

Illuminating the Global South?*

Giorgio Chiovelli[†]
Universidad de Montevideo

Stelios Michalopoulos[‡]
Brown University, CEPR and NBER

Elias Papaioannou[§]
London Business School, CEPR

Tanner Regan[¶]
George Washington University

June 30, 2025

Abstract

Satellite images of nighttime lights are commonly used to proxy local economic conditions. Despite their popularity, there are concerns about how accurately they capture local development in different settings and scales. We compile an annual series of comparable nighttime lights globally from 1992 to 2023 by applying adjustments that consider key factors affecting accuracy and comparability over time: top coding, blooming, and variations in satellite systems (DMSP and VIIRS). Applied to various low-income settings, the adjusted luminosity series outperforms the unadjusted series as a predictor of local development, particularly over time and at higher spatial resolutions.

KEYWORDS: Night Lights, Economic Development, Measurement, Africa.

JEL CLASSIFICATION: O1, R1, E01, I32

*We thank Gonzalo Ferrés, Federico Ferro, Mallory Perillo, Juan Pablo Tizon, and Camila Varela for their superb research assistance. We are grateful to four anonymous referees and Sascha O. Becker for their valuable feedback and comments. We thank Sam Bazzi, Roland Hodler, and Paul Raschky for sharing their data, and Sam Asher and Paul Novosad for making their granular data on development across rural and urban India publicly available. All errors are our responsibility.

[†]Giorgio Chiovelli. Universidad de Montevideo, Department of Economics, Prudencio de Pena 2440, Montevideo, 11600, Uruguay; gchiovelli@um.edu.uy. Web: <https://sites.google.com/view/giorgiochiovelli/>

[‡]Stelios Michalopoulos. Brown University, Department of Economics, 64 Waterman Street, Robinson Hall, Providence, RI, 02912, United States; smichalo@brown.edu. Web: <https://sites.google.com/site/stelioecon/>

[§]Elias Papaioannou. London Business School, Economics Department, Regent's Park. London NW1 4SA. United Kingdom; eliasp@london.edu. Web: <https://sites.google.com/site/papaioannouelias/home>

[¶]Tanner Regan. George Washington University, Department of Economics, Monroe Hall 2115 G Street NW, Washington, DC, 20052. United States; tanner.regan@gwu.edu. Web: <https://sites.google.com/site/tannerregan>

1 Introduction

A considerable literature in economics, political science, and remote sensing employs satellite nighttime lights (luminosity) to proxy development (Donaldson and Storeygard, 2016; Levin *et al.*, 2020).¹ The use of satellite data appears *a priori* helpful for low and middle-income countries with weak state capacity and recurrent conflict. Earlier research reveals that luminosity is a valuable proxy for cross-country GDP (Henderson *et al.*, 2012; Chen and Nordhaus, 2011); hence, researchers use luminosity to correct for inconsistencies and noise in country-level statistics stemming from challenges measuring output in economies specializing in agriculture, with a large informal economy, and underfunded statistical agencies (Pinkovskiy and Sala-i Martin, 2016). Luminosity appears helpful in correcting inflated output statistics that non-democratic governments often produce (Martinez, 2022) and quantifying profit shifting by multinationals (Bilicka and Seidel, 2022). Increasingly, applied research uses luminosity to measure regional well-being, as data unavailability and error-in-variables are more serious at the local level. Many papers use luminosity to proxy development across administrative units (Hodler and Raschky, 2014; Alesina *et al.*, 2016), cities (Storeygard, 2016), historical ethnic homelands (Michalopoulos and Papaioannou, 2013, 2014), and grid-squares (Henderson *et al.*, 2018).²

However, there are still open issues regarding how accurately luminosity predicts development. First, while some studies validate the correlation between luminosity and development measures, there are concerns about the strength of the association in broader samples. Second, researchers have leeway on the validation, for example, selecting the development proxy and the spatial units, which sometimes are large areas and, in other cases, small grid squares. Third, while some papers use unadjusted series (e.g., Michalopoulos and Papaioannou, 2013), others adjust for top coding and the tendency of light to spill to neighboring areas (e.g., Henderson *et al.*, 2018). Fourth, mapping the lower resolution and coarser pre-2013 series (DMSP-OLS) to the post-2013 (VIIRS) series is not straightforward. Most papers abstain from dealing with this issue using either the pre- or the post-2013 series. Moreover, as most studies are cross-sectional, it is still unclear how strongly changes in luminosity reflect changes in local well-being over time. Consequently, despite the widespread use of luminosity data in economics and political science, its effectiveness in capturing development in low-income settings remains ambiguous, particularly concerning the scale, location, and period of

¹This literature comprises papers on Africa (Hjort and Poulsen, 2019; Storeygard, 2016; Michalopoulos and Papaioannou, 2014; Dreher *et al.*, 2019; Henderson *et al.*, 2017), Asia (Harari, 2020; Chodorow-Reich *et al.*, 2020; Baum-Snow *et al.*, 2017), Europe (Gibson, 2021), North America (Bleakley and Lin, 2012), and worldwide (Ch *et al.*, 2021; Henderson *et al.*, 2018; Pinkovskiy, 2017; Martinez, 2022).

²A related research stream combines light density with daytime imagery and other data to better proxy local activity (Jean *et al.*, 2016; Yeh *et al.*, 2020; Khachiyan *et al.*, 2022; Rossi-Hansberg and Zhang, 2025). However, while the methods are usually made public, daytime input data is typically proprietary. Besides, even when daytime data are freely available (e.g., landsat), replicating the methods is prohibitive for typical economic research projects. Future work could use and blend the data produced here with local wealth estimates from daytime imagery.

analysis.

Recent papers paint a somewhat conflicting picture. On the one hand, India-based tabulations suggest that lights proxy well regional economic conditions in levels and changes (Asher *et al.*, 2021). On the other hand, studies from Indonesia, China, and South Africa suggest that variation of GDP does not correlate strongly with lights outside cities, at least for the early period, 1992 – 2013 (Gibson *et al.*, 2021). Some other papers reveal strong cross-sectional but weak panel associations (Chen *et al.*, 2024).

In this paper, we build a global dataset of luminosity from 1992 to 2023, and then explore the association between the newly compiled series and various proxies of local development. For our validation analysis, we focus on Africa and other low-income regions, where the output data quality is poor, and there is relatively limited regional data on economic activity, especially at granular levels.³ As research has moved from cross-country designs to *meso* approaches that exploit variation within-country across regions (administrative units, ethnic homelands, etc.) or grid squares of various sizes, we also explore the luminosity-development nexus across areas of varying spatial resolution.

Our first contribution is to create a standardized panel of global nightlights over three decades (1992 – 2023), integrating the pre-2013 (DMSP) and post-2013 (VIIRS) series after various adjustments to reduce measurement error.⁴ While the VIIRS data series offers various improvements (Gibson *et al.*, 2021), research often requires the longest possible time series to study economic phenomena in the developing world, for example, the impact of democratic transitions, trade-induced shocks, and financial liberalization. Adjusting and integrating the old luminosity series with the newer ones can allow exploring a wider range of questions that exploit not only cross-sectional but also time series variation. While other papers have made adjustments to specific aspects of nighttime lights data (Cao *et al.*, 2019; Nechaev *et al.*, 2021; Bluhm and Krause, 2022), we apply blooming and topcoding corrections to the DMSP series and extend the series by fusing it with the VIIRS data.

Our second contribution is to validate the newly constructed luminosity series as a proxy for local development in low-income environments, zooming in on the spatial dimension of aggregation and the granularity of the analysis. We compare the performance of the unadjusted luminosity series with our adjusted and harmonized one, examining their correlation with education, household wealth, electrification, and other development proxies in various settings: across gridcells of different sizes using 139 georeferenced DHS surveys from 34 African countries; tabulating all Mozambican censuses, across coarse and fine administrative units; using Indonesian surveys spanning thousands

³As more data become available, such as urban structures from daytime satellite images and mobile telephony and large-sample surveys are used more, researchers will increasingly have more tools to capture local development.

⁴The replication code for the dataset is available on GitHub at github.com/tannerregan/world_nightlights. This open-source code allows researchers to customize the production of adjusted nightlights series, and the GitHub page also includes a link to download the global dataset directly.

of villages with multiple development outcomes, and using rich surveys across Indian villages and towns. The analysis of these data, spanning the DMSP and the VIIRS periods, reveals four main takeaways. First, the new luminosity series correlates strongly with local development in the cross-section and over time. The correlation is, however, far from perfect. As lights are very skewed, the binary (lit/unlit) transformation (stemming from the many zeros), which empirical studies often take, cannot perfectly capture the significant variation in regional assets, education, and public goods. Nonetheless, luminosity correlates significantly with numerous local well-being proxies. Second, the correlation is stronger, especially in panel estimation, with the adjusted and harmonized series, telling of a reduction in measurement error from our adjustments of the noisier DMSP series and their merging with the higher quality VIIRS. Third, the luminosity-development association retains significance across spatial units of various sizes, even when exploiting localized variation. Fourth, the adjustments yield stronger correlations between luminosity and local development at higher levels of spatial resolution, and especially when comparing small regions and gridcells. In contrast, across coarse administrative units or large grid squares, the adjustments yield minor improvements in the strength of the luminosity-development link. So, as empirical papers are increasingly using local identification schemes, such as spatial Regression Discontinuity Designs (RDD), researchers should consider our adjustments.

Structure Section 2 details our methodology for correcting the DMSP-OLS series and merging it with the downgraded VIIRS series. Then, we explore the cross-sectional and over time association of luminosity with development across countries (Section 3), gridcells in Africa (Section 4), and administrative units in Mozambique, Indonesia, and India (Section 5). In Section 6, we replicate two pan-African studies and one global study, comparing the initial results with the unadjusted series to the ones with the adjusted and merged nighttime lights series. Section 7 concludes, discussing avenues for future research.

2 Methods and New Series

This section presents the data and our approach to compiling a new global time series of nighttime lights at a pixel level from 1992 to 2023. Unless otherwise noted, by ‘pixels’ we refer to the source spatial resolution of the DMSP data: equally spaced by 30-arc seconds, equivalent to a $926m \times 926m$ grid at the equator (the distance in meters shrinks towards the poles). First, we discuss the adjustments to the DMSP series. Second, we present the machine learning (ML) approach that converts and merges the new VIIRS series to the *adjusted* DMSP.

2.1 Adjusted DMSP Series

The DMSP data from 1992 – 2013 has three deficiencies, which are by now well understood: cross-sensor calibration, top coding, and blooming.⁵ As earlier studies have addressed these issues individually, we discuss them only briefly.

Cross-sensor Accuracy As the DMSP series comes from six satellites, luminosity readings vary. Sometimes, values differ even for the same satellite due to the sensor’s degradation. A long-standing remote sensing literature has addressed this problem. Li *et al.* (2020), for instance, provide a downloadable DMSP series, which integrates the cross-sensor correction from Li and Zhou (2017). Using data from sensors with overlapping or nearby years, these studies estimate a second-order polynomial to map values across sensors. In this paper, we refer to the Li *et al.* (2020) data as the ‘unadjusted’ series and compare it to the adjustments we make in the following steps.

Top Coding The DMSP data are 8-bit integers, Digital Numbers (DN), ranging from 0 to 63. This limits stored information. In addition, as sensors are calibrated to detect clouds they miss brighter lights, so a DN of 63 corresponds to a range of actual radiance. DMSP pixels with DNs in the mid-50s also suffer from ‘implicit’ top coding, as they are averages of multiple potentially top-coded values (Bluhm and Krause, 2022).⁶ Alternatively, a ‘radiance-calibrated’ (RC) vintage of DMSP, which is not top coded, is available for only a sub-sample of seven years (Hsu *et al.*, 2015). We adjust the series by combining the DMSP with the RC data using the approach of Bluhm and Krause (2022). First, for each year, we identify pixels (number N_t) with $DN \geq 55$ for replacement. Second, we rank the N_t pixels using the RC series from the nearest year. Third, we generate “structural values” from a truncated Pareto distribution: $f(x) = \frac{\alpha L^\alpha x^{-\alpha-1}}{1-(L/H)^\alpha}$.⁷ Fourth, we replace the N_t top-coded pixels so that the pixel with the i -th highest rank is replaced by the i -th highest “structural value”.

Blooming Due to weak spatial accuracy, the DMSP data suffer from blooming (sometimes called ‘blurring’ or ‘bleeding’).⁸ The sensor records a window where central pixels cover less space on the

⁵Another issue is ‘bottom coding’ of low light areas. We are unaware of an approach to adjust for this. However, as discussed below, our corrections substantially strengthen the light-development elasticities even in rural African regions with very low light levels.

⁶In practice, many developing countries are not particularly susceptible to topcoding issues. For instance, most DMSP pixels in Africa are completely unlit: 98.4% in 1992, 97.4% in 2002, and 96.8% in 2012, and even among lit pixels the share of top-coded ones ($DN = 63$) is tiny: 0.98% in 1992, 1.4% in 2002, and 1.7% in 2012, and the share of lit pixels with values close to top-coding ($DN \geq 55$) is 2.8% in 1992, 5.6% in 2002, and 10% in 2012.

⁷We use the parameter values from Bluhm and Krause (2022): $\alpha = 1.5$, lower bound threshold $L = 55$, and upper bound threshold $H = 2000$.

⁸Some papers suggested that blooming may be more important close to the sea and lakes, but Gibson *et al.* (2021) show that this is not the case. Nevertheless, we allow the blooming correction to differ flexibly in coastal areas and across broad global regions.

ground than pixels along the edges, ‘stretching out’ the edge pixels (Gibson *et al.*, 2021). The sensor may also be displaced up to 3km. We follow Cao *et al.* (2019), who model blooming as spatial spillovers and remove them. The background light is identified based on all lit pixels that neighbor at least one unlit pixel; pseudo-light pixels (PLPs). The method works as follows: First, identify PLPs, lit pixels ($DN > 0$) with one of their neighboring pixels dark ($DN = 0$). Second, for each PLP take the inverse squared-distance weighted sum of light in a surrounding 7×7 window, excluding pixels with less light.⁹ Third, run an OLS regression of the PLP light on the sum of its neighbors’ lights: $DN_p = \alpha + \beta \sum_q \frac{DN_q}{d_{q,p}^2}$.¹⁰ Fourth, for lit pixels, remove blooming by subtracting the model-predicted lights; replace pixels’ light with $DN'_p = DN_p - \hat{\alpha} - \hat{\beta} \sum_q \frac{DN_q}{d_{q,p}^2}$. Fifth, smooth each pixel’s value with the mean of it and its eight nearest neighbors, then set all negative values to zero. The blooming correction increases the share of unlit pixels, which is already large, as it removes light that spills over from nearby, more lit pixels. The global share of unlit pixels in 1992 rises from 92% to 95% and 2012 from 88% to 91%. However, as shown below, the adjusted series correlates more strongly with local development despite the higher share of unlit pixels; this is especially true across small units and pixels and when exploring granular variation.

2.2 Harmonizing the DMSP and VIIRS Series

While VIIRS does not suffer from top coding, blooming, and sensor degradation, it becomes available only after 2013 thus limiting the scope of research questions. The VIIRS series (we use the VIIRS VNL V2 series, Elvidge *et al.* (2021)) is not directly comparable to the DMSP. First, VIIRS records 14-bit DN, allowing for a wider range and more distinct luminosity values than DMSP. Second, VIIRS is recorded at a finer spatial resolution than DMSP (15 vs. 30 arc seconds). Third, the quality of the sensors is (much) better. Consequently, almost all studies rely on one of the two series. We develop a method to integrate the VIIRS with the DMSP series, creating an uninterrupted panel of comparable nightlights over three decades. Since the VIIRS data does not suffer from these issues, our preferred series downgrades VIIRS to match the DMSP series. We merge VIIRS to four versions of DMSP (unadjusted, blooming only, topcoding only, blooming and topcoding), and, hence, we end up with four extended series over 1992-2023.

Related Literature Parallel studies also “downgrade” the VIIRS series to make it comparable to previous data (Li *et al.*, 2020; Nechaev *et al.*, 2021). Li *et al.* (2020) use a sigmoid function calibrated in overlapping DMSP-VIIRS years. But, as the authors acknowledge, their series are comparable *only* for pixels with high light values, which is not that relevant for the low luminosity

⁹The 7×7 window follows the approach of Cao *et al.* (2019) and works from the assumption that blooming does not ‘stretch out’ from an origin pixel across more than ≈ 3 adjacent pixels (i.e., ≈ 3 km).

¹⁰We estimate the spatial decay separately for broad world regions (see Appendix Table A1), although this does not affect the estimates much.

areas in Africa and other low-income regions. Nechaev *et al.* (2021) use a Convolutional Neural Network (CNN) to downgrade the VIIRS series. While their approach is comparable to ours, they only apply it to the ‘unadjusted’ DMSP data, which suffers from blooming and top-coding, and as such, the merging is not conceptually very appealing. As shown below, the unadjusted series does a poorer job explaining local development. We take a machine learning (ML) approach agnostic to the functional form of the two series’ mapping. The concordance of the two series is a problem well-suited to an ML design, which discovers complex structures not specified in advance (Mullainathan and Spiess, 2017).

2.2.1 Method

To implement the downgrading of VIIRS and its merging with the DMSP series, we use an *ensemble method*, considered ‘state-of-the-art’ in machine learning, Sagi and Rokach (2018).¹¹ Ensemble methods combine different models to improve out-of-sample performance over a single model, (Athey and Imbens, 2019). We use an ‘extremely randomized trees,’ an averaging ensemble method that combines many decision trees, (Geurts *et al.*, 2006). The principle behind averaging methods is to build multiple independent models and then average their predictions. Random forests combine decision trees, each built from a random sample of observations and features (covariates/predictors); the decision trees use the best splits from the respective samples (Breiman, 2001). Extremely randomized trees take an extra step: instead of picking the ‘best’ thresholds from the sample of observations and features, pick them randomly, as doing so improves accuracy and computational efficiency (Geurts *et al.*, 2006).

