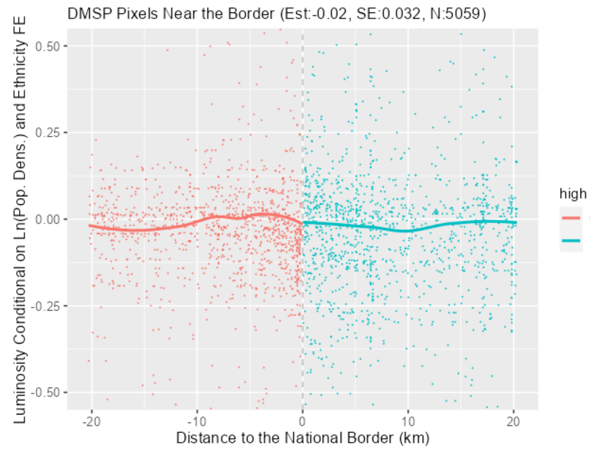
A Appendix Figures and Tables

Appendix Figure 1: Regression discontinuity: $\ln(1 + NTL)$, local quadratic

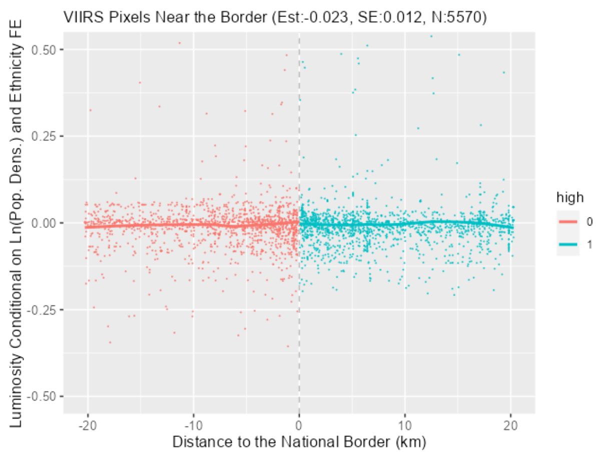
DMSP: Rule of law



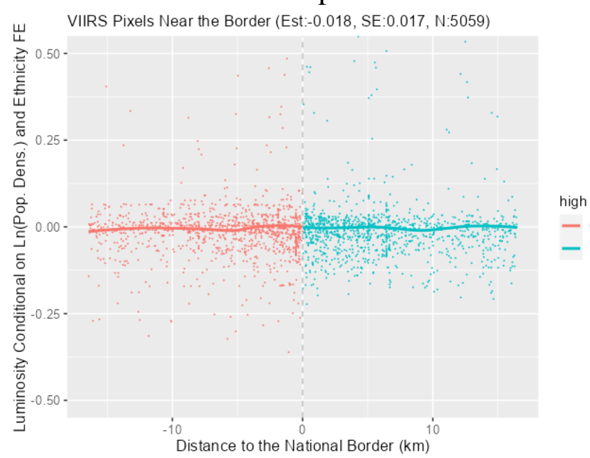
Control of corruption



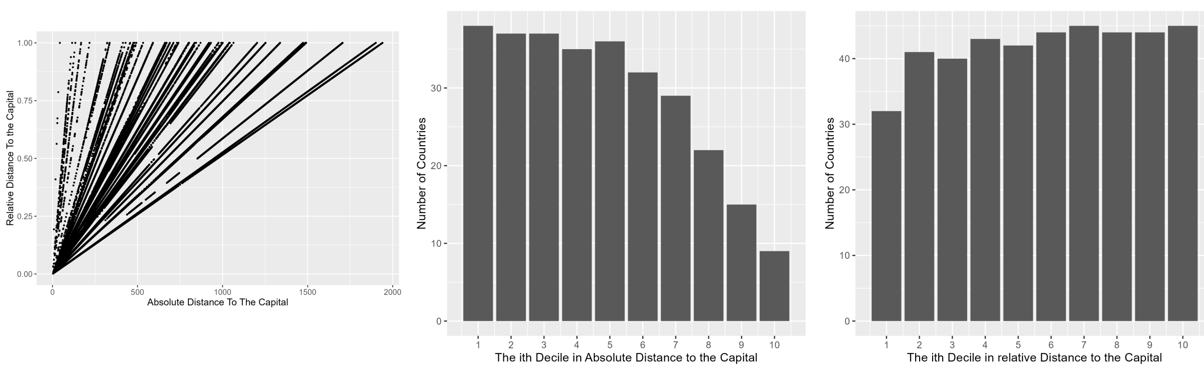