We aim to predict DMSP-like values for 2014 onward using VIIRS data from those years. We train the model using data from 2013, the only *full* year of overlapping DMSP and VIIRS, and 30-arc-second pixels, which matches the DMSP resolution. As the VIIRS spatial resolution is 15-arc seconds, our pixels are aggregates of four VIIRS pixels. We use the following features: (i) pixel statistics (mean, median, min, and max of VIIRS-pixels); (ii) statistics of nearby pixels (mean and variance within windows of varying widths)¹²; and (iii) indicators for broad world regions.¹³ The ‘decision trees’ allow for all complex interactions between features. The extremely randomized tree has a set of (regularization) parameters that must be calibrated.¹⁴ We implement

¹¹Mullainathan and Spiess (2017) write that ‘*while it may be unsurprising that such ensembles perform well on average... it may be more surprising that they come on top in virtually every prediction competition.*’

¹²The window widths, in number of pixels, are 3, 4, 7, 9, 11, 13, 17, 21. So, for example, the first window is $3 \times 3 = 9$ pixels centered on and including the own-pixel.

¹³All features are listed in Appendix Table A1 along with the ‘feature importance’. The latter captures the total reduction of the squared error from a single feature, for each of the four models.

¹⁴The full set of parameters is: `n_estimators` (number of trees), `min_samples_split` (the minimum number of observations required to split an internal node), `min_samples_leaf` (the minimum number of samples required at a leaf node), `max_features` (the number of features to consider when looking for the best split), `max_depth` (the maximum tree length), and whether bootstrap samples are used when building trees or the full dataset is used for each tree.

a randomized search across parameter values evaluated by a randomized cross-fold validation to avoid misjudgment; we use the ‘*scikit learn*’ library in *python* that picks parameters that maximize the out-of-sample R^2 .

2.2.2 Performance

To judge performance, we retrain the model using data from 2012, as there is a partial overlap between DMSP and VIIRS, and then compare the out-of-sample predictions for 2013 to the actual DMSP values.¹⁵ The downgrading of VIIRS to the adjusted DMSP performs well both in absolute terms and compared to the recent efforts of Li *et al.* (2020) and Nechaev *et al.* (2021). For the global set of pixels, Figure 1 reports the results of our method’s out-of-sample performance (in 2013) and compares with Li *et al.* (2020) and Nechaev *et al.* (2021). In Appendix Figure A2, we repeat the analysis for Africa only.

First, the left panels plot the scatters of the predicted values (in the vertical axis) against the actual, blooming corrected, DMSP values (in the horizontal axis) for our method (panel a), the Li *et al.* (2020) sigmoid function method (panel c), and the Nechaev *et al.* (2021) convolutional neural network approach (panel e). Our method’s root mean square error (RMSE) is 1.50, considerably lower than the 3.27 of the approach of Li *et al.* (2020) and lower than the 1.57 of Nechaev *et al.* (2021).¹⁶ Our method performs better across the luminosity distribution but especially at the low to middle end, a feature particularly useful when studying regions in a low-income or (lower) middle-income setting.

Second, because low-light regions are common in Africa and many papers apply binary transformations, we also examine performance at the extensive margin. Panels (b), (d), and (f) report “confusion matrices” of lit and unlit pixels with the three methods. The rows correspond to actual adjusted DMSP values, and the columns to out-of-sample predicted DMSP values for 2013. The top-left counts pixels classified correctly as unlit, and the bottom-right counts pixels correctly classified as lit. For our method, the share of lit pixels correctly classified (*recall*) is 0.95 [17634634/(17634634 + 998461)]; the share of all predicted as lit pixels correctly classified (*precision*) is 0.58 [17634634/(17634634 + 12872108)]. These statistics depend on the distribution of lit and unlit pixels, which is highly skewed. The actual share of lit pixels is just 8.6%. Simply

¹⁵The only full (12 months) year where both Version 4 DMSP-OLS and VIIRS data are available is 2013. In 2012, the two datasets overlapped for eight months. For this reason, we use 2013 to train our main model (best overlap). We use 2012 to retrain our model for the out-of-sample performance check, as this gives us an estimate of model performance with partial training data. The Appendix shows no discontinuities between 2012 and 2013, the years of the DMSP and VIIRS overlap. Appendix Figure A1 uncovers a strong co-evolution between the harmonized luminosity series and the share of the population with electricity in Mozambique, Kenya, the Democratic Republic of Congo, Ghana, Tanzania, and Nigeria without a jump when moving from DMSP to VIIRS (see also Henderson *et al.* (2012)).

¹⁶Results are similar for Africa (Appendix Figure A2), albeit with lower absolute values due to lower lights: Our method’s RMSE is 0.715 vs 3.10 for Li *et al.* (2020) and 0.727 for Nechaev *et al.* (2021).

classifying all pixels as lit would get a recall score of 100%, but a precision of just 8.6%. The figure thus also reports the $F1$ score, a widely used metric to evaluate the success of binary classifiers when one class is rare (Lipton *et al.*, 2014). The $F1$ score takes the harmonic mean of the recall and precision scores; $F1 = 2 * recall * precision / (recall + precision)$. Higher $F1$, bounded between 0 and 1, indicates a better accuracy. Our method yields an $F1$ of 0.72, much higher than the 0.51 of the sigmoid approach and slightly higher than the 0.71 of the CNN approach.¹⁷

3 Cross-Country Patterns

While our focus is on the use of lights to capture local economic activity, we begin the analysis examining the cross-country association between nighttime luminosity and GDP for a global set of countries, using data from the World Bank’s World Development Indicators Database.¹⁸ Since much of our focus is on Africa, we repeat our analyses for African countries in Appendix Section B.1; see Appendix Figure B3 and Appendix Tables B3-B6.

Cross-sectional Association Figure 2 Panels (a) and (b) illustrate the strong cross-sectional association between GDP and the sum of the harmonized nighttime lights in 2005 (adjusted DMSP) and 2015 (downgraded VIIRS to the adjusted DMSP). Appendix Table B2 - Panel A explores in more detail the cross-sectional association. The highly significant coefficient shows that luminosity is a good output proxy across 173 countries. The fit is strong in both periods with an adjusted R^2 around 0.9. The elasticity is stable across periods (around 0.85); in Africa, the elasticity is 0.7, and the R^2 is around 0.9.

Within-Country over Time Association We run panel and long-run differences specifications to examine the dynamic association between nighttime lights and GDP. Figure 2 - Panels (c) and (d) plot the association between changes in GDP and the adjusted and harmonized series over 1992 – 2019 (to avoid capturing the pandemic and the recovery) and over 1992 – 2013, the DMSP period. The highly significant elasticity is similar over both periods, around 0.25 and 0.24, respectively (median regression estimates are identical). These estimates are similar, though slightly lower, to the ones [0.30 – 0.33] reported by Henderson *et al.* (2012) across 188 countries for 1992/3 – 2005/6.¹⁹ Panel (e) illustrates the panel association at the yearly frequency, plotting the residuals of GDP and luminosity, netting out country and year constants. As the noise in GDP and nighttime lights gets magnified at the yearly frequency, panel (f) gives the elasticity using five-year

¹⁷Results on $F1$ scores are slightly better for Africa (see Appendix Figure A2): The $F1$ of our ensemble method is 0.77, while for the sigmoid function approach is 0.17 and 0.75 for Convolutional Neural Networks method.

¹⁸We require that every country is observed in the DMSP and the VIIRS era. We drop Equatorial Guinea, an evident outlier; see also Henderson *et al.* (2012).

¹⁹Our specification in panels (c) and (d) are comparable to Table 3 column (4) in Henderson *et al.* (2012), with an estimate of 0.32.

averages of GDP and luminosity. The elasticity is around 0.15 – 17 with the merged series. Besides, as shown in the Appendix, the estimates with the adjusted light series are not dissimilar to the unadjusted ones in the global and African samples.

Finally, to grasp what is lost when downgrading and merging VIIRS to DMSP, we have run the 2014-2019 ‘long-difference’ regressions. With the original VIIRS data, the long-run difference GDP-luminosity long-run elasticity is 0.4 (0.076) for Africa, and 0.15 (0.057) for the global sample. With the downgraded series, it is 0.35 (0.094) for Africa, and 0.19 (0.046) across the 173 countries.

4 Local African Development

As applied research on African (long-run) development, economic history, and political economy has moved over the past years from cross-country approaches to designs that exploit spatial variation within countries (Michalopoulos and Papaioannou, 2018), we explore the potential of luminosity to capture well-being across African regions using georeferenced Demographic and Health Surveys (DHS).²⁰ As research moves from coarser units (like admin-1 units) to finer, more granular units, we explore the cross-sectional and dynamic patterns across spatial units of various sizes. Across all tests, we compare the newly compiled adjusted and merged VIIRS-DSMP lights series with unadjusted ones, as doing so illustrates a part of our contribution.

4.1 Data and Specification

We obtained all geo-referenced Demographic and Health Surveys from Africa. Appendix Table B7 reports the country-survey years, while Appendix Table B8 gives summary statistics. The 34 countries are from all parts of the continent, relatively richer and poorer, former British, French, Belgian, and German colonies. Most surveys were conducted in the 2000s and 2010s, but we also have over a dozen surveys in the 1990s. We extract four development indicators: the mean years of schooling of respondents aged 15-39 (*Adult Years Schl.*),²¹ the mean DHS composite household wealth index (*Wealth Index*),²² the share of households with access to improved sanitation (*Improved Sanitation*),²³ and the share of households with an electricity connection (*HH has Elect.*).

The geo-referenced DHS gives information across survey clusters (enumeration areas), typically cities, towns, and villages. We create an arbitrary grid of Africa, where each gridcell has a resolution

²⁰This follows a considerable body of research that uses DHS data to proxy African development; see, for example, Young (2012), Lu and Vogl (2023), Lowes and Montero (2021b), Ashraf *et al.* (2020), Fenske (2015), Michalopoulos and Papaioannou (2016), and Michalopoulos and Papaioannou (2020) for a review.

²¹We choose the 15-39 age range because, in the panel specifications, we want to capture the ‘flow’ of education. We get similar results using an upper age of 65.

²²The index is based on a principal component aggregation of household characteristics like the quality of the household roof and the ownership of assets.

²³The DHS definition of improved sanitation includes flush toilets to sewers, septic tanks, and also ventilated and slab pit latrines, while unimproved includes open pits, bucket toilets, and open defecation.

of 900 arc seconds, 30 times the resolution of the DMSP pixels; therefore, each gridcell contains 900 DMSP pixels. At the equator, this corresponds to a grid of $28km$ by $28km$; see Appendix Figure B4 for an example. We then aggregate the DHS data to these gridcells by matching the DHS cluster to the gridcell where their GPS coordinate falls, then aggregate by cell-year.

We associate the development proxies with luminosity, running the following specification:

$$Y_{g,c,t} = \beta NL_{g,c,t} + \gamma \ln(area)_g + \mu_{c(g),t} [+ \delta_g] + \epsilon_{g,c,t} \quad (1)$$

$Y_{g,c,t}$ denotes the average of the socio-economic outcome (e.g., schooling, composite wealth index, access to electricity) in gridcell g in country c , in a survey conducted in year t . To enable coefficient comparisons, we standardize all outcomes to have a mean of zero and a standard deviation of one.²⁴ $NL_{g,c,t}$ is either the log sum of nightlights plus half of the minimum positive value or an indicator that equals one if the gridcell is lit. The specifications include country-year fixed effects $\mu_{c(g),t}$, as our objective is to explore the usefulness of luminosity in capturing regional African development in a given country period. The cross-sectional specifications also control for the logarithm of the gridcell’s/unit’s area, $\ln(area)_g$, which is absorbed by the unit fixed effects $[\delta_g]$ in the panel estimation.

4.2 Cross-sectional Estimates

Figure 3 panels (a)-(b) plot the cross-sectional estimates with the adjusted series (red diamonds) and the unadjusted ones (blue squares). Two results emerge. First, nightlights are a suitable proxy for local development, as we obtain significant correlations across all specifications with both the harmonized and adjusted series and the unadjusted ones. The estimates in panel (b) hover around 0.6, suggesting that lit areas have a little more than half a standard deviation higher schooling and access to electricity and improved sanitation; the coefficient on the composite wealth index, which minimizes error on household assets and public goods access, suggests differences of one standard deviation. The correlation, however, is far from perfect as the binary luminosity index cannot fully capture the considerable spatial variation in development (Appendix Figure B5). Second, with all outcomes, we obtain *more robust and less noisy* correlations with the newly compiled, adjusted, and merged VIIRS-DMSP series compared to the unadjusted ones. Appendix Figure B9 Panels (a) and (b) examine this further by breaking down results by each type of luminosity correction. Most gains come from the blooming rather than the topcoding correction, which is consistent with our African context, where topcoding is rare. However, papers in other settings, countries, and

²⁴Note that this standardization is done for the main text figures, which plot coefficients from models with different development outcomes. For the appendix tables and figures with coefficients all from models with the same outcome, outcomes are left in their native units.

(sub)-continents might find that topcoding is more relevant. Finally, when we look at the intensive margin of luminosity (panel a), all coefficients increase with the harmonized series, consistent with a measurement error interpretation, as using less noisy explanatory variables yields less attenuated estimates and a higher R^2 (Wooldridge, 2010).

4.3 Panel Estimates

To examine the dynamic correlation between the DHS development proxies and luminosity, we augmented the cross-sectional specification with gridcell fixed effects (δ_g). The panel specifications, reported in panels (c) and (d) of Figure 3, yield positive coefficients. [Appendix Table B10 reports the estimates for each adjustment separately.] The coefficients with the adjusted series are *always* larger than the analogous ones with the unadjusted ones, telling of the reduction in measurement error that our adjustments to the DMSP series and merging to the downgraded VIIRS achieve. For example, the estimate for log lights in the specification with years of schooling is around 0.02 when we use the newly compiled series, which is about double the coefficient of the unadjusted series. Similar results hold when the outcome is the composite wealth index or the share of households in the gridcell with electricity access. The comparison of the specifications examining the association between development and the extensive margin of lights with the new and the unadjusted series in panel (d) yields starker patterns. Most specifications with the unadjusted series yield small and indistinguishable from zero estimates, while all permutations with the adjusted series yield much larger and statistically significant correlations. These results align with the fact that error in variables is often more significant when expressing the empirical model in differences. The harmonized series suggests that mean years of schooling increase, on average, by 0.05 standard deviations in gridcells turning lit, compared to unlit; this translates into 0.125 schooling years (Appendix Table B10). When cells turn lit with the newly compiled series, the Wealth Index and Electricity Access increase by around 0.05 standard deviations. Appendix Figure B9 Panels (c) and (d) show that, similarly to the cross-sectional results, most of the improvement of the corrected series comes from the blooming correction.

4.4 Further Evidence

4.4.1 Spatial Aggregation

Applied research uses luminosity data across spatial units of various sizes, some coarse (Alesina *et al.*, 2016 at admin-1 and admin-2 units and linguistic areas), some granular (Henderson *et al.*, 2018 at small grid-square level; Storeygard, 2016 at city-level). Figure 4 panels (a) and (b) provide graphical illustrations of the luminosity-composite wealth index elasticity across spatial units of various sizes (blocks of gridcells) to explore the implications of aggregation. The furthest to the left

gives the coefficient from a specification across relatively small blocks, 2×2 (four gridcells). As one moves towards the right along the x-axis, the data is aggregated into larger units, with the largest 12×12 (144 gridcells). The cross-sectional estimates in panel (a) are highly significant, around 0.18, across all aggregation levels. All panel estimates in panel (b) are highly significant, showing that luminosity approximates the variation in household asset changes and public goods access. The coefficients are fairly stable across the aggregation levels, around 0.075, although slightly stronger at larger spatial units. The estimates are almost always larger with the newly compiled adjusted luminosity series (red diamonds) compared to the unadjusted (blue squares); this is especially the case at the highest levels of spatial resolution, suggesting that the reduction in measurement error becomes more important at more disaggregated levels. Research should use the harmonized series and carefully consider measurement error as it moves into more granular analyses. In contrast, aggregating light data to coarser spatial units reduces noise in the unadjusted light series; a pattern that echoes the cross-country analyses showing minimal differences in GDP-luminosity elasticity using the adjusted and unadjusted light series.

4.4.2 Local Variation

Researchers commonly use identification designs, such as local fixed effects models (Wantchekon *et al.*, 2015) or spatial RDD (Michalopoulos and Papaioannou, 2014; Lowes and Montero, 2021a), to advance on causation by comparing proximate locations. We thus assess how well luminosity captures local development, focusing on estimates within increasingly proximate areas. We augment the empirical model with fixed effects of increasing spatial resolution (in the panel estimates interacted with year constants) to partial out localized hard-to-account-for features related to geography, ecology, and culture.

Figure 4 panels (c) and (d) plot the luminosity-wealth correlation with fixed effects of various sizes. Moving along the x-axis, we plot estimates with larger (coarser) fixed effects. The furthest to the left specification includes fixed-effects for blocks of 2×2 (four gridcells); the furthest to the right specification includes fixed-effects for blocks of 12×12 (144 gridcells). The cross-sectional estimates (panel c) are stable; the coefficients with the adjusted series are around 0.17, slightly larger than with the unadjusted series, about 0.15. The within-gridcell over time estimates in panel (d) highlight the improvement from our adjustments. The coefficient with the harmonized and adjusted series is significantly positive and very stable, although the proximity of the comparisons changes considerably as the fixed-effects increase in size. Even when comparing nearby areas, changes in luminosity correlate significantly with changes in household wealth. In contrast, the coefficients of the unadjusted luminosity series are smaller and, until fixed effects are at least 7×7 (roughly $200km \times 200km$ at the equator), statistically indistinguishable from zero. As shown in the Appendix Section B.2, the patterns are similar when using schooling and access to electricity

to proxy local well-being. The adjustments to the luminosity series strengthen considerably the cross-sectional and over time correlation with schooling and electricity access at granular levels of disaggregation and when exploiting (very) localized variation.