VIIRS: Rule of law



Control of corruption



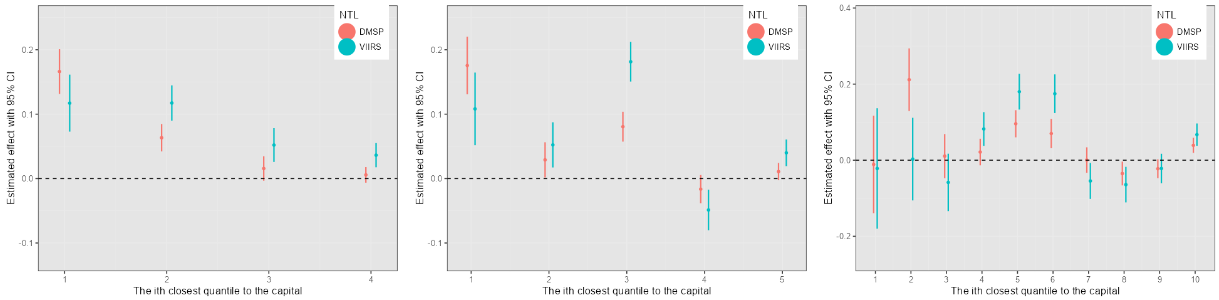
Appendix Figure 2: Absolute distance vs. relative distance



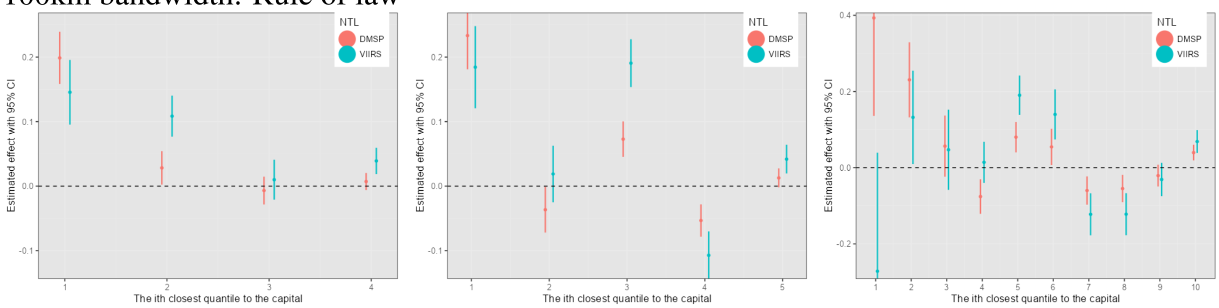
Appendix Figure 3: Heterogeneity by relative distance to the capital cities (binary NTL): AAO data

Quartiles Quintiles Deciles

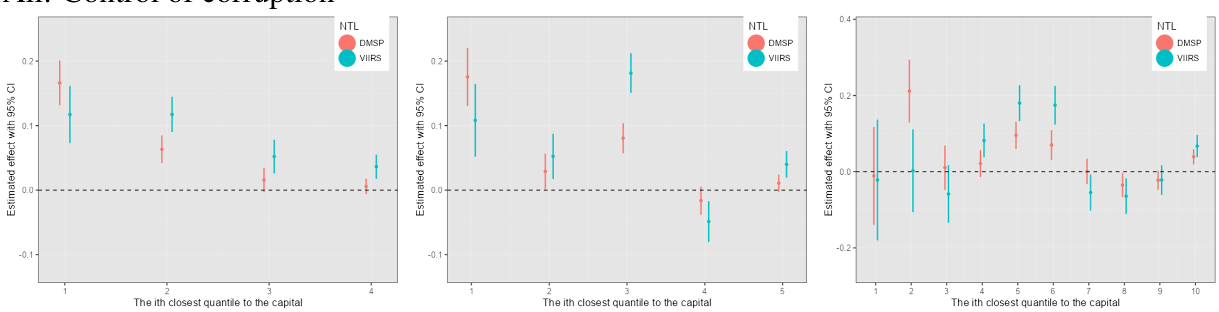
All: Rule of law



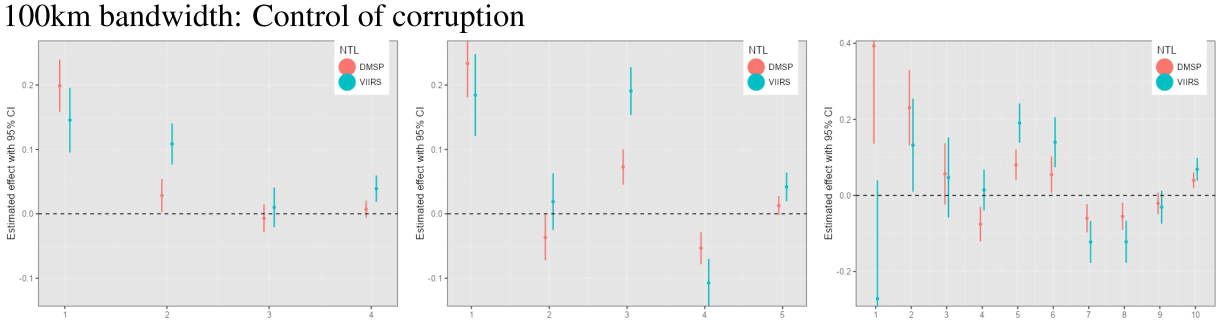
100km bandwidth: Rule of law



All: Control of corruption



100km bandwidth: Control of corruption

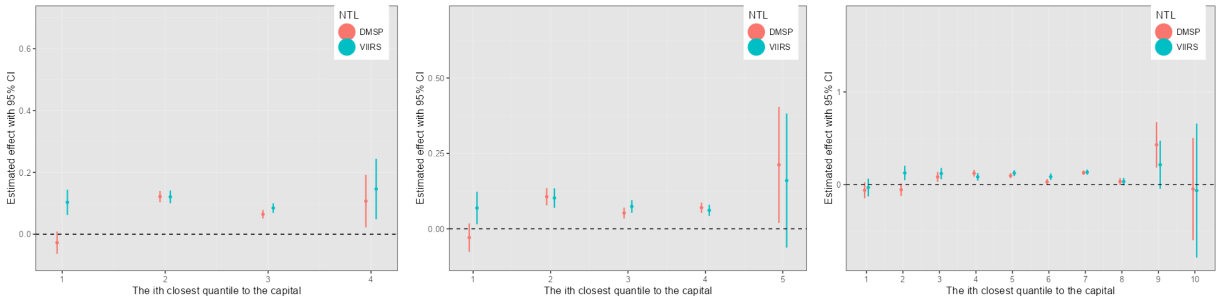


The bar indicates the 95% confidence intervals.

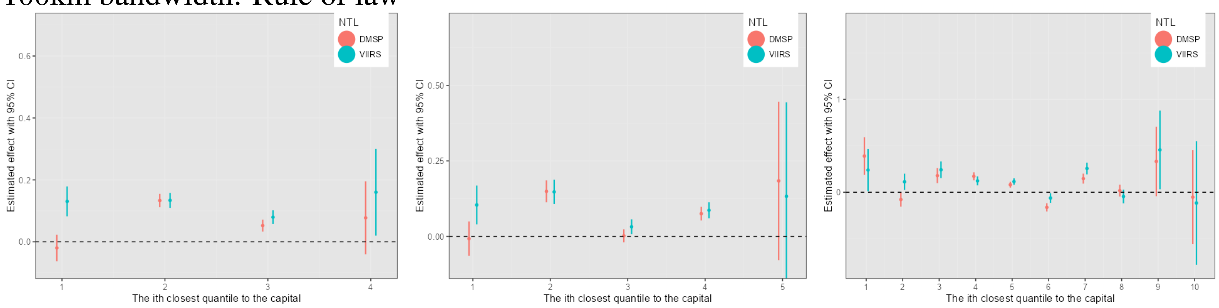
Appendix Figure 4: Heterogeneity by absolute distance to the capital cities (binary NTL)