4.4.3 Urban and Rural

Another issue with light data regards its accuracy in explaining well-being within urban and rural areas. Figure 5 plots the luminosity coefficients for the development outcomes separately for urban and rural households (using the DHS classification). All cross-sectional estimates are highly significant, suggesting that nighttime lights proxy well schooling, household wealth, and service access across both urban (higher luminosity) and rural (lower luminosity) areas. Besides, the estimates are similar in the rural and urban samples. There is some evidence that the adjustment is more critical across rural areas. The panel specifications yield somewhat different patterns. First, the coefficients are statistically significant only when using the newly compiled harmonized and adjusted for top-coding, blooming, and sensor calibration lights series. Second, the estimates of the urban sample are consistently larger than those of the rural sample, showing that the development-luminosity nexus is more substantial in urban areas, an asymmetry that echoes the recent findings in India of Asher *et al.* (2021).

5 Country-Specific Case Studies

Many studies use luminosity from specific countries to explore various inquiries, such as the geographic impact of demonetization in India (Chodorow-Reich *et al.*, 2020), landmine clearance in Mozambique (Chiovelli *et al.*, 2025), and the flattening of the government hierarchy in China (Li *et al.*, 2016). In this Section, we examine the development-luminosity nexus at a granular level, focusing on low-income settings. First, we zoom in on Mozambique using Census data, which is less noisy than surveys. Second, we look across dozens of thousands of Indonesian villages. Third, we turn to more than half a million Indian rural and urban settlements.

5.1 Mozambique (Census-based Estimates)

We examine the association between the newly compiled luminosity series and local development using all post-civil war Mozambican censuses that allow us to zoom in at high spatial resolution with many observations²⁵

Approach and Sample We estimate linear specifications across Mozambican administrative units linking development to luminosity. We have retrieved, processed, and digitized the entire

²⁵Recent work in Namibia shows that luminosity tracks local development better in census compared to survey data (Maatta *et al.*, 2022).

censuses of 1997, 2007, and 2017, available across admin-4 level units. We estimate the following specification:

$$Y_{i,t} = \beta L_{i,t} + \gamma_a \ln(\text{area})_i [+ \delta_i] + \mu_{j(i),t} + \epsilon_{i,t} \quad (2)$$

$Y_{i,t}$ denotes the average years of schooling of Mozambicans 15-39 years and non-agriculture employment of 15-24 year-olds in administrative unit i in period t .²⁶ $L_{i,t}$ is either the log sum of nightlights plus half of the minimum positive value or a lit indicator. We include fixed effects for year-by-admin unit (at various levels), $\mu_{j(i),t}$. The cross-sectional specification also controls for log geographic area, $\ln(\text{area})_i$. Administrative unit fixed effects, δ_i , account for geography, location, and other time-invariant factors in the panel estimation. Appendix Table B11 gives summary statistics across 1,126 admin-4 units (*localities*).

Cross-Sectional Estimates Figure 6 panels (a) and (b) report the cross-sectional estimates at the locality level with the two transformations of luminosity. The specifications reveal a significant luminosity-development correlation, further illustrating the usefulness of luminosity to approximate localized differences in education and employment in the “modern” sector. Besides, the coefficients with the harmonized and adjusted series are stronger than the analogous ones with the “unadjusted” light data, showing the reduction in measurement error from blooming. Estimates imply that years of schooling and employment in the modern sectors are (at least) half a standard deviation higher in lit compared to unlit localities, about 0.5 years and 10 percentage points, respectively.

The luminosity-development correlation is also present when we add admin-3 unit fixed effects (*postos*) to exploit localized variations across proximate localities. Appendix Tables B12-B13 reports cross-sectional estimates also across 142 admin-2 areas (*distritos*), including (10) admin-1 (provinces) fixed effects and across 403 admin-3 units (*postos*), including admin-2 fixed effects. Luminosity is a significant proxy of education and non-agriculture employment across all administrative splits. As with the DHS analysis, the improvement in estimates from the harmonized and adjusted for top-coding, blooming, and sensor calibration series is mostly noticeable at finer spatial resolution; when the units are large, then the adjustments do not matter much.

Dynamic Correlations Figure 6 panels (c) and (d) give panel estimates (with admin-4 unit fixed effects) that explore the dynamic association between luminosity and development. Within-locality changes in luminosity correlate significantly with swings in schooling and out-of-agriculture employment. As shown in Appendix Tables B14-B15, the correlation is strong across all levels of spatial aggregation. Luminosity co-moves with schooling and modern-sector employment even when

²⁶Employment status is not available for the 2017 census, so any specifications with this outcome include only data from 1997 and 2007. Also, similar to our years of schooling outcomes, we measure employment for the sample of respondents aged 19-24 in order to better approximate changes in economic conditions in the panel estimation.

we augment the specifications with interactions between census-year constants and admin-3 unit fixed effects that allow us to zoom within geographically proximate areas and account for quite localized unobserved trends. Besides, the panel estimates for non-agricultural employment are considerably larger with the newly compiled fused VIIRS to the adjusted DMSP series. Appendix Figure B10 illustrates the within-locality patterns plotting the increase in schooling years for four groups of localities; initially (in 1997) unlit admin-4 units that either stay unlit (by 2007 or 2017) or turn lit, and initially lit localities that either stay lit or turn unlit. The difference in schooling is about half a year when comparing localities turning lit (from unlit) or staying lit (rather than becoming unlit), even when we compare nearby localities with the inclusion of admin-3 fixed-effects.

5.2 Indonesia

Approach and Sample We then examine the link between luminosity and local development, using very granular village-level data from Indonesia. We rely on the Village Potential Statistics Census (PODES, *Pendataan Potensi Desa*), conducted every three to four years since 1996, and the associated village-level shapefile from 2000.²⁷ This high-quality dataset has been used to examine the development and political economy impact of large school construction programs (Martinez-Bravo, 2017), ethnic mixing’s role in nation-building (Bazzi *et al.*, 2019), and administrative decentralization on local public goods (Cassidy and Velayudhan, 2025), among others. We use the 1996, 1999, 2002, 2005, 2008, 2011, 2014, and 2018 waves. Due to evolving village boundaries, some villages could not be reliably merged across years. Nonetheless, we retain over 60,000 villages.²⁸ The PODES data help quantify the luminosity-development nexus at a very granular level. We associate various public goods measures (presence of a primary, secondary school, or kindergarten, access to drinking water, availability of doctors) with the log luminosity and an indicator for lit areas. The specification always includes survey (year)-specific (3,737) admin-3 fixed effects. The cross-sectional specifications also condition on the log village area.²⁹ As PODES includes various potential outcomes, we aggregate them via principal components, making the estimates comparable to those with the DHS wealth index.³⁰ Appendix Table B16 gives summary statistics across 61,601 villages.

²⁷This data was kindly shared by Samuel Bazzi.

²⁸The shapefile includes multiple polygons for some villages, particularly those on islands. These were consolidated using administrative codes to ensure a single geometry per village. Some villages appear only in one dataset, with 4,679 unique to the shapefile and 410 unique to the survey. The shapefile does not cover the Papua province.

²⁹The average (median) size of a village is 21.87 (5.68) km^2 with a standard deviation of 80.25 $.km^2$. The average (median) village has a population of 3,739.77 (2,369) with a standard deviation of 5,454.364, as recorded in 2011.

³⁰The public goods and local development proxy measures include: binary measures for access to formal garbage disposal, use of toilet facilities, access to drinking water, use of gas or electricity for cooking, and presence of paved roads, and the number of kindergartens, primary, middle, and secondary schools.

Results Figure 7 Panels (a)-(d) report the cross-sectional correlations between the various public goods measures and the unadjusted and adjusted luminosity series. The top row correlates the two luminosity series with the first principal component, which captures the overall level of public goods.³¹ The correlation is highly significant with both nighttime lights series and both transformations of luminosity. However, the coefficient is much larger with the adjusted and harmonized luminosity series, indicating a reduction in measurement error. The adjusted night light series (almost) consistently outperforms the unadjusted version with all public goods proxies. The panel estimates, reported in panels (c) and (d), also tell of the reduction in noise from the adjustments to the DMSP series and its harmonization with the VIIRS. The coefficient on the adjusted nighttime lights with the composite wealth index suggests a significantly positive correlation, while with the unadjusted luminosity series, the estimate is, counterintuitively, negative. Turning to the various public goods proxies, the estimate with the adjusted series is mostly positive and significant, while this is rarely the case with the unadjusted luminosity.

5.3 India

Approach and Sample We then turn to India, relying on the recently compiled Socioeconomic High-Resolution Rural-Urban Geographic Dataset on India (SHRUG), compiled by Asher *et al.* (2021). This is a rich dataset recording various social, economic, political, and development features for the *universe*, more than 550,000 of municipalities (towns and villages) and legislative constituencies in India. For example, the villages have a median population of about 832, and the towns a median just short of 15,000.³² This dataset –and some of its core components– have been used to study the role of irrigation on structural transformation (Asher *et al.*, 2023), the role of transportation investments on education (Asher and Novosad, 2020), and the impact of local government size on public goods (Narasimhan and Weaver, 2024).

We followed the luminosity validation in Asher *et al.* (2021) and examined the association between various development proxies from the 1991, 2001, and 2011 Population Census of India (population count) as well as variables from the 1990, 1998, 2005, and 2013 Economic Censuses (like total non-farm, total manufacturing, and total services employment) and the logarithm of luminosity.³³ The specification always includes census (year)-specific admin-3 (subdistricts) fixed effects.

³¹All public goods measures load positively in the first principal component, which captures about a third of the common variance, with an eigenvalue close to 3.5.

³²The open-access geospatial data portal, maintained by the Development Data Lab (<https://www.devdatalab.org/>), integrates many high-resolution socioeconomic, demographic, political, and remote sensing data for over 600,000 villages and towns with a harmonized spatial identifier from 1990 to 2018.

³³In contrast to the original study (Asher *et al.*, 2021), which used SHRUG version 1.5, we used the latest SHRUG version 2.1 (Pakora 2.1) due to the unavailability of spatial data for version 1.5. Night light data is available from 1992 onwards thus we used the 1992 night light data to represent 1990 and 1991, to align with the PCA and EC data. We have also dropped the variables on electrification, given the known “questionable relationship with the actual availability of electricity,” reported in the SHRUG’s codebook. We merged our luminosity data using SHRUG’s unique location identifiers.

The cross-sectional specifications also condition on the log of the municipality area. Appendix Table B17 Panel D reports descriptive statistics across the 550,000 municipalities.

Results Figure 8 Panel (a)-(d) follows Table 3 and 4 from Asher *et al.* (2021). Panels (a) and (b) present the cross-sectional estimates of the log luminosity at the town (urban) and village (rural) levels, respectively. Panels (c) and (d) give the corresponding panel estimates. With the exception of the cross-section in the urban areas where the unadjusted series delivers larger coefficient estimates, across most outcomes, especially non-farm, manufacturing, and services employment, the adjusted series regressions yield economically and statistically stronger associations, suggesting a stronger link between luminosity and economic activity. This likely reflects improved measurement accuracy in the adjusted data, which explicitly corrects for blooming and top-coding issues that may have attenuated estimates in the original DMSP series.

6 Applications

We now re-examine three papers that proxy local African and global development with luminosity due to the unavailability of regional data on income/output.

6.1 Precolonial Ethnic Institutions and Contemporary African Development

Inquiry and Approach Michalopoulos and Papaioannou (2013) examine the within-country long-run correlation between precolonial political centralization and contemporary regional development, blending ethnographic information (from Murdock (1959, 1967)) on ethnicities' political organization and their spatial distribution at the onset of European colonization, and luminosity, conditioning on country constants, geographical, locational, and ecological features. Table 1 replicates their core results, spanning 682 country-ethnic homelands. Michalopoulos and Papaioannou (2013) use the (logarithm of) mean luminosity value in 2007 and 2008, using the original DMSP series without any adjustment. The primary explanatory variable is an ordered index proxying ethnic level jurisdictional hierarchy beyond the local level. The measure equals zero for acephalous, fragmented societies only organized at the village/settlement level; one and two indicate petty and large chiefdoms, while three and four indicate kingdoms and large states. Following Gennaioli and Rainer (2007), they also use a binary political centralization index where one indicates politically centralized ethnic groups (large chiefdoms and states) and zero indicates non-centralized groups and those organized as small chiefdoms.

Results In Table 1, we compare their findings using the raw DMSP/OLS series to results based on the adjusted series. The dependent variable is the logarithm of average luminosity, defined

as $\log(0.01 + \text{mean luminosity})$ at the country-ethnic homeland level.³⁴ Columns (1) and (2) reproduce the original estimates using the jurisdictional hierarchy index and a binary measure of political centralization (columns (4) and (8) in the original paper, respectively), conditioning on country constants, the log population density, and a rich set of geographic controls. Within African countries, one sees higher levels of local development, as reflected in nighttime lights, in the ancestral homelands of politically centralized groups.³⁵ Columns (3) and (4) report otherwise identical specifications, using the adjusted for sensor inter-calibration, top-coding, and blooming DMSP series in 2007-2008. The coefficients on the 0 – 4 jurisdictional hierarchy index and the binary political centralization are highly significant. The adjustment of the luminosity series is not associated with a major change in the estimates’ economic effect; the coefficients are somewhat smaller, but luminosity is also on average smaller as the new series adjust for blooming and, hence, the number (share) of unlit country-ethnic observations increases from 164 (24%) to 197 (29%).

The fact that the raw DMSP-OLS and the adjusted DMSP series yield similar results is not surprising, as the analysis’s units (country-ethnic areas) are very large (the average (median) area is 25,547 (9,327) squared kilometers) and, as such, the adjustments do not matter much as the noise “averages out”. Columns (6)-(8) extend the analysis to the whole 1992–2023 period. Looking at a longer horizon assuages concerns that the estimates pick up idiosyncrasies in 2007 and 2008; besides, there are fewer observations with zero lights (12%), and hence the logarithmic transformation is more appropriate. Across all specifications, the coefficients on the proxies of precolonial political centralization retain economic and statistical significance. Since the country-ethnic homelands are pretty large, the results with the adjusted and unadjusted series are similar.

6.2 National Institutions and Sub-national Development in Africa

Inquiry and Approach Michalopoulos and Papaioannou (2014) examine the link between national institutions and regional development in Africa, exploiting the fact that African borders, designed in late 19th century in European capitals at a time when imperial powers had limited knowledge of local conditions, partitioned many ethnic groups at independence. By overlaying Murdock (1959) map delineating the historical homelands of 825 ethnic groups at the onset of colonization with contemporary country borders, which to a large extent, follow colonial ones, they identify 220 systematically partitioned ethnic groups (where at least 20% of a group’s homeland falls into more than two countries). Michalopoulos and Papaioannou (2014) then associate country-ethnic regional development, as reflected in DMSP nighttime lights series (in 2007-2008), to

³⁴As in the original analysis, standard errors are double-clustered across countries and ethno-linguistic families.

³⁵Murdock’s data are noisy and unavailable for some parts of the continent. Michalopoulos and Papaioannou (2015) show that the strong correlation between luminosity and precolonial political statehood emerges with an alternative to Murdock’s proxies of political centralization. Besides, the correlation strengthens in areas far from the capital cities, as there national institutions appear to matter more (Michalopoulos and Papaioannou, 2014).

national institutions proxies, conditioning on ethnic homeland fixed-effects. By exploiting within-ethnicity variation, their design accounts for cultural and geographic differences, which are strong cross-country correlates of GDP and national institutions. Besides analyzing country-ethnic (large) regions, they also run spatial RDD models, narrowing the comparison to grid squares of 0.125×0.125 decimal degrees.

Their analysis yields three results. First, when pooling across all partitioned ethnic areas, national institutions do not systematically correlate with local development/luminosity. Second, the null average effect masks considerable heterogeneity. Third, proximity to the capital, coupled with the weak state capacity of African states and their chronic inability to broadcast power in the periphery, partly explains the weak association between national institutions and at-the-border development. When zooming into split ethnic homelands where both areas are close to the respective capital cities, a significant correlation between national institutions and regional development emerges.

Results Table 2 reexamines some of the results, comparing the estimates with the original raw DMSP-OLS (Panel *A*) and the adjusted ones (Panel *B*). The dependent variable takes the value of one when the gridcell belonging to the ancestral homeland of partitioned ethnic groups is lit and zero otherwise. As misclassifications in binary outcome variables always yield non-classical measurement error (e.g., Aigner (1973), Meyer and Mittag (2017)), Panel *C* pools all yearly observations from 1992-2023 with the adjusted and harmonized series, as this accounts for noise in the outcome variable and reduces the number of zeros (the share of lit pixels doubles). Columns (1)-(2) report specifications across all gridcells of partitioned ethnic homelands (analogous to Table IV-Panel B, Columns (1) and (2) of the paper). Column (1) conditions on the pixel log land area and population.

The cross-sectional specifications with both the original and the adjusted for blooming, sensor quality, and top coding series yield significantly positive coefficients, suggesting that a one-point increase in the rule-of-law index (which ranges globally from -2.5 to 2.5) moves in tandem with a ten percentage points higher likelihood that the gridcell will be lit. Column (2) adds the fixed effects of the ethnic homeland, which allow for a comparison of the regional development of the partitioned areas of the same ethnicity. The within-ethnicity coefficient drops by more than half, to about 0.025. With the adjusted series, the estimate passes (marginally) the 90% statistical significance cutoff, illustrating the benefits of the noise reduction the new series achieves; with the unadjusted series, the estimate’s p -value (t-stat) was around 0.15 (1.55). Columns (3)-(4) report local regressions that restrict the estimation to gridcells within $50km$ from each side of the border (Table V-Panel B Column (1) and (2) of the paper).

The significantly positive cross-sectional coefficient (in (3)) drops considerably when adding the ethnicity constants, implying much weaker economic effects. The within-ethnicity estimate

is statistically indistinguishable from zero with the “raw” DMSP lights data (column (4)). The estimate is larger and weakly (at the 90%) significant with the adjusted luminosity series that reduces noise. Columns (5)-(6) give pooled across all split ethnic homelands spatial RDD estimates (adding an ethnic-country-specific third and fourth order RD polynomial on distance to the border), which approximate the effect of national institutions at the border (Table VI Panel A Columns (3) and (4) in the paper). The RD estimates are small and statistically insignificant for both luminosity series, suggesting that the economic impact of national institutions at the African border is, on average, small.