Quartiles Quintiles Deciles

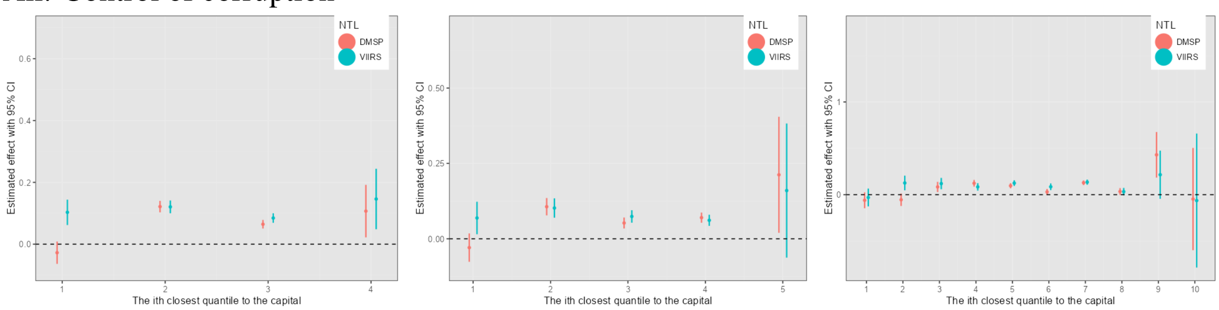
All: Rule of law



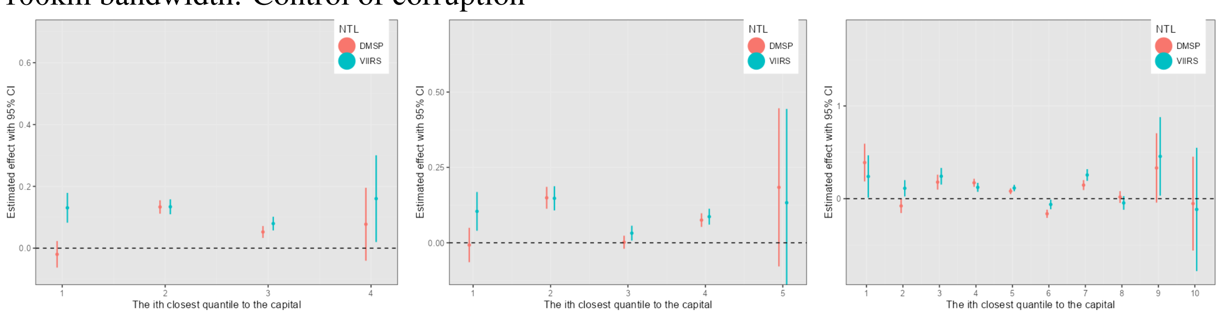
100km bandwidth: Rule of law



All: Control of corruption



100km bandwidth: Control of corruption

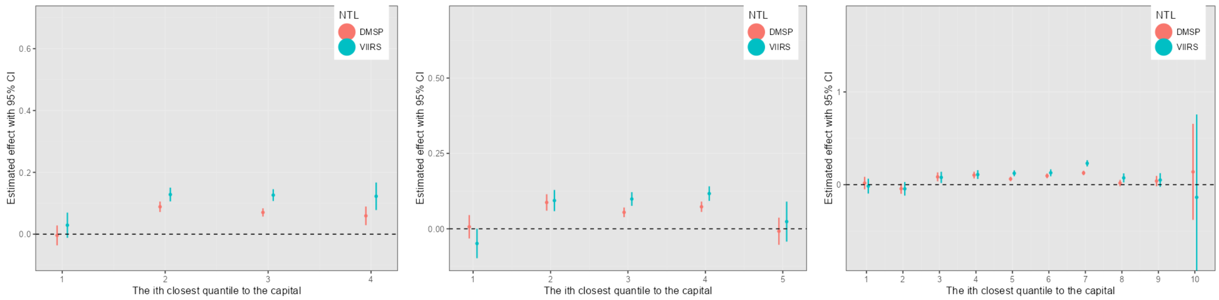


The bar indicates the 95% confidence intervals.

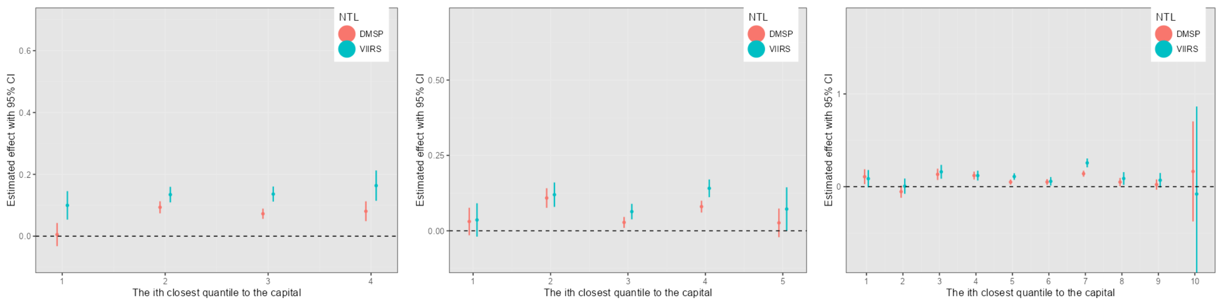
Appendix Figure 5: Heterogeneity by absolute distance to the capital cities (binary NTL): AAO data