In columns (7)-(8), we distinguish between split ethnic groups where both partitioned areas are relatively close to the respective capital cities and those where both ethnic regions are far from the capitals (Table VIII Panel B Columns (5) and (6) in the paper). The analysis with the unadjusted luminosity series in Panel A yields a positive and marginally significant coefficient for partitioned ethnic homelands where both split areas are close to the respective capitals. The estimate with the adjusted series (in Panel B) yields a much more precise estimate. The estimate is highly significant when pooling across all years (1992 – 2023), implying that national institutions matter for local development when looking at areas not far from the capitals. However, when looking at areas far from the respective capitals, the correlation between national institutions and local luminosity is nil, both with the unadjusted and the less noisy adjusted and harmonized across satellite systems series.

6.3 Regional Favoritism

Inquiry and Approach Hodler and Raschky (2014) take a systematic global viewpoint on regional favoritism, which numerous case studies reveal in specific settings. They examine how development changes at the birthplace of national leaders when they take on power or leave office. Their design exploits yearly variation in luminosity and national leaders’ home-place (from the *Archigos* database) across 38,427 subnational regions from 126 countries from 1992 to 2009. Their panel specification conditions on region fixed-effects, which absorb geographic, locational, and other time-invariant factors, and country-year fixed-effects, which account for the impact of all national policies and institutions. Hodler and Raschky (2014) show that, controlling for local regional features and the general yearly developments in the country, the luminosity of leaders’ home region increases (falls) when they take on (leave) office. In addition, this result is pronounced in countries with weak institutions and low human capital, and years of sizable foreign aid flows.³⁶

³⁶In subsequent work, Luca *et al.* (2018) show similar patterns of ethnic favoritism with nighttime luminosity intensity increasing by 7 – 10% in the national leaders’ ethnic homelands

Results Table 3 reassesses their influential results with both the original raw DMSP-OLS and the adjusted luminosity series; a nice feature for our purposes of their study is the fact that they exploit within-region over-time variation, which often magnifies error-in-variables. Panel (a) reproduces the main results of Hodler and Raschky (2014) (Table 2 in the original paper) with the raw DMSP-OLS luminosity series. The authors take a log transformation of luminosity after adding a small number to avoid losing data from regions where the satellites do not detect any lights. We retain the original paper’s clustering at the leader’s level. A statistically positive within-region association emerges across all permutations (controlling for inertia in nighttime lights, the leader’s role, and population). The linear probability model estimate in column (6) suggests an increase in the likelihood that the leader’s hometown is lit of about three percentage points in the year after the leader gets into power. Panel (b) gives identical specifications but with the adjustments for top-coding, sensor inter-calibration, and blooming DMSP series. In line with the original analysis, the coefficients on the leader’s home region are always positive and at least two standard errors larger than zero. This applies in all permutations. The coefficients are larger and more precisely estimated. For example, in specification (1), the coefficient on the lagged leader increases from 0.038 to 0.058. Similar gains in magnitude and precision are evident across most specifications.³⁷ The adjustment in the linear probability model estimates in (7) is important as a binary transformation of a noisy outcome measure yields non-classical measurement error (e.g., Aigner (1973), Meyer and Mittag (2017)). The coefficient on the (lagged) leader suggests a seven percentage point higher likelihood of his/her birthplace being lit after he/she resumes office, much higher than the analogous estimate with the plain/raw nighttime lights series. Overall, the adjusted series, if anything, strengthens the evidence for regional favoritism by political leaders, likely due to improvements in spatial accuracy and correction for light “blooming, making the signals of localized development more distinct and robust.

7 Conclusion

While satellite nighttime lights have gained widespread popularity in applied research, it is still unclear when and where the luminosity data is a dependable proxy of economic development. The debate is especially pertinent in high-resolution analyses in low-development regions, where a significant pixel share is dark. But it is in low-income areas that satellite imagery is *a priori* needed, as in such environments there is limited high-quality data. Accounting for noise in the nighttime light series is especially relevant in low-development regions and high levels of granularity, as measurement error is more pertinent, and not necessarily classical, when applying binary transformations.

³⁷Notably, specification (6), which drops zero-luminosity regions due to the log transformation without a constant, shows a slight discrepancy in the number of observations across panels, reflecting differences in the density of non-zero grid squares between datasets.

In the first part, we compile a novel annual series of gridded global nighttime light data at a spatial resolution of 30 arc seconds (roughly one square kilometer at the equator). Employing ensemble ML methods, we standardize luminosity data from various satellites with sensors of differing resolutions and accuracy. The new series accounts for various intricacies related to variations in sensor quality, top-coding, blooming phenomena, and the transition from the DMSP-OLS to the VIIRS satellite systems in 2013.

The second part of our study explores the relationship between luminosity and local development proxies such as education, household wealth, sectoral employment, and access to public services. We commence the validation analysis drawing upon geo-referenced survey data from 34 African countries. By harnessing both cross-sectional and temporal variation, we show that the newly harmonized luminosity series effectively encapsulates local African development (dynamics). This holds even at highly granular levels, where the challenge of excess-zero observations is pronounced. The adjusted and harmonized luminosity series correlate much more strongly with all development proxies, even more so in changes, at granular levels, and when exploiting localized variation, telling of the reduction in measurement error that the newly compiled series achieves. We then zoom into three (large) countries: Mozambique, using all post-independence Censuses (1997, 2007, 2017); Indonesia, using a rich administrative triennial survey of public goods spanning over 60,000 villages since the mid-1990s; and India, relying on a recently compiled dataset recording local employment across more than half a million villages and towns. Although the settings, development proxies, and data collection differ, the country-level analyses yield similar patterns. While the correlations are far from perfect, the newly compiled adjusted data for top-coding, sensor inter-calibration, and blooming and harmonized across satellites are significant correlates of local development and public goods. The adjustments yield stronger (cross-sectional and over-time) correlations when focusing on proximate districts and villages.

Lastly, we re-examine the baseline results of Michalopoulos and Papaioannou (2013, 2014) on the lasting importance of precolonial ethnic political institutions on contemporary African regional development alongside the small role of national institutions within split-by-the-border ethnic homelands, and the findings of Hodler and Raschky (2014) on regional favoritism across the world with the newly compiled luminosity series. The significantly positive (within-country) association between precolonial political centralization and regional development across ancestral ethnic homelands is similarly strong with the adjusted and the unadjusted series; as the county-ethnic homelands are large, the various adjustments do not matter much. With the less noisy adjusted luminosity series, the overall weak correlation between national institutions proxies and development across split ethnic homelands somewhat strengthens, although the effect at the border continues to be tiny and statistically insignificant. We also reaffirm with the new luminosity series the results of Hodler and Raschky (2014) showing an increase in luminosity and a higher likelihood of being

lit at the national leaders' home districts, once they take office. If anything, the panel correlations strengthen with the adjusted and harmonized series, telling of the reduction in noise, often more relevant when exploiting within-region over-time variation.

We view our work as offering insights into the ongoing discourse concerning the reliability and utility of satellite nighttime light data in proxying economic development and local public goods in regions grappling with data limitations. However, while satellite imagery of nighttime lights appears helpful in reflecting local well-being across granular low-income settings, the correlations are far from perfect. Besides, as more satellite data are becoming available to researchers, future research can blend the newly compiled nighttime lights series with other granular satellite data, such as daytime economic activity (traffic) and imagery of structures, to provide high-quality mappings of well-being in low-income settings.

References

- Aigner, D.J. (1973). ‘Regression with a binary independent variable subject to errors of observation’, *Journal of Econometrics*, vol. 1(1), pp. 49–59, doi:10.1016/0304-4076(73)90026-6.
- Alesina, A., Michalopoulos, S. and Papaioannou, E. (2016). ‘Ethnic inequality’, *Journal of Political Economy*, vol. 124(2), pp. 428–488.
- Asher, S., Champion, A., Gollin, D. and Novosad, P. (2023). ‘The long-run development impacts of agricultural productivity gains: Evidence from irrigation canals in india’, Unpublished manuscript.
- Asher, S., Lunt, T., Matsuura, R. and Novosad, P. (2021). ‘Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform’, *World Bank Economic Review*, vol. 35(4).
- Asher, S. and Novosad, P. (2020). ‘Rural roads and local economic development’, *American Economic Review*, vol. 110(3), pp. 797–823, doi:10.1257/aer.20180268.
- Ashraf, N., Bau, N., Nunn, N. and Voena, A. (2020). ‘Bride price and female education’, *Journal of Political Economy*, vol. 128(2).
- Athey, S. and Imbens, G.W. (2019). ‘Machine learning methods that economists should know about’, *Annual Review of Economics*, vol. 11(1), pp. 685–725.
- Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A. and Zhang, Q. (2017). ‘Roads, railroads, and decentralization of chinese cities’, *The Review of Economics and Statistics*, vol. 99(3), pp. 435–448.
- Bazzi, S., Gaduh, A., Rothenberg, A.D. and Wong, M. (2019). ‘Unity in diversity? how intergroup contact can foster nation building’, *American Economic Review*, vol. 109(11), pp. 3978–4025, doi:10.1257/aer.20180762.
- Bilicka, K. and Seidel, A. (2022). ‘Measuring Firm Activity from Outer Space’, NBER WP 29945.
- Bleakley, H. and Lin, J. (2012). ‘Portage and path dependence’, *Quarterly Journal of Economics*, vol. 127, pp. 587–644.
- Bluhm, R. and Krause, M. (2022). ‘Top lights: Bright cities and their contribution to economic development’, *Journal of Development Economics*, vol. 157.
- Breiman, L. (2001). ‘Random forests’, *Machine Learning*, vol. 45, pp. 5–32.

- Cao, X., Hu, Y., Zhu, X., Shi, F., Zhuo, L. and Chen, J. (2019). ‘A simple self-adjusting model for correcting the blooming effects in dmsp-ols nighttime light images’, *Remote Sensing of Environment*, vol. 224, pp. 401–411.
- Cassidy, T. and Velayudhan, T. (2025). ‘Government fragmentation and economic growth’, *Review of Economics and Statistics*, *Forthcoming*.
- Ch, R., Martin, D.A. and Vargas, J.F. (2021). ‘Measuring the size and growth of cities using nighttime light’, *Journal of Urban Economics*, vol. 125.
- Chen, X. and Nordhaus, W.D. (2011). ‘Using luminosity data as a proxy for economic statistics’, *Proceedings of the National Academy of Sciences*, vol. 108(21), pp. 8589–8594.
- Chen, Y., Ursavaş, U. and Mendez, C. (2024). ‘Can higher-quality nighttime lights predict sectoral gdp across subnational regions? urban and rural luminosity across provinces in türkiye’, *Letters in Spatial and Resource Sciences*, vol. 17(12), doi:10.1007/s12076-024-00339-6.
- Chiovelli, G., Michalopoulos, S. and Papaioannou, E. (2025). ‘Landmines and Spatial Development’, *Econometrica*, *forthcoming*.
- Chodorow-Reich, G., Gopinath, G., Mishra, P. and Narayanan, A. (2020). ‘Cash and the economy: Evidence from india’s demonetization’, *The Quarterly Journal of Economics*, vol. 135(1).
- Donaldson, D. and Storeygard, A. (2016). ‘The view from above: Applications of satellite data in economics’, *Journal of Economic Perspectives*, vol. 30(4), pp. 171–198.
- Dreher, A., Fuchs, A., Hodler, R., Parks, B.C., Raschky, P.A. and Tierney, M.J. (2019). ‘African leaders and the geography of china’s foreign assistance’, *Journal of Development Economics*, vol. 140, pp. 44–71.
- Elvidge, C.D., Zhizhin, M., T., G., FC, H. and J, T. (2021). ‘Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019’, *Remote Sensing of Environment*, vol. 68(1), pp. 77–88.
- Fenske, J. (2015). ‘African polygamy: Past and present’, *Journal of Development Economics*, vol. 117, pp. 58–73.
- Gennaioli, N. and Rainer, I. (2007). ‘The modern impact of precolonial centralization in africa’, *Journal of Economic Growth*, vol. 12(3), pp. 185–234, doi:10.1007/s10887-007-9017-z.
- Geurts, P., Ernst, D. and Wehenkel, L. (2006). ‘Extremely randomized trees’, *Machine Learning*, vol. 63, pp. 3–42.

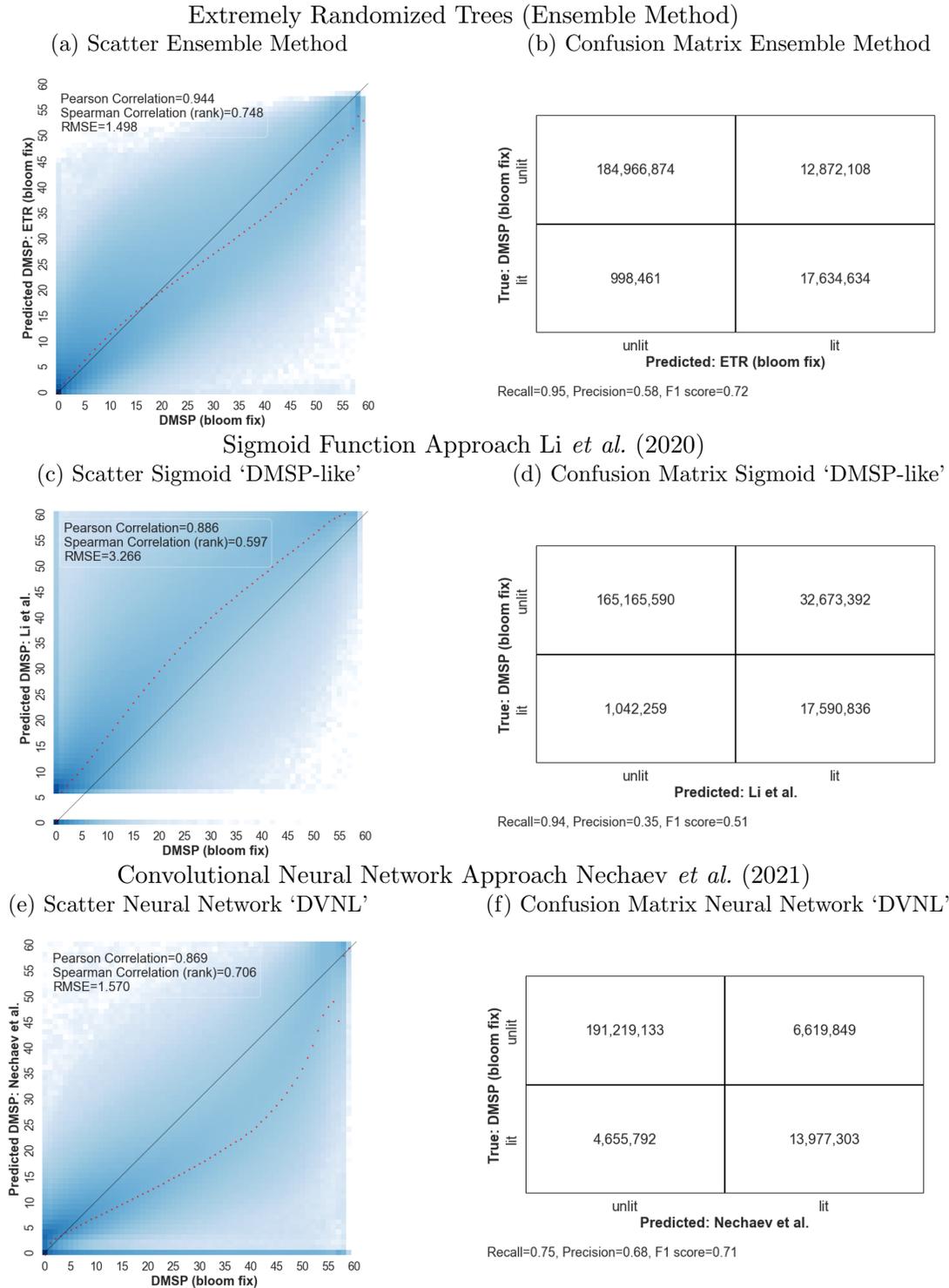
- Gibson, J. (2021). ‘Better night lights data, for longer’, *Oxford Bulletin of Economics and Statistics*, vol. 83(3), pp. 770–791.
- Gibson, J., Olivia, S., Boe-Gibson, G. and Li, C. (2021). ‘Which night lights data should we use in economics, and where?’, *Journal of Development Economics*, vol. 149, pp. 102–602.
- Harari, M. (2020). ‘Cities in bad shape: Urban geometry in india’, *American Economic Review*, vol. 110, pp. 2377–2421.
- Henderson, J.V., Squires, T., Storeygard, A. and Weil, D.N. (2018). ‘The global distribution of economic activity: Nature, history, and the role of trade’, *The Quarterly Journal of Economics*, vol. 133(1), pp. 357–406.
- Henderson, J.V., Storeygard, A. and Deichmann, U. (2017). ‘Has climate change driven urbanization in africa?’, *Journal of Development Economics*, vol. 124, pp. 60–82.
- Henderson, J.V., Storeygard, A. and Weil, D.N. (2012). ‘Measuring economic growth from outer space’, *American Economic Review*, vol. 102(2), pp. 994–1028.
- Hjort, J. and Poulsen, J. (2019). ‘The arrival of fast internet and employment in africa’, *American Economic Review*, vol. 109(3), pp. 1032–1079.
- Hodler, R. and Raschky, P. (2014). ‘Regional favouritism’, *Quarterly Journal of Economics*, vol. 129(2), pp. 995–1023.
- Hsu, F.C., Baugh, K.E., Ghosh, T., Zhizhin, M. and Elvidge, C.D. (2015). ‘Dmsp-ols radiance calibrated nighttime lights time series with intercalibration’, *Remote Sensing*, vol. 7(2).
- Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B. and Ermon, S. (2016). ‘Combining satellite imagery and machine learning to predict poverty’, *Science*, vol. 353(6301), pp. 790–794.
- Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A. and Rosing, T. (2022). ‘Using neural networks to predict micro-spatial economic growth’, *American Economic Review: Insights*, forthcoming.
- Levin, N., Kyba, C.C., Zhang, Q., Sánchez de Miguel, A., Román, M.O., Li, X., Portnov, B.A., Molthan, A.L., Jechow, A., Miller, S.D., Wang, Z., Shrestha, R.M. and Elvidge, C.D. (2020). ‘Remote sensing of night lights: A review and an outlook for the future’, *Remote Sensing of Environment*, vol. 237.
- Li, P., Lu, Y. and Wang, J. (2016). ‘Does flattening government improve economic performance? evidence from china’, *Journal of Development Economics*, vol. 123, pp. 18–37.