Quartiles Quintiles Deciles

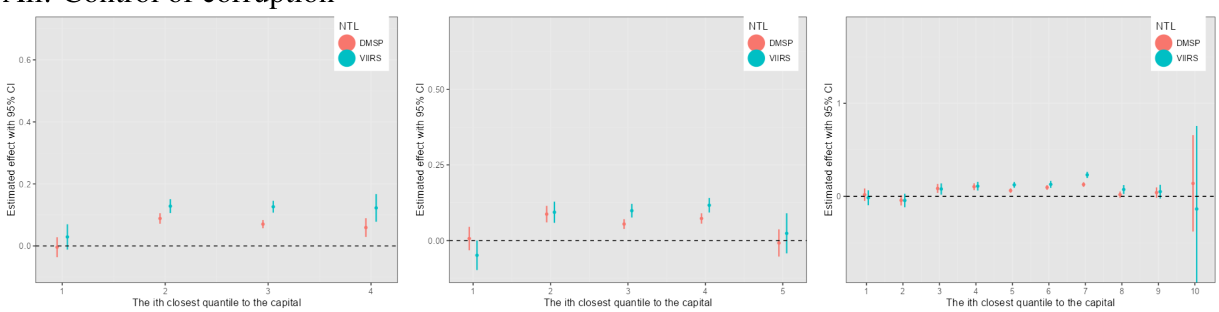
All: Rule of law



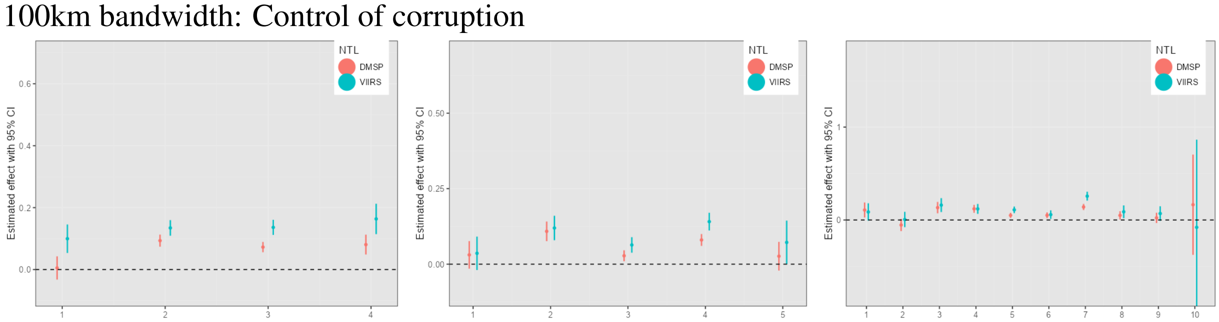
100km bandwidth: Rule of law



All: Control of corruption



100km bandwidth: Control of corruption

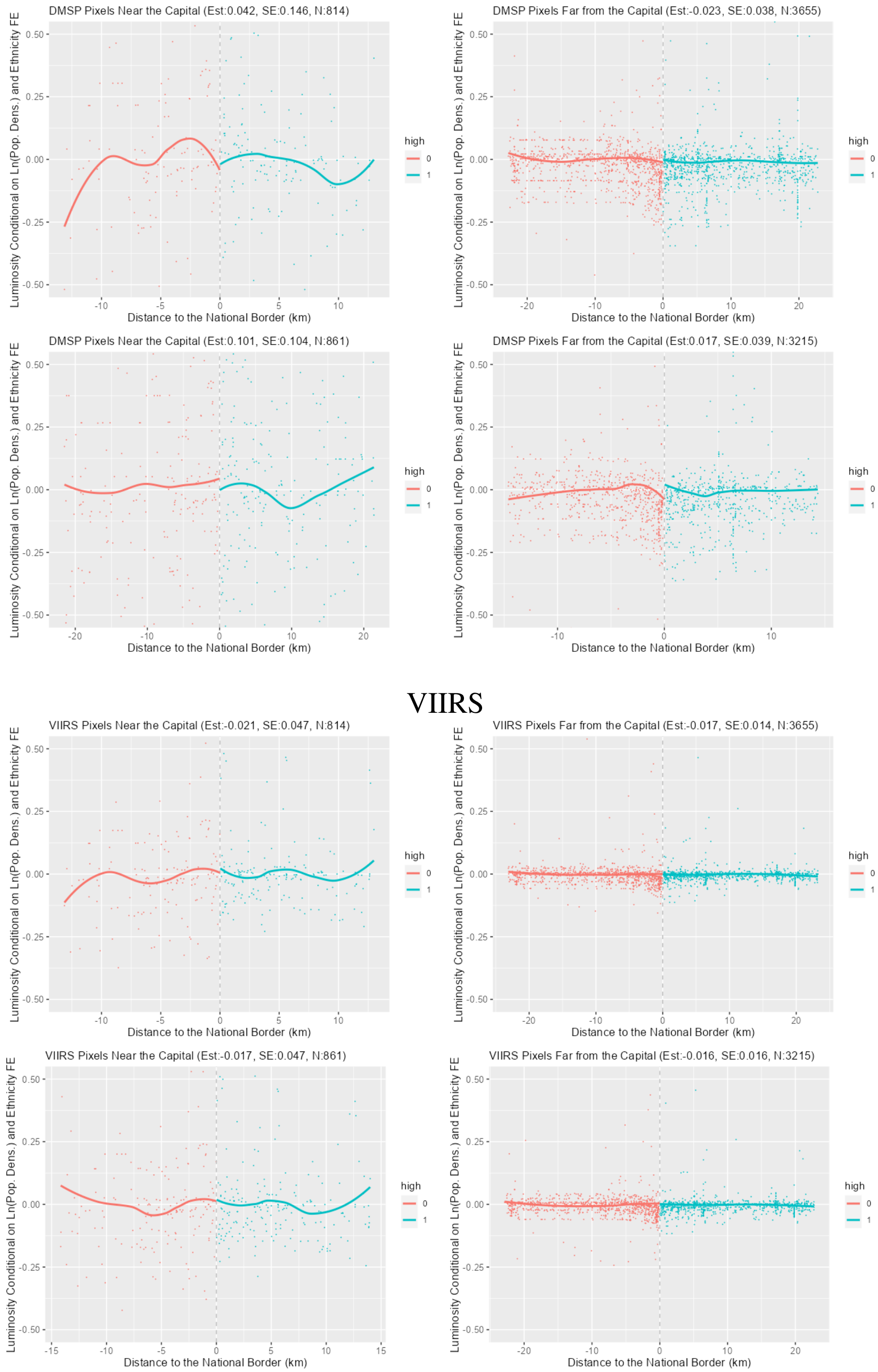


The bar indicates the 95% confidence intervals.

Appendix Figure 6: Regression Discontinuity by distance to the capital cities (continuous outcome)

Close to capital Far from capital

DMSP



Appendix Table 1: Summary Statistics at pixel levels using AAO data

Statistic	N	Mean	St. Dev.	Min	Max
Light density (VIIRS2014)	48,733	0.046	0.663	0.000	70.018
Light density (DMSP2007)	48,733	0.228	1.798	0.000	62.825
Light density (DMSP2008)	48,733	0.215	1.747	0.000	61.919
Ln (0.01+Light density) (VIIRS2014)	48,733	−4.339	0.816	−4.605	4.249
Ln (0.01+Light density) (DMSP2007)	48,733	−4.223	1.312	−4.605	4.141
Ln (0.01+Light density) (DMSP2008)	48,733	−4.235	1.284	−4.605	4.126
Ln (0.01+Population density)	48,733	1.293	2.193	−4.605	9.044
Distance to the capital city	48,733	618.657	387.778	1.371	1,935.361
Distance to the sea coast	48,733	658.511	442.998	0.045	1,810.693
Distance to the border	48,733	75.671	73.092	0.001	448.213
Rule of law	48,733	−0.868	0.623	−2.197	0.615
Control of corruption	48,733	−0.746	0.558	−1.664	0.814

Appendix Table 2: DMSP 2013 data

(A) Country-ethnic homeland level analysis

Night time light	Rule of law				Corrupt			
Institutional quality	0.702	0.257	0.649	0.393	0.925	0.411	0.802	0.530
Double-clustered std. err	(0.188)	(0.203)	(0.177)	(0.215)	(0.210)	(0.221)	(0.193)	(0.232)
Observations	507	507	507	507	507	507	507	507
R2 Adj.	0.306	0.661	0.422	0.672	0.324	0.664	0.429	0.674
Controls								
Ethnicity fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location controls	No	No	Yes	Yes	No	No	Yes	Yes
Geographic controls	No	No	Yes	Yes	No	No	Yes	Yes

(B) Pixel level analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Night time light	Rule of law				Corrupt			
Institutional quality	0.097	0.043	0.073	0.037	0.123	0.060	0.096	0.045
Double-clustered std. err	(0.034)	(0.021)	(0.028)	(0.021)	(0.040)	(0.026)	(0.035)	(0.027)
Observations	49832	49832	49832	49832	49832	49832	49832	49832
R2 Adj.	0.114	0.289	0.167	0.297	0.124	0.290	0.173	0.297
Controls								
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

Appendix Table 3: Pixel level data AAO (binary outcomes)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Night time light	DMSP				VIIRS			
Rule of law	0.086	0.042	0.072	0.041	0.184	0.087	0.141	0.072
Double-clustered std. err	(0.031)	(0.021)	(0.028)	(0.022)	(0.046)	(0.036)	(0.038)	(0.031)
Observations	48733	48733	48733	48733	48733	48733	48733	48733
Adjusted R-squared	0.120	0.313	0.171	0.320	0.212	0.370	0.253	0.380
Controls								
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Night time light	DMSP				VIIRS			
Corrupt	0.109	0.058	0.093	0.051	0.215	0.116	0.165	0.090
Double-clustered std. err	(0.038)	(0.027)	(0.035)	(0.028)	(0.056)	(0.045)	(0.048)	(0.040)
Observations	48733	48733	48733	48733	48733	48733	48733	48733
Adjusted R-squared	0.129	0.313	0.177	0.320	0.218	0.370	0.255	0.379
Controls								
Ethnicity FE	No	Yes	No	Yes	No	Yes	No	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location&geography	No	No	Yes	Yes	No	No	Yes	Yes

The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

Appendix Table 4: DMSP 2013 RD polynomial

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Data	Rule of Law						Corrupt					
Band Width	All pixels		100km		50km		All pixels		100km		50km	
RD polynominal	3rd	4th	3rd	4th	3rd	4th	3rd	4th	3rd	4th	3rd	4th
Institutional quality	0.018	0.011	-0.009	-0.002	0.002	-0.027	0.010	-0.004	-0.014	-0.010	-0.013	-0.032
Double-clustered std. err	(0.015)	(0.016)	(0.018)	(0.020)	(0.018)	(0.020)	(0.016)	(0.017)	(0.014)	(0.015)	(0.015)	(0.017)
Observations	49710	49710	21145	21145	13390	13390	49710	49710	21145	21145	13390	13390
Adjusted R-squared	0.280	0.280	0.269	0.269	0.285	0.285	0.281	0.281	0.269	0.269	0.284	0.284

The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the ethnic homeland fixed effects, logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

Appendix Table 5: Heterogeneity in association due to distance to the capital ($\ln(1 + NTL)$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(1 + NTL)	DMSP		VIIRS		DMSP		VIIRS	
All	Absolute distance to the capital				Relative distance to the capital			
	Close	Far	Close	Far	Close	Far	Close	Far
Rule of law	0.083	0.037	0.009	0.009	0.170	0.006	0.051	-0.002
Double-clustered std. err	(0.135)	(0.171)	(0.699)	(0.225)	(0.002)	(0.574)	(0.002)	(0.614)
Observations	24916	24916	24916	24916	24904	24928	24904	24928
Adjusted R-squared	0.430	0.099	0.227	0.029	0.402	0.274	0.187	0.092

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(1 + NTL)	DMSP		VIIRS		DMSP		VIIRS	
100km bandwidth	Absolute distance to the capital				Relative distance to the capital			
	Close	Far	Close	Far	Close	Far	Close	Far
Rule of law	0.101	0.016	0.024	0.004	0.073	0.002	0.015	-0.002
Double-clustered std. err	(0.018)	(0.443)	(0.117)	(0.416)	(0.046)	(0.880)	(0.420)	(0.624)
Observations	10630	10631	10630	10631	10618	10643	10618	10643
Adjusted R-squared	0.387	0.147	0.204	0.045	0.363	0.235	0.167	0.094