- Li, X. and Zhou, Y. (2017). ‘A stepwise calibration of global dmsp/ols stable nighttime light data (1992–2013)’, *Remote Sensing*, vol. 9(6).
- Li, X., Zhou, Y., Zhao, M. and Zhao, X. (2020). ‘A harmonized global nighttime light dataset 1992–2018’, *Nature: Scientific Data*, vol. 7(168).
- Lipton, Z.C., Elkan, C. and Naryanaswamy, B. (2014). ‘Optimal thresholding of classifiers to maximize f1 measure’, in (T. Calders, F. Esposito, E. Hüllermeier and R. Meo, eds.), *Machine Learning and Knowledge Discovery in Databases*, pp. 225–239, Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lowes, S. and Montero, E. (2021a). ‘Concessions, violence, and indirect rule: Evidence from the congo free state’, *Quarterly Journal of Economics*, vol. 136(4), pp. 2047–2091, doi:10.1093/qje/qjab021.
- Lowes, S. and Montero, E. (2021b). ‘The Legacy of Colonial Medicine in Central Africa’, *American Economic Review*, vol. 11(4), pp. 1284–1314.
- Lu, F. and Vogl, T. (2023). ‘Intergenerational persistence in child mortality’, *American Economic Review: Insights*, vol. 5(1), p. 93–110.
- Luca, G.D., Hodler, R., Raschky, P.A. and Valsecchi, M. (2018). ‘Ethnic favoritism: An axiom of politics?’, *Journal of Development Economics*, vol. 132, pp. 115–129, doi:10.1016/j.jdeveco.2017.12.002.
- Maatta, I., Ferreira, T. and Leßmann, C. (2022). ‘Nighttime lights and wealth in very small areas’, *Review of Regional Research*, vol. 42, pp. 161–190.
- Martinez, L. (2022). ‘How much should we trust the dictator’s gdp growth estimates?’, *Journal of Political Economy*, vol. 130(10), pp. 2731–2769.
- Martinez-Bravo, M. (2017). ‘The local political economy effects of school construction in indonesia’, *American Economic Journal: Applied Economics*, vol. 9(2), pp. 256–289, doi:10.1257/app.20150563.
- Meyer, B.D. and Mittag, N. (2017). ‘Misclassification in binary choice models’, *Journal of Econometrics*, vol. 200(2), pp. 295–311, doi:10.1016/j.jeconom.2017.06.004.
- Michalopoulos, S. and Papaioannou, E. (2013). ‘Pre-colonial ethnic institutions and contemporary African development’, *Econometrica*, vol. 81(1), pp. 113–152.
- Michalopoulos, S. and Papaioannou, E. (2014). ‘National institutions and sub-national development in africa’, *Quarterly Journal of Economics*, vol. 129(1), pp. 151–213.

- Michalopoulos, S. and Papaioannou, E. (2015). ‘Further evidence on the link between precolonial political centralization and contemporary african development’, *Economics Letters*, vol. 126(1), pp. 57–62, doi:10.1016/j.econlet.2014.10.008.
- Michalopoulos, S. and Papaioannou, E. (2016). ‘The long-run effects of the scramble for africa’, *American Economic Review*, vol. 106(7), pp. 1802–1848.
- Michalopoulos, S. and Papaioannou, E. (2018). ‘Spatial patterns of development: A meso approach’, *Annual Review of Economics*, vol. 10, pp. 383–410.
- Michalopoulos, S. and Papaioannou, E. (2020). ‘Historical legacies and african development’, *Journal of Economic Literature*, vol. 58(1), pp. 53–128, doi:10.1257/jel.20181446.
- Mullainathan, S. and Spiess, J. (2017). ‘Machine learning: An applied econometric approach’, *Journal of Economic Perspectives*, vol. 31(2), pp. 87–106.
- Murdock, G.P. (1959). *Africa: Its Peoples and Their Culture History*, New York, NY: McGraw-Hill.
- Murdock, G.P. (1967). ‘Ethnographic atlas: A summary’, *Ethnology*, vol. 6(2), pp. 109–236, doi:10.2307/3772751.
- Narasimhan, V. and Weaver, J. (2024). ‘Polity size and local government performance: Evidence from india’, *American Economic Review*, vol. 114(11), pp. 3385–3426, doi:10.1257/aer.20220796.
- Nechaev, D., Zhizhin, M., Poyda, A., Ghosh, T., Hsu, F.C. and Elvidge, C. (2021). ‘Cross-sensor nighttime lights image calibration for dmsp/ols and snpp/viirs with residual u-net’, *Remote Sensing*, vol. 13(24).
- Pinkovskiy, M. (2017). ‘Growth discontinuities at borders’, *Journal of Economic Growth*, vol. 22, pp. 145–192.
- Pinkovskiy, M. and Sala-i Martin, X. (2016). ‘Lights, camera . . . income! illuminating the national accounts-household surveys debate’, *The Quarterly Journal of Economics*, vol. 131(2), pp. 579–631.
- Rossi-Hansberg, E. and Zhang, J. (2025). ‘Local gdp estimates around the world’, (33458), doi:10.3386/w33458.
- Sagi, O. and Rokach, L. (2018). ‘Ensemble learning: A survey’, *WIREs Data Mining Knowl Discov.*, vol. 8.
- Storeygard, A. (2016). ‘Farther on down the road: transport costs, trade and urban growth’, *Review of Economic Studies*, vol. 83(3), pp. 1263–1295.

- Wantchekon, L., Klačnja, M. and Novta, N. (2015). 'Education and human capital externalities: Evidence from colonial benin', *The Quarterly Journal of Economics*, vol. 130(2), pp. 703–758.
- Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.
- Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S. and Burke, M. (2020). 'Using publicly available satellite imagery and deep learning to understand economic well-being in africa', *Nature Communications*, vol. 11(2583).
- Young, A. (2012). 'The african growth miracle', *Journal of Political Economy*, vol. 120(4), pp. 696–739.

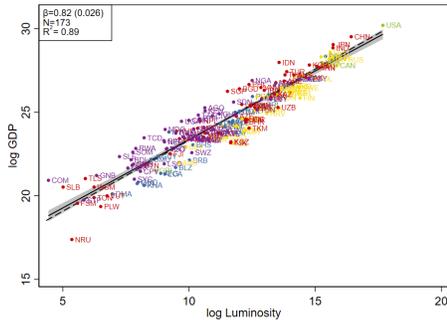
Figure 1: Mapping Nighttime Lights. Global VIIRS and DMSP in 2013



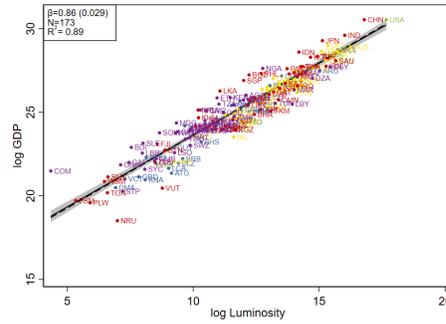
This figure gives illustrations of the pixel-level mapping of downgraded VIIRS to original DMSP in 2013. Panels (a)-(b) report estimates with our extremely randomized trees, ensemble method. The downgraded data presented here are from a model trained in 2012 (when VIIRS data is available for only part of the year) and predicted ‘out-of-sample’ in 2013 (when VIIRS data is available for the full year). Panels (c)-(b) give estimates with the sigmoid function of Li *et al.* (2020) that yields ‘DMSP-like’ downgraded VIIRS series. Panels (e) and (f) give the tabulation of the convolutional neural network method of Nechaev *et al.* (2021), ‘DVNL’. Panels (b), (d), and (f) give confusion matrices looking at the extensive margin of luminosity in 2013.

Figure 2: GDP-Luminosity Elasticity. Global Country-level Estimates

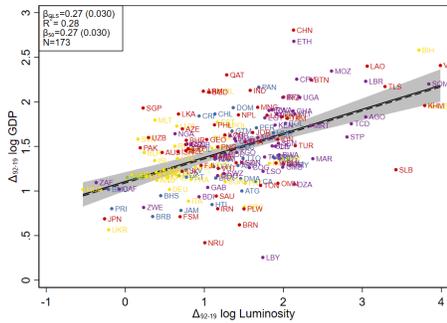
(a) Cross-sectional - log GDP 2005



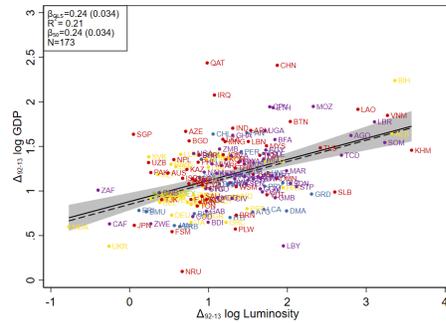
(b) Cross-sectional - log GDP 2015



(c) Long-Differences - Δ 1992-2019

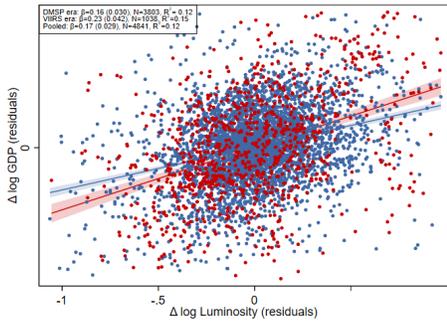


(d) Long-Differences - Δ 1992-2013

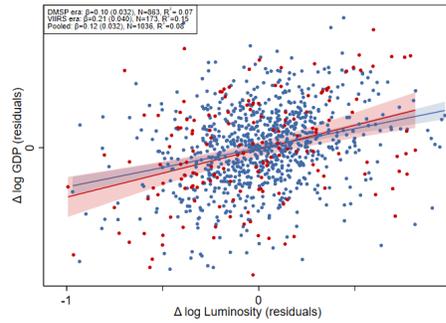


• North America • South America • Europe • Asia and Oceania • Africa

(e) Panel Estimates - Annual



(f) Panel Estimates - 5-year

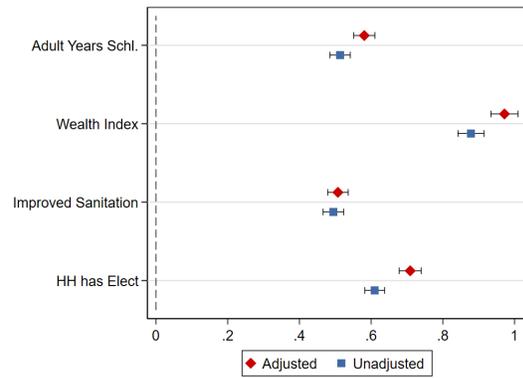
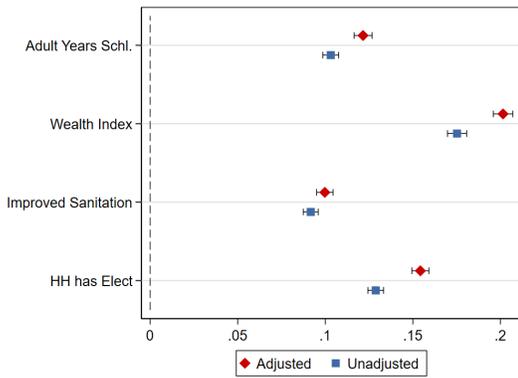


— DMSP era — VIIRS era

Panels (a) and (b) plot the cross-country association between the log of GDP and the log of nighttime lights (luminosity) across all countries in 2005 (DMSP period) and in 2015 (VIIRS period), alongside the LS regression line (solid line) and the median regression line (dashed). Panels (c) and (d) plot the long-difference association between log GDP and log luminosity over 2019-1992 and 2013-1992, respectively. Countries in panels (a) - (d) are colored according to their broad global region (North America, South America, Europe, Asia and Oceania, and Africa). Panels (e) and (f) plot the panel association between log GDP and the log nighttime lights (luminosity) across all countries and years during the period 1992-2019. Panel (e) uses all observations; panel (f) uses five-year average values. The panels plot the residuals of log GDP and log luminosity on country fixed-effects and year fixed-effects. Blue dots indicate country-year residuals during 1992-2012 (DMSP period) and the red dots indicate residuals during 2013-2019 (VIIRS period). All specifications use the (logarithm of the) newly compiled luminosity series harmonized VIIRS-DMSP after adjusting the DMSP data for top coding, sensor calibration, and blooming.

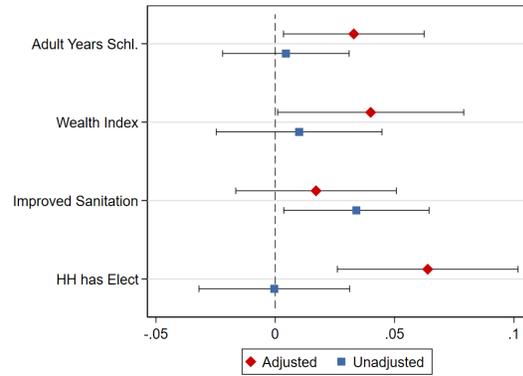
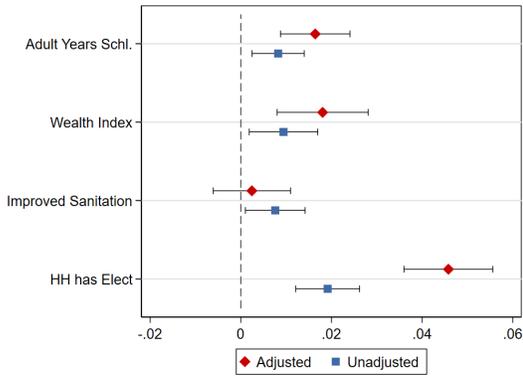
Figure 3: Local (standardized) Development-Luminosity Correlation. DHS Analysis

(a) Cross-sectional Estimates - Log Nightlights (b) Cross-sectional Estimates - Lit Indicator



(c) Panel Estimates - Log Nightlights

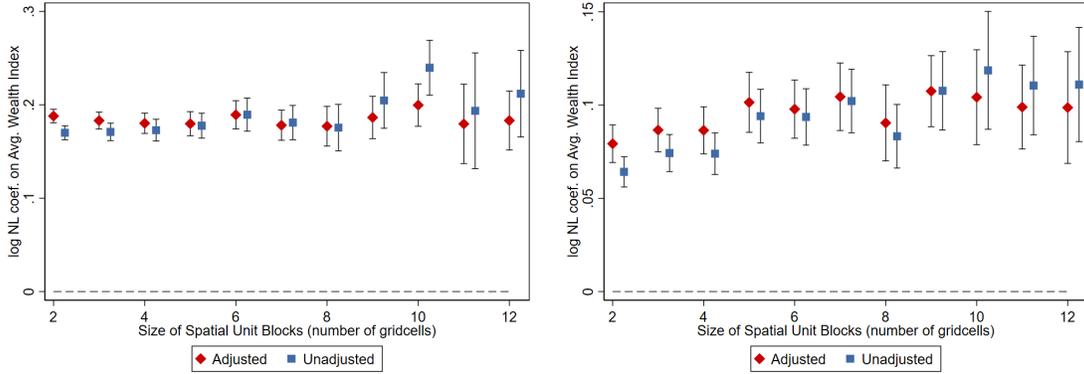
(d) Panel Estimates - Lit Indicator



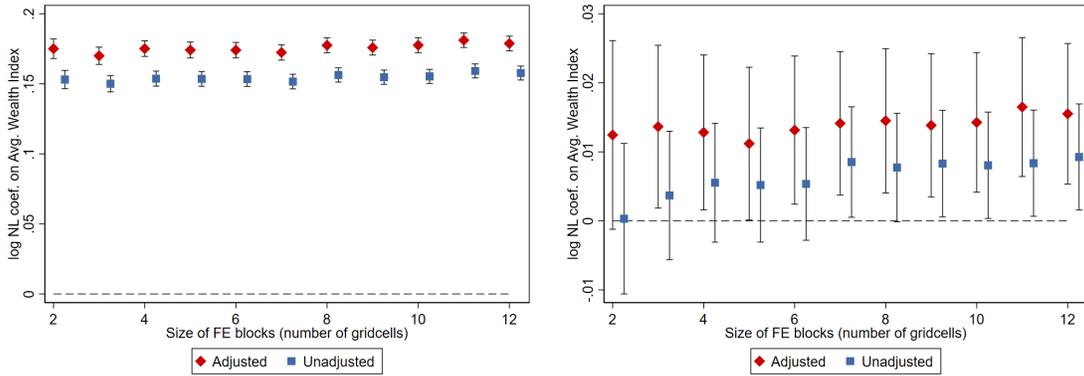
The figure plots coefficients from regressions associating proxies of development from the Demographic and Health Surveys (DHS) on night-time lights (luminosity). All outcomes are standardized to have a mean of zero and a standard deviation of one. Panels (a) and (c) use log luminosity; panels (b) and (d) use an indicator that equals one when the gridcell is lit and zero otherwise. Panels (a) and (b) control for country-survey-year fixed effects and log cell area. Panels (c) and (d) also include gridcell fixed effects. The unit of analysis is always gridcell-years. The bars give 95% confidence intervals based on standard errors clustered at the gridcell level.

Figure 4: Household Wealth-Luminosity Correlation. Further Evidence

(a) Cross-sectional, Varying Spatial Unit Size (b) Panel Estimates, Varying Spatial Unit size

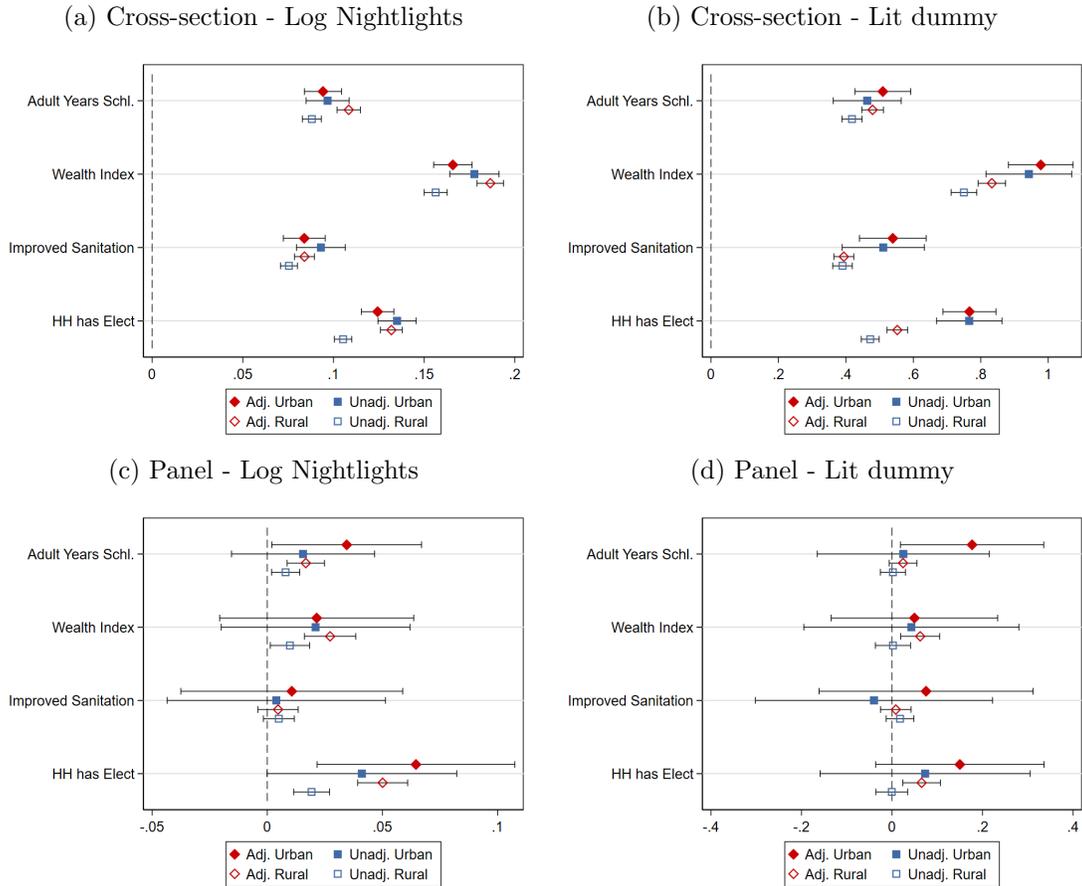


(c) Cross-sectional, Varying Fixed-Effects Size (d) Panel Estimates, Varying Fixed-Effects Size



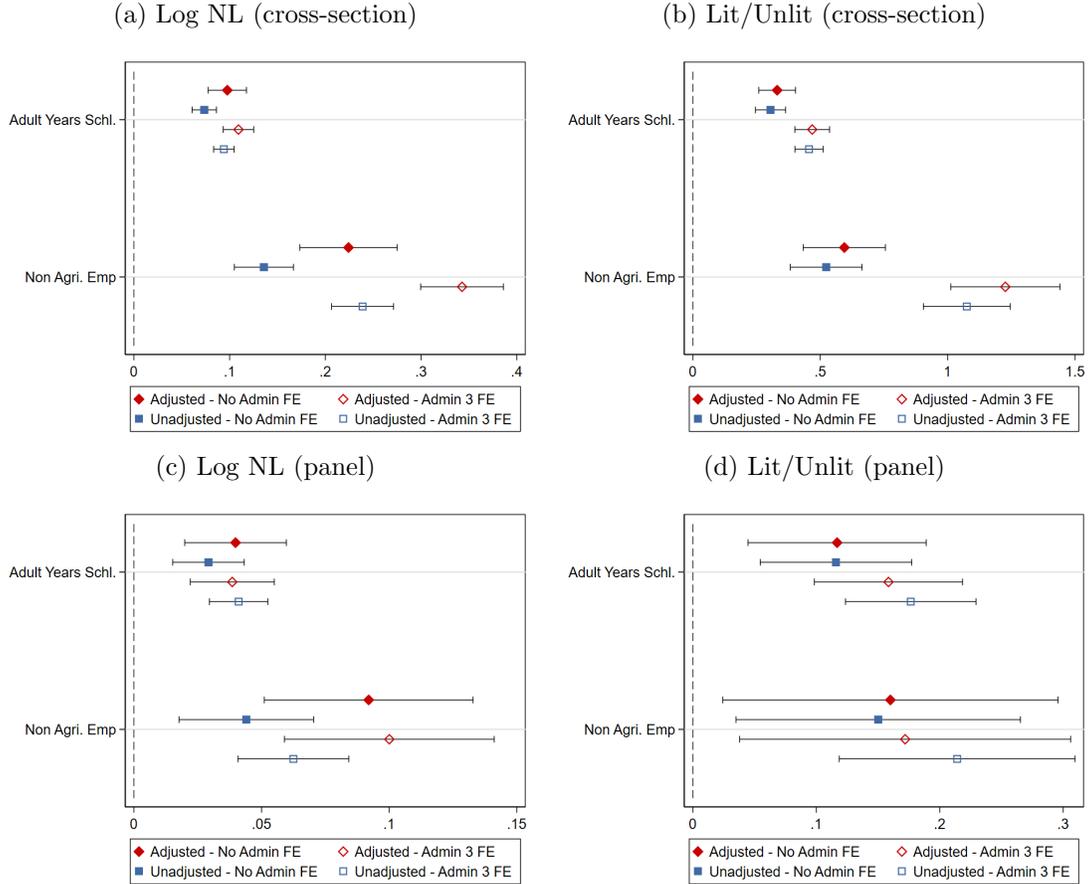
The figure plots coefficients from regressions associating the DHS Composite Wealth Index on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity holding gridcell size fixed and augmenting the specification with block fixed effects of increasing size. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of increasing sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the gridcell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

Figure 5: Local (standardized) Development - Luminosity Association. Urban and Rural Areas



Notes: This figure plots coefficients from regressions of standardized DHS measures on nightlights. All outcomes are standardized to have a mean of zero and a standard deviation of one. For the luminosity variables, panels (a) and (c) use log nightlights and panels (b) and (d) use an indicator equal to one for positive lights and zero otherwise. The top panels (a) and (b) control for country by year fixed effects. Panels (a) and (b) control for log gridcell area, while panels (c) and (d) control for gridcell fixed effects. In each panel, estimates from the subset of urban gridcells (red diamonds) are compared with estimates from the subset of rural gridcells (blue squares). The solid markers denote estimates using our corrected nightlight series, and hollow markers denote estimates using the unadjusted series. The bars represent 95% confidence intervals, and standard errors are clustered at the gridcell level.

Figure 6: Luminosity and (standardized) Local Development. Mozambique Census Analysis

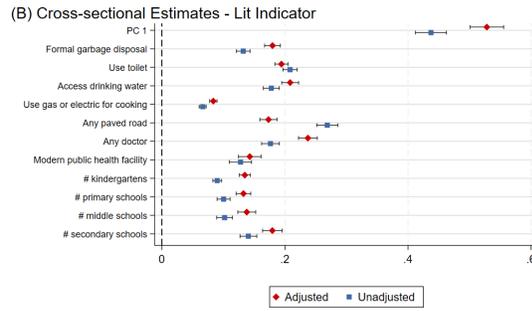
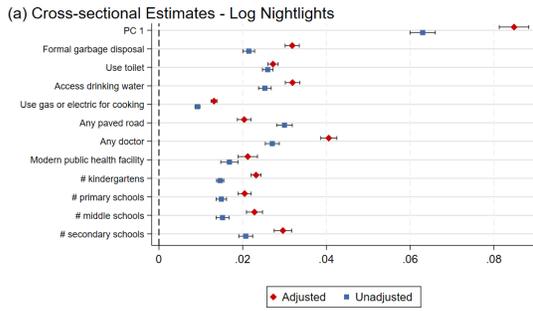


The figure plots coefficients from regressions associating mean years of schooling of individuals aged 15-39 and employment outside agriculture (in services, manufacturing, and mining) of individuals aged 15-24 with night-time lights luminosity across Mozambican localities, level-4 administrative units. The outcomes are standardized to have a mean of zero and a standard deviation of one. Panels (a) and (c) use the natural logarithm of nightlights, adding a small number. Panels (b) and (d) employ a luminosity indicator variable that equals one if the administrative unit is lit and zero otherwise. Schooling years are computed for all three Mozambican censuses (1997, 2007, and 2017). The share of non-agricultural employment is calculated using the 1997 and 2007 censuses. Panels (a) and (b) give cross-sectional estimates. Solid red diamonds and solid blue squares condition on the log admin area and year constants. The hollow circle/square also conditions on interactions between year constants and admin-3 fixed effects. Panels (c) and (d) give panel estimates with locality (admin-4) fixed-effects and year constants. The hollow circle/square specifications also condition on interactions between year constants and admin-3 fixed effects. The bars represent 95% confidence intervals, and standard errors are clustered at the admin-3 level.

Figure 7: Local (standardized) Development - Luminosity Association. Indonesia (PODES)

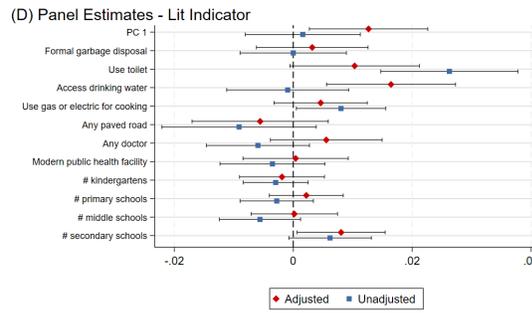
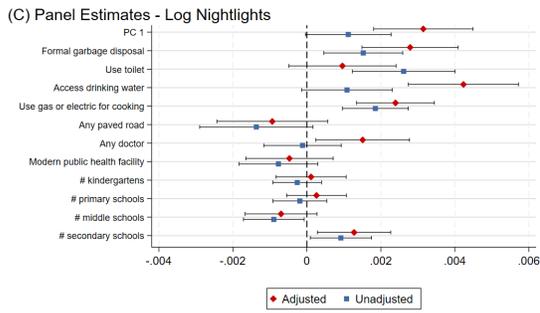
(a) Cross-section - Log Nightlights

(b) Cross-section - Lit dummy



(c) Panel - Log Nightlights

(d) Panel - Lit dummy

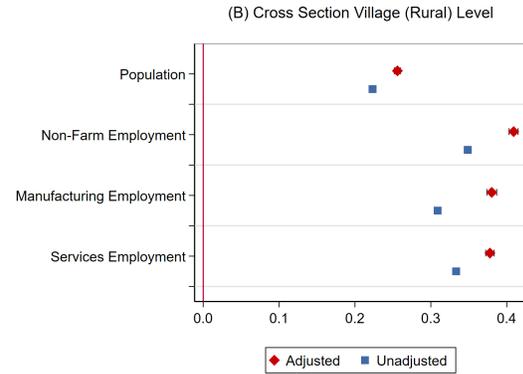
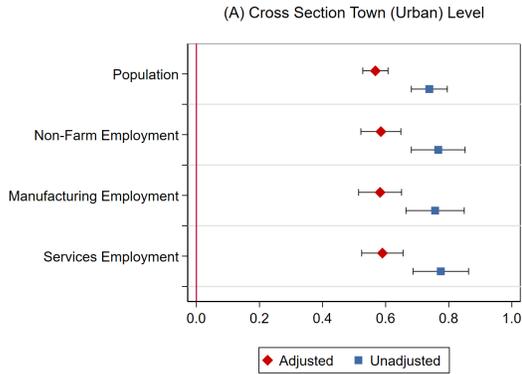


Notes: This figure plots coefficients from regressions of PODES measures on nightlights at the village (DESA) admin-4 level. The outcomes are standardized to have a mean of zero and a standard deviation of one. For the luminosity variables, panels (a) and (c) use log nightlights, adding a small number, and panels (b) and (d) use an indicator equal to one for positive lights and zero otherwise. Panels (a) and (b) control for log village area and admin-3 \times period fixed effects, while panels (c) and (d) control for village and admin-3 \times period fixed effects. The red diamonds denote estimates using our corrected nightlight series, and blue squares denote estimates using the unadjusted series. The bars represent 95% confidence intervals, and standard errors are clustered at the village level.

Figure 8: Local Development - Luminosity Association. India (SHRUG). Villages and Towns

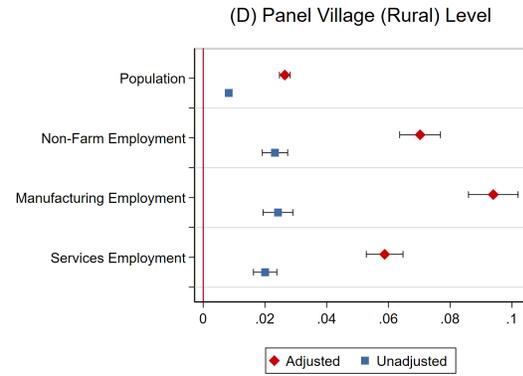
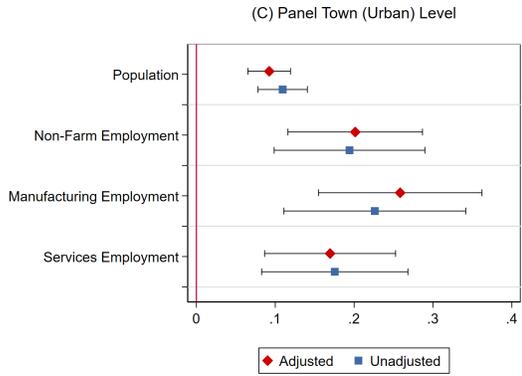
(a) Cross-Section. Town (urban) level

(b) Cross-Section. Village (rural) level



(c) Panel. Town (Urban) Level.

(d) Panel. Village (Rural) Level.



Notes: This figure plots coefficients from regressions of SHRUG measures on nightlights at the town (urban) and village (rural) level. For the luminosity variables, all panels use log nightlights. Panels (a) and (b) control for town (village) area and admin-3 (subdistrict) \times period fixed effects, while panels (c) and (d) control for town (village) and admin-3 \times period fixed effects. The red diamonds denote estimates using our corrected nightlight series, and blue squares denote estimates using the unadjusted series. The bars represent 95% confidence intervals, and standard errors are clustered at the town (village level).

Table 1: Precolonial Political Institutions and Contemporary African Development
 Michalopoulos and Papaioannou (2013)

	(1)	(2)	(3)	(4)	(5)	(6)
	Original Data	Original Data	Adjusted 07/08	Adjusted 07/08	Adjusted 92/23	Adjusted 92/23
Jurisdictional Hierarchy	0.1766*** (0.0501)		0.1471*** (0.0453)		0.1341*** (0.0512)	
Binary Political Centralization		0.3086*** (0.0972)		0.2376*** (0.0893)		0.2183** (0.0962)
Adj. R-squared	0.661	0.659	0.620	0.617	0.645	0.643
N	682	682	682	682	682	682
Mean Dep. Var.	-2.951	-2.951	-3.537	-3.537	-3.351	-3.351
SD Dep. Var.	1.697	1.697	1.407	1.407	1.474	1.474
Non-lit obs.	164	164	197	197	81	81
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The Table reports within-country OLS estimates associating regional development with pre-colonial ethnic institutions. The dependent variable is the log (0.01 + light density at night from satellite) at the ethnicity-country level. In columns (1) and (2), we use the original luminosity data employed by the authors (raw DMSP-OLS): average luminosity in 2007 and 2008. In columns (3) to (6), we use the adjusted DMSP-VIIRs series: average luminosity in 2007 and 2008 (columns (3) and (4)) and average luminosity between 1992 and 2023 (columns (5) and (6)). In columns (1), (3), and (5), we measure pre-colonial ethnic institutions using Murdock’s (1967) jurisdictional hierarchy beyond the local community index. In columns (2), (4), and (6), we use a binary political centralization index that is based on Murdock’s (1967) jurisdictional hierarchy beyond the local community variable. Following Gennaioli and Rainer (2007), this index takes on the value of zero for stateless societies and ethnic groups that were part of petty chiefdoms and one otherwise (for ethnicities that were organized as paramount chiefdoms and ethnicities that were part of large states). The set of control variables includes the distance of the centroid of each ethnicity-country area from the respective capital city, the distance from the sea coast, the distance from the national border, log (1 + area under water (lakes, rivers, and other streams)), log (surface area), land suitability for agriculture, elevation, a malaria stability index, a diamond mine indicator, and an oil field indicator. Below the estimates, we report in parentheses double-clustered standard errors at the country and the ethno-linguistic family dimensions.