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(1 + NTL)	DMSP		VIIRS		DMSP		VIIRS	
All	Absolute distance to the capital				Relative distance to the capital			
	Close	Far	Close	Far	Close	Far	Close	Far
Control of Corruption	0.110	0.050	0.005	0.014	0.231	0.014	0.060	0.000
Double-clustered std. err	(0.205)	(0.089)	(0.897)	(0.106)	(0.005)	(0.412)	(0.013)	(0.984)
Observations	24916	24916	24916	24916	24904	24928	24904	24928
Adjusted R-squared	0.431	0.099	0.227	0.029	0.402	0.274	0.187	0.092

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(1 + NTL)	DMSP		VIIRS		DMSP		VIIRS	
100km bandwidth	Absolute distance to the capital				Relative distance to the capital			
	Close	Far	Close	Far	Close	Far	Close	Far
Control of Corruption	0.158	0.028	0.034	0.007	0.129	0.009	0.024	0.000
Double-clustered std. err	(0.002)	(0.187)	(0.116)	(0.219)	(0.012)	(0.577)	(0.338)	(0.987)
Observations	10630	10631	10630	10631	10618	10643	10618	10643
Adjusted R-squared	0.390	0.148	0.204	0.045	0.364	0.235	0.168	0.094

The table reports the estimated coefficients of the regression, with standard errors doubly-clustered at the country and ethnic homeland dimensions in parentheses. The control variables include the ethnic homeland fixed effects, logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.

Appendix Table 6: Heterogeneity in association due to distance to the capital (binary NTL): AAO data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DMSP		VIIRS		DMSP		VIIRS	
	Absolute distance to the capital				Relative distance to the capital			
All	Close	Far	Close	Far	Close	Far	Close	Far
Rule of law	0.058	0.046	0.086	0.082	0.094	0.019	0.096	0.041
p-value	(0.161)	(0.110)	(0.008)	(0.004)	(0.030)	(0.269)	(0.021)	(0.079)
FDR q-value	[0.138]	[0.571]	[0.017]	[0.017]	[0.131]	[0.571]	[0.028]	[0.042]
Observations	24916	24916	24916	24916	24904	24928	24904	24928
Adjusted R-squared	0.389	0.124	0.412	0.164	0.363	0.261	0.402	0.277
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.729		0.745		0.107		0.255	
100km bandwidth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rule of law	0.076	0.015	0.125	0.058	0.034	0.012	0.056	0.052
p-value	(0.165)	(0.529)	(0.004)	(0.032)	(0.352)	(0.520)	(0.118)	(0.016)
FDR q-value	[0.138]	[0.710]	[0.014]	[0.034]	[0.241]	[0.710]	[0.054]	[0.025]
Observations	10630	10631	10630	10631	10618	10643	10618	10643
Adjusted R-squared	0.384	0.181	0.429	0.230	0.383	0.244	0.418	0.297
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.229		0.126		0.587		0.924	
All	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corruption	0.078	0.066	0.099	0.118	0.120	0.032	0.111	0.065
p-value	(0.170)	(0.042)	(0.038)	(0.004)	(0.031)	(0.180)	(0.035)	(0.048)
FDR q-value	[0.138]	[0.511]	[0.033]	[0.017]	[0.131]	[0.571]	[0.033]	[0.034]
Observations	24916	24916	24916	24916	24904	24928	24904	24928
Adjusted R-squared	0.389	0.125	0.412	0.165	0.362	0.261	0.401	0.278
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.630		0.803		0.145		0.456	
100km bandwidth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corruption	0.124	0.030	0.158	0.077	0.066	0.026	0.070	0.071
p-value	(0.043)	(0.237)	(0.002)	(0.028)	(0.131)	(0.311)	(0.106)	(0.013)
FDR q-value	[0.131]	[0.571]	[0.013]	[0.034]	[0.138]	[0.571]	[0.054]	[0.025]
Observations	10630	10631	10630	10631	10618	10643	10618	10643
Adjusted R-squared	0.386	0.182	0.429	0.231	0.384	0.245	0.418	0.298
$H_0 : \gamma_{Close} = \gamma_{Far}$	0.074		0.080		0.424		0.985	

The table reports the estimated coefficients of the regression, with p values based on the standard errors clustered at the country and ethnic homeland in parentheses. The control variables include the ethnic homeland fixed effects, logarithm of population density, logarithm of the pixel area, and the location and geography variables such as the distance to the boundary, distance to the capital city, distance to the sea coast, logarithm of the water area, pixel means of land suitability for agriculture, malaria suitability index, elevation, and indicators for diamond mine and petroleum.