Table 2: National Institutions and Subnational Development in Africa. Michalopoulos and Papaioannou (2014)

	Panel A: Original Data							
	Pixel Level			Ethnic-Specific RD 100 km			Relative Distance	
					3rd-order	4th-order	Close	Far
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Institutional Quality	0.107** (0.040)	0.025 (0.016)	0.099** (0.038)	0.017 (0.018)	0.012 (0.014)	0.005 (0.016)	0.114* (0.063)	-0.001 (0.023)
Mean Lit	0.124	0.124	0.123	0.123	0.123	0.123	0.218	0.086
Adj. R-squared	0.149	0.327	0.134	0.312	0.313	0.313	0.436	0.214
Observations	42709	42709	21289	21289	21289	21289	4655	10987
	Panel B: Adjusted 2007-2008							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Institutional Quality	0.090** (0.034)	0.027* (0.014)	0.080*** (0.030)	0.025* (0.013)	0.012 (0.013)	0.007 (0.014)	0.095* (0.052)	0.014 (0.016)
Mean Lit	0.092	0.092	0.091	0.091	0.091	0.091	0.164	0.056
Adj. R-squared	0.140	0.297	0.120	0.272	0.272	0.272	0.379	0.159
Observations	42709	42709	21289	21289	21289	21289	4655	10987
	Panel C: Adjusted 1992/2023							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Institutional Quality	0.108*** (0.037)	0.010 (0.022)	0.111*** (0.035)	0.027 (0.019)	0.005 (0.016)	-0.017 (0.022)	0.147*** (0.051)	0.008 (0.027)
Mean Lit	0.205	0.205	0.221	0.221	0.221	0.221	0.331	0.175
Adj. R-squared	0.191	0.317	0.185	0.325	0.327	0.327	0.414	0.269
Observations	42709	42709	21289	21289	21289	21289	4655	10987
Ethnicity FE	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Pop. dens. and area	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location and geography	No	No	No	No	No	No	Yes	Yes

Note: The table reports cross-sectional, within-ethnicity OLS, and regression discontinuity (RD) estimates associating regional development with contemporary national institutions, as reflected in the World Bank's Governance Matters rule of law index, averaged between 1996 and 2006 in areas of partitioned ethnicities at the grid-square level. The dependent variable is an indicator that takes on the value of one if the grid square (of 0.125 x 0.125 decimal degrees) is lit and zero otherwise. Columns (1) and (3) report cross-sectional specifications. All other columns give within-ethnicity estimates, including a vector of ethnicity fixed effects (constants not reported). In columns (3) to (8), we focus on ethnic areas within 50 kilometers of each side of the national border (a total of 100 kilometers). In Columns (5) and (6), we include a global (common to all partitioned ethnicities) RD polynomial in the distance of the centroid of each grid square to the national border, allowing the polynomial terms to differ on the two sides of the border. In columns (7) and (8), we focus on grid squares that are relatively close or far from the corresponding capital city on both sides of the border. We use the median relative distance to the capital within each country as a cutoff. We control for log population density in 2000 and log land area in all specifications. In columns (7) and (8), we further control for distance from the coast, distance from the national border, an indicator for pixels with a water body (lakes, rivers, and other streams), land suitability for agriculture, elevation, a malaria stability index, a diamond mine indicator, and an oil field indicator. Below the estimates, we report double-clustered standard errors in the country and ethnolinguistic family dimensions in parentheses.

Table 3: Regional Favoritism. Hodler and Raschky (2014)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Light _{ict}	Light0 _{ict}	Lightpc _{ict}				
Panel A: Original data							
Leader _{ict-1}	0.038*** (0.014)			0.019* (0.010)	0.061*** (0.010)	0.029** (0.013)	0.062** (0.025)
Leader _{ict}		0.039** (0.015)					
Leader _{ict-2}			0.041*** (0.013)				
Light _{ict-1}				0.400*** (0.024)	0.962*** (0.004)		
Pop _{ict}							-0.958*** (0.068)
Number of regions	38,427	38,427	38,427	38,427	38,427	36,204	37,475
Observations	690,495	690,495	689,870	652,362	652,362	619,208	673,382
R-squared	0.97	0.97	0.97	0.98	0.96	0.98	0.93
Panel B: Adjusted DMSP-VIIRS							
Leader _{ict-1}	0.058*** (0.016)			0.034*** (0.010)	0.049*** (0.008)	0.070*** (0.018)	0.093*** (0.025)
Leader _{ict}		0.044*** (0.016)					
Leader _{ict-2}			0.057*** (0.014)				
Light _{ict-1}				0.438*** (0.032)	0.976*** (0.003)		
Pop _{ict}							-0.888*** (0.089)
Number of regions	38,427	38,427	38,427	38,427	38,427	35,204	37,475
Observations	690,495	690,495	689,870	652,362	652,362	594,983	673,382
R-squared	0.98	0.98	0.98	0.98	0.97	0.97	0.94

Note: The Table reports fixed effect regressions (except for column (5), which is standard OLS) using annual data for subnational regions between 1992 and 2009. Panel A reports the original estimates in Hodler and Raschky (2014) using the raw DMSP-OLS. Panel B gives the results with the adjusted DMSP-VIIRS series. The original column 8 was omitted (the outcome was not luminosity). $Light_{ict}$ is the log of average nighttime light intensity plus 0.01. $Light0_{ict}$ is the log of average nighttime light intensity (without adding a constant). $Lightpc_{ict}$ is the log of nighttime light intensity per capita plus 0.01. $Leader_{ict}$ is a dummy variable equal to 1 if region i is the birth region of the political leader in country c in year t , and zero otherwise. Pop_{ict} is the log of regional population. All specifications control for region and country-year fixed effects, except for specification (5), which includes only country-year FE. Standard errors are adjusted for leader clustering. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

Supplementary Data Appendix

Illuminating the Global South

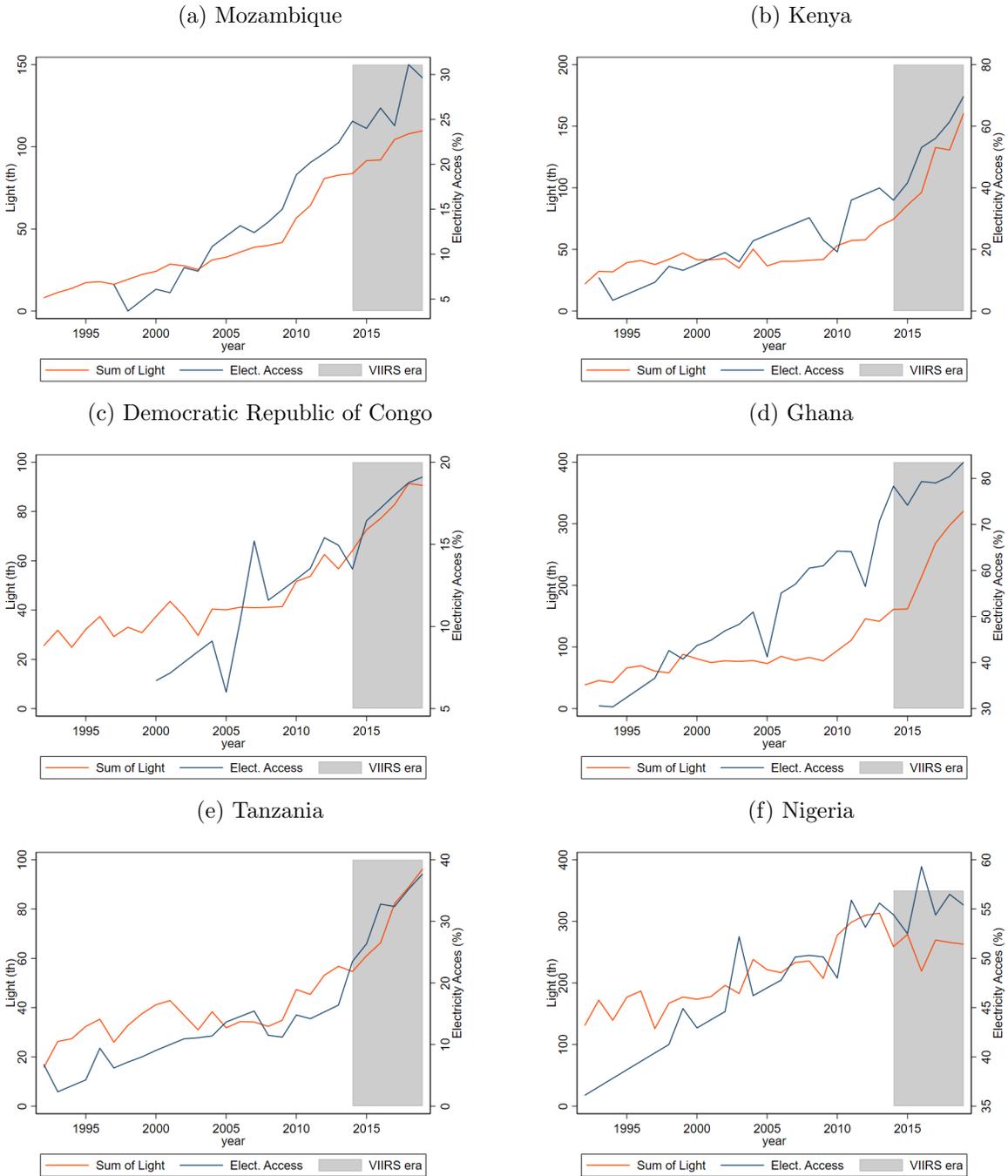
Giorgio Chiovelli Stelios Michalopoulos
Elias Papaioannou Tanner Regan

June 30, 2025

A Data and Methodology

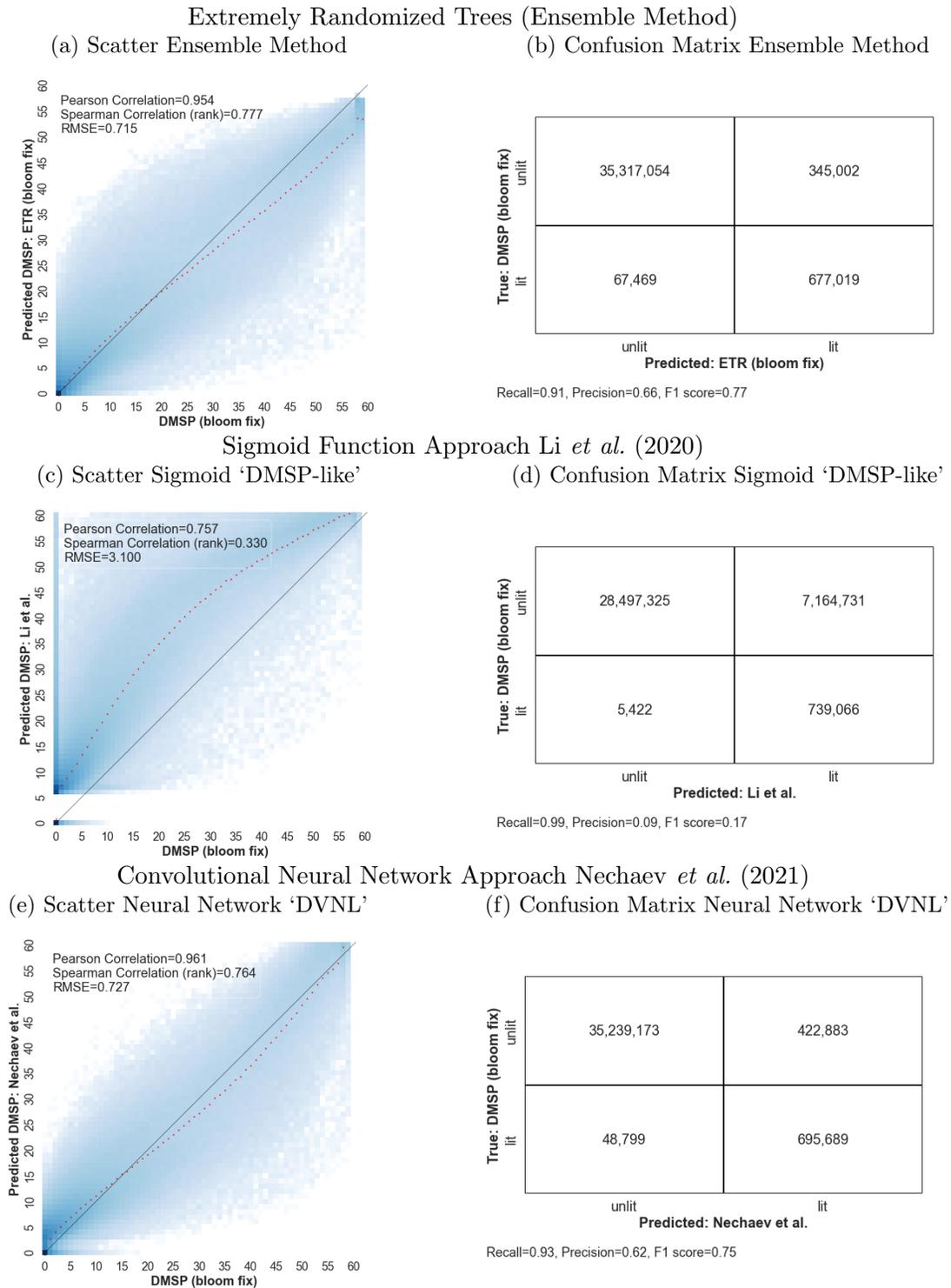
Appendix Figure A1 demonstrates the significant within-country over-time correlation between the newly-compiled harmonized luminosity series, based on the downgrading and merging of the VIIRS series to the adjusted for sensor quality, top-coding, and blooming DMSP series. The figure plots the harmonized nighttime light data alongside the share of the population with electricity in Mozambique, Kenya, the Democratic Republic of Congo, Ghana, Tanzania, and Nigeria using data from the World Bank's Development Indicators Database. Two results emerge. First, luminosity correlates with electricity access in all countries. Second, there was no major change in luminosity from 2012 to 2014, when we switched from the DMSP to the VIIRS satellite system.

Figure A1: Electricity Access and Harmonized Luminosity Series across African Countries



The figure plots trends in Electricity Access (as a share of the total population) and the sum of the harmonized and adjusted VIIRS-DSMP luminosity series (in thousands of DN) across six African countries. Data on electricity access come from World Bank's World Development Indicators Database.

Figure A2: Mapping Nighttime Lights. Africa VIIRS and DMSP in 2013



This figure gives illustrations of the pixel-level mapping of downgraded VIIRS to original DMSP in 2013. Panels (a)-(b) report estimates with our extremely randomized trees, ensemble method. The downgraded data presented here are from a model trained in 2012 (when VIIRS data is available for only part of the year) and predicted 'out-of-sample' in 2013 (when VIIRS data is available for the full year). Panels (c)-(b) give estimates with the sigmoid function of Li *et al.* (2020) that yields 'DMSP-like' downgraded VIIRS series. Panels (e) and (f) give the tabulation of the convolutional neural network method of Nechaev *et al.* (2021), 'DVNL'. Panels (b), (d), and (f) give confusion matrices looking at the extensive margin of luminosity in 2013.

Table A1: Feature Importances in the ERT models

Name	Feature Importances			
	No fix	Bloom only	Topcode only	Both fixes
VIIRS mean in 9x9 pixel window	0.112	0.121	0.088	0.092
VIIRS mean in 13x13 pixel window	0.107	0.136	0.078	0.081
VIIRS mean in 7x7 pixel window	0.105	0.111	0.105	0.104
VIIRS mean in 11x11 pixel window	0.101	0.089	0.099	0.075
VIIRS mean in 17x17 pixel window	0.089	0.074	0.081	0.088
VIIRS mean in 5x5 pixel window	0.087	0.103	0.085	0.093
VIIRS mean in 3x3 pixel window	0.073	0.057	0.068	0.046
VIIRS mean in 21x21 pixel window	0.062	0.087	0.075	0.084
VIIRS minimum in pixel	0.058	0.057	0.093	0.057
VIIRS mean in pixel	0.045	0.032	0.063	0.029
VIIRS median in pixel	0.036	0.044	0.071	0.031
Western Europe dummy	0.021	0.012	0.004	0.003
VIIRS maximum in pixel	0.020	0.022	0.028	0.023
Southern Europe dummy	0.011	0.006	0.002	0.002
VIIRS var. in 13x13 pixel window	0.011	0.004	0.004	0.005
VIIRS var. in 7x7 pixel window	0.007	0.008	0.004	0.006
VIIRS var. in 5x5 pixel window	0.007	0.005	0.003	0.004
VIIRS var. in 9x9 pixel window	0.007	0.004	0.004	0.006
VIIRS var. in 11x11 pixel window	0.005	0.004	0.005	0.005
Asiatic Russia dummy	0.005	0.002	0.002	0.001
VIIRS var. in 3x3 pixel window	0.004	0.003	0.003	0.003
Northern Europe dummy	0.004	0.001	0.002	0.004
VIIRS var. in 17x17 pixel window	0.003	0.003	0.004	0.006
Eastern Asia dummy	0.003	0.003	0.001	0.002
Eastern Europe dummy	0.003	0.001	0.001	0.001
Southern Asia dummy	0.002	0.001	0.001	0.001
VIIRS var. in 21x21 pixel window	0.001	0.002	0.004	0.006
Off-Coast dummy	0.001	0.002	0.001	0.022
Western Asia dummy	0.001	0.001	0.005	0.003
European Russia dummy	0.001	0.001	0.003	0.012
Northern America dummy	0.001	0.001	0.003	0.003
Northern Africa dummy	0.001	0.001	0.001	0.097
South America dummy	0.001	0.001	0.001	0.001
Australia/New Zealand dummy	0.001	0.001	0.000	0.000
Middle Africa dummy	0.001	0.000	0.000	0.000
Western Africa dummy	0.001	0.000	0.000	0.000
Eastern Africa dummy	0.001	0.000	0.000	0.000
Southeastern Asia dummy	0.000	0.000	0.006	0.003
Polynesia dummy	0.000	0.000	0.000	0.000
Micronesia dummy	0.000	0.000	0.000	0.000
Melanesia dummy	0.000	0.000	0.000	0.000
Southern Africa dummy	0.000	0.000	0.000	0.000
Central America dummy	0.000	0.000	0.000	0.000
Central Asia dummy	0.000	0.000	0.000	0.000
Caribbean dummy	0.000	0.000	0.000	0.000

Note: This table lists all features that are used in the Extremely Randomized Trees model to downgrade VIIRS. Each row corresponds to a feature and the columns show the feature importances for each of the four models that we implement. Feature importance is measured by the normalized total reduction of the model sum of squared errors from a feature, i.e. if the feature is removed from the trained model how much would model performance fall. The normalization ensures all feature importances sum to one.

B Supplementary Evidence

B.1 Cross Country Luminosity - GDP Elasticity. Further Evidence

This Appendix Section complements the cross-country correlation between GDP (Gross Domestic Product) and the newly compiled luminosity series that fuse the VIIRS data (2013 – 2023) to the DMSP data (1992 – 2013). The estimates, therefore, complement the (graphical) analysis of the cross-country GDP-luminosity elasticity in Section 3 of the main paper. First, we report additional cross-sectional patterns. Second, we give additional results of the within-country over time association between luminosity and GDP. Third, we report additional long-difference estimates.

Cross-sectional Patterns Panel A of Appendix Table B2 (global sample) and Appendix Table B3 (Africa sample) report pooled across years, cross-sectional regression estimates of the following form:

$$\ln GDP_{c,t} = \beta \ln NL_{c,t} + \gamma_a \ln(area)_c + \gamma_p \ln pop_{c,t} + \mu_t + \epsilon_{c,t} \quad (3)$$

$\ln(GDP)_{c,t}$ denotes the logarithm of current GDP (in PPP terms) of country c in year t ; $\ln(NL)_{c,t}$ is the log sum of the merged DMSP-VIIRS nightlights; $\ln(pop)_{c,t}$ is the log of population while $\ln(area)_c$ denotes log land area. All specifications include year constants, μ_t , that capture the increase in development and luminosity over time. Column (1) gives the lights-GDP elasticity across the entire period, 1992–2019; columns (2) and (3) look at the periods where only DMSP and only VIIRS are available, respectively. The luminosity GDP elasticity is about 0.61 for the global sample (Appendix Table B2) and 0.53 for the Africa sample (Appendix Table B3); the coefficient is highly significant, showing that luminosity is a good proxy of output in both samples. The fit is strong with an adjusted R^2 consistently above 0.9. As shown in columns (4)-(6) of Tables B2 and B3, the lights-GDP elasticity of the merged series adjusting for the deficiencies of DMSP is somewhat smaller as compared to the unadjusted DMSP data (about 0.61 vs 0.67 and 0.53 vs 0.62, respectively), though the fit is similarly strong. This alludes to a key finding throughout the paper, which is that the relative value of the corrected series is most substantial at fine levels of spatial analysis.

We then examine the correlation across African countries. Appendix Figure B3, panels (a) and (b) give a graphical illustration in 2005 (DMSP) and 2015 (VIIRS). As shown in Appendix Table B4, the GDP-luminosity elasticity is stable when we augment the specifications with broad regional constants (southern, central, western, eastern, and northern Africa) or exclude island nations (Comoros, Equatorial Guinea, Mauritius, Mayotte, Reunion, Sao Tome and Principe, and Seychelles).

Panel Estimates Appendix Table B3 - Panel *B* examines the dynamic association between nighttime lights and GDP for Africa. To do so, we augment the cross-sectional specification with country-fixed effects (δ_c). The regression equation reads:

$$\ln GDP_{c,t} = \beta \ln NL_{c,t} + \mu_t + \delta_c + \epsilon_{c,t} \quad (4)$$

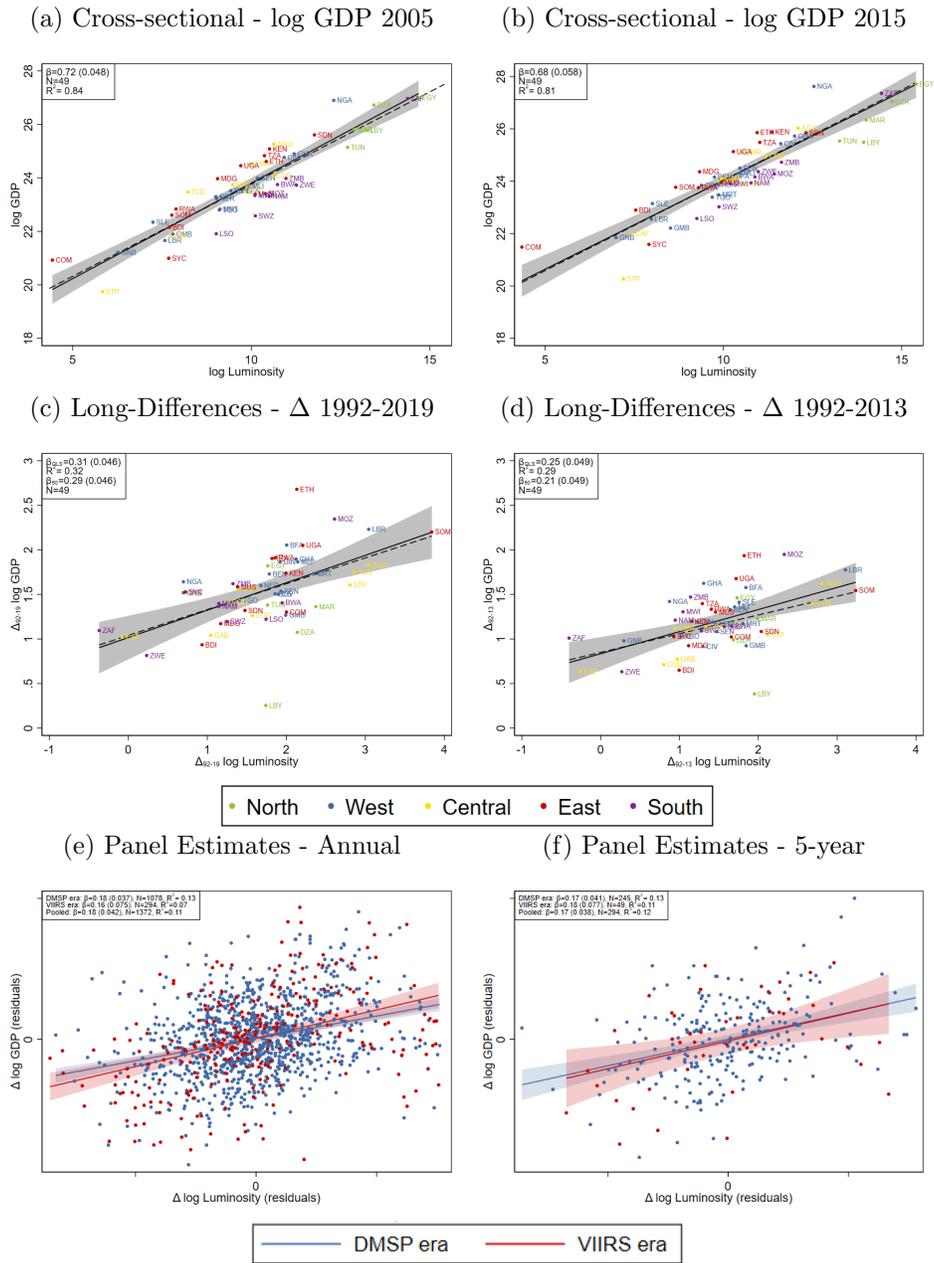
The regression analysis, therefore, complements the graphical illustrations in Figure B3 panels (c)-(d). GDP-luminosity elasticity with the newly compiled harmonized luminosity series that fuses the higher quality and more granular VIIRS series to the DMSP series adjusted for top-coding, blooming, and sensor quality in columns (1)-(3) hovers around 0.15 – 0.2. The elasticity is relatively stable when we drop island nations or interact the year constants with broad African region fixed effects to account for regional trends (Appendix Table B5). When we use the merged VIIRS to the unadjusted DMSP (in (4)-(6)), we obtain again a highly significant elasticity, which is somewhat higher (around 0.17–0.28). Appendix Figure B3, panels (e)-(f), plot the panel correlation across African countries, which is similar. Appendix Table B4-Panel (b) gives the corresponding regression estimates.

Long-Run Differences Appendix Table B3-Panel (c) plots the correlation between long-run changes in GDP and night-time lights in long differences for African countries. Taking the long-run differences (over 2019-1992 and 2013-1992) reduces noise in both GDP and nighttime lights, which is likely considerable in the annual frequency. The specification reads.

$$\Delta \ln GDP_c = \beta \Delta \ln NL_c + \delta \Delta \ln Pop_c + [\mu_r] + \epsilon_c \quad (5)$$

where $\Delta \ln(GDP)$ is the change in log GDP, $\Delta \ln(NL)_c$ the change in log nightlights, $\Delta \ln(Pop)_c$ the change in log population, and $\mu_{r(c)}$ are broad region r constants. Consistent with the graphical illustrations in Figure B3, panels (c)-(d), the elasticity hovers around 0.27 (columns (1)-(3)), close to the ones estimated for our global sample and those reported by Henderson *et al.* (2012). The median regression estimates, reported in Appendix Table B6, are pretty similar (around 0.2).

Figure B3: GDP-Luminosity Elasticity. Africa Country-level Estimates



Panels (a) and (b) plot the cross-country association between the log of GDP and the log of nighttime lights (luminosity) across African countries in 2005 (DMSP period) and in 2015 (VIIRS period), alongside the LS regression line (solid line) and the median regression line (dashed). Panels (c) and (d) plot the long-difference association between log GDP and log luminosity over 2019-1992 and 2013-1992, respectively. Countries in panels (a) - (d) are colored according to their broad African region (East, West, North, Central, and Southern). Panels (e) and (f) plot the panel association between log GDP and the log nighttime lights (luminosity) across African countries and years during the period 1992-2019. Panel (e) uses all observations; panel (f) uses five-year average values. The panels plot the residuals of log GDP and log luminosity on country fixed-effects and year fixed-effects. Blue dots indicate country-year residuals during 1992-2012 (DMSP period) and the red dots indicate residuals during 2013-2019 (VIIRS period). All specifications use the (logarithm of the) newly compiled luminosity series harmonized VIIRS-DMSP after adjusting the DMSP data for top coding, sensor calibration, and blooming.

Table B2: GDP - Luminosity Elasticity. Global Cross-Country Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) ln(GDP)	(5) ln(GDP)	(6) ln(GDP)
Panel A: Cross-sectional estimates						
ln(NL)	0.609*** (0.0186)	0.607*** (0.0190)	0.620*** (0.0205)	0.665*** (0.0196)	0.665*** (0.0205)	0.675*** (0.0203)
ln(area)	-0.0581* (0.0315)	-0.0468 (0.0304)	-0.101*** (0.0389)	-0.0881** (0.0417)	-0.0731* (0.0409)	-0.145*** (0.0465)
ln(Pop.)	0.443*** (0.0408)	0.430*** (0.0399)	0.486*** (0.0497)	0.441*** (0.0515)	0.421*** (0.0514)	0.508*** (0.0563)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	4841	3803	1038	4841	3803	1038
Countries	173	173	173	173	173	173
R ²	0.956	0.956	0.956	0.946	0.944	0.950
Panel B: Panel Estimates						
ln(NL)	0.174*** (0.0290)	0.151*** (0.0310)	0.100*** (0.0306)	0.218*** (0.0381)	0.184*** (0.0445)	0.130*** (0.0384)
ln(Pop.)	0.280*** (0.106)	0.346*** (0.132)	0.151 (0.179)	0.278*** (0.105)	0.322** (0.134)	0.0965 (0.182)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	4841	3803	1038	4841	3803	1038
Countries	173	173	173	173	173	173
R ²	0.996	0.997	0.999	0.996	0.997	0.999
within R ²	0.186	0.170	0.053	0.206	0.174	0.054
Panel C: Long-difference Estimates						
Δ ln(NL)	0.258*** (0.0342)	0.225*** (0.0389)	0.188*** (0.0493)	0.334*** (0.0348)	0.323*** (0.0465)	0.183*** (0.0400)
Δ ln(Pop.)	0.259** (0.126)	0.359** (0.165)	0.249 (0.289)	0.257** (0.125)	0.325** (0.164)	0.327 (0.297)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Countries	173	173	173	173	173	173
R ²	0.322	0.303	0.149	0.375	0.358	0.144

Note: This table presents regressions of (adjusted for PPP) log national GDP from the World Bank on nightlights for all available countries. Panel A shows cross-sectional estimates, while panels B and C shows panel and long differences estimates, respectively. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as the log sum of light in a country and the log sum of population. Panel A also includes the log area of the country. Panel A includes fixed effects for year, while Panel B includes fixed effects for year and country, and Panel C includes fixed effects for region. Standard errors in parentheses are clustered at the country level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: GDP - Luminosity Elasticity. Africa Cross-Country Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) ln(GDP)	(5) ln(GDP)	(6) ln(GDP)
Panel A: Cross-sectional estimates						
ln(NL)	0.529*** (0.0324)	0.540*** (0.0350)	0.496*** (0.0332)	0.626*** (0.0311)	0.635*** (0.0327)	0.597*** (0.0357)
ln(area)	-0.0205 (0.0441)	-0.00387 (0.0446)	-0.0808 (0.0524)	-0.0118 (0.0406)	0.00545 (0.0402)	-0.0745 (0.0487)
ln(Pop.)	0.414*** (0.0710)	0.386*** (0.0729)	0.509*** (0.0780)	0.370*** (0.0645)	0.345*** (0.0660)	0.456*** (0.0694)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1344	1056	288	1344	1056	288
Countries	48	48	48	48	48	48
R ²	0.931	0.929	0.931	0.941	0.941	0.934
Panel B: Panel Estimates						
ln(NL)	0.173*** (0.0446)	0.197*** (0.0390)	0.151** (0.0648)	0.277*** (0.0475)	0.278*** (0.0456)	0.173** (0.0750)
ln(Pop.)	0.888*** (0.284)	0.866*** (0.265)	1.378** (0.521)	0.747*** (0.250)	0.795*** (0.253)	1.372** (0.530)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1344	1056	288	1344	1056	288
Countries	48	48	48	48	48	48
R ²	0.993	0.995	0.999	0.993	0.995	0.999
within R ²	0.302	0.320	0.178	0.351	0.340	0.173
Panel C: Long-difference Estimates						
Δ ln(NL)	0.288*** (0.0610)	0.247*** (0.0482)	0.267*** (0.0746)	0.366*** (0.0882)	0.350*** (0.0764)	0.228** (0.0916)
Δ ln(Pop.)	0.423 (0.255)	0.574** (0.283)	0.342 (0.438)	0.357 (0.295)	0.433 (0.339)	0.425 (0.468)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Countries	48	48	48	48	48	48
R ²	0.491	0.433	0.544	0.476	0.418	0.488

Note: This table presents regressions of (adjusted for PPP) log national GDP from the World Bank on nightlights for all available countries. Panel A shows cross-sectional estimates, while panels B and C shows panel and long differences estimates, respectively. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as the log sum of light in a country and the log sum of population. Panel A also includes the log area of the country. Panel A includes fixed effects for year, while Panel B includes fixed effects for year and country, and Panel C includes fixed effects for region. Standard errors in parentheses are clustered at the country level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: GDP - Luminosity Elasticity. Sensitivity. Africa Cross-Sectional Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) ln(GDP)	(5) ln(GDP)	(6) ln(GDP)
Panel A: All African countries, adding Region FEs						
ln(NL)	0.605*** (0.0581)	0.613*** (0.0596)	0.606*** (0.0592)	0.737*** (0.0496)	0.747*** (0.0489)	0.710*** (0.0592)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1344	1056	288	1344	1056	288
Countries	48	48	48	48	48	48
R ²	0.949	0.949	0.949	0.961	0.962	0.956
Panel B: All African countries, adding Region FEs and Island Dummy						
ln(NL)	0.592*** (0.0647)	0.602*** (0.0674)	0.587*** (0.0615)	0.745*** (0.0604)	0.765*** (0.0587)	0.693*** (0.0634)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1344	1056	288	1344	1056	288
Countries	48	48	48	48	48	48
R ²	0.950	0.950	0.951	0.961	0.963	0.957
Panel C: Dropping island nations						
ln(NL)	0.529*** (0.0272)	0.544*** (0.0309)	0.483*** (0.0282)	0.607*** (0.0325)	0.616*** (0.0362)	0.574*** (0.0292)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1232	968	264	1232	968	264
Countries	44	44	44	44	44	44
R ²	0.927	0.923	0.933	0.933	0.929	0.937

Note: This table presents regressions of log national GDP from the World Bank on nightlights. All specifications include nightlights as the log sum of light in a country, the log area of the country, the log sum of population, and fixed effects for year. Each panel is done for a different sample or set of controls: panel A uses the sample of all African countries and adds region (central, east, north, south, and west) fixed effects, panel B adds region FEs and a dummy for island nations (Comoros, Mauritius, Mayotte, Reunion, Sao Tome and Principe, and Seychelles), and panel C restricts the sample to Africa countries excluding island nations with baseline controls. Columns 1-3 use nightlights that have been adjusted to fix blooming and topcoding including also cross-sensor calibration and the downgrading of VIIRS. Columns 4-6 use nightlights that have only been adjusted for cross-sensor calibration. Standard errors in parentheses are clustered at the country level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: GDP - Luminosity Elasticity. Sensitivity. Africa Panel Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1) ln(GDP)	(2) ln(GDP)	(3) ln(GDP)	(4) ln(GDP)	(5) ln(GDP)	(6) ln(GDP)
Panel A: All African countries						
ln(NL)	0.173*** (0.0446)	0.197*** (0.0390)	0.151** (0.0648)	0.277*** (0.0475)	0.278*** (0.0456)	0.173** (0.0750)
ln(Pop.)	0.888*** (0.284)	0.866*** (0.265)	1.378** (0.521)	0.747*** (0.250)	0.795*** (0.253)	1.372** (0.530)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1344	1056	288	1344	1056	288
Countries	48	48	48	48	48	48
R ²	0.993	0.995	0.999	0.993	0.995	0.999
within R ²	0.302	0.320	0.178	0.351	0.340	0.173
Panel B: exclude island nations						
ln(NL)	0.192*** (0.0462)	0.208*** (0.0409)	0.225*** (0.0745)	0.331*** (0.0412)	0.304*** (0.0462)	0.264*** (0.0848)
ln(Pop.)	1.052*** (0.308)	0.960*** (0.290)	1.843*** (0.524)	0.873*** (0.249)	0.888*** (0.263)	1.877*** (0.514)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1232	968	264	1232	968	264
Countries	44	44	44	44	44	44
R ²	0.991	0.993	0.998	0.992	0.994	0.998
within R ²	0.345	0.342	0.282	0.411	0.371	0.281
Panel C: region-year FEs and exclude islands						
ln(NL)	0.245*** (0.0414)	0.197*** (0.0429)	0.182** (0.0732)	0.319*** (0.0477)	0.265*** (0.0525)	0.240** (0.0934)
ln(Pop.)	1.109*** (0.321)	1.631*** (0.401)	0.927 (0.732)	1.127*** (0.293)	1.610*** (0.379)	1.108 (0.678)
Sample Years	1992-2019	1992-2013	2014-2019	1992-2019	1992-2013	2014-2019
Cntry-yrs	1232	968	264	1232	968	264
Countries	44	44	44	44	44	44
R ²	0.993	0.995	0.999	0.993	0.995	0.999
within R ²	0.380	0.414	0.117	0.401	0.429	0.128

Note: This table presents regressions of log national GDP from the World Bank on nightlights. Panel A uses all available countries in Africa, panel B excludes island nations (Comoros, Equatorial Guinea, Mauritius, Mayotte, Reunion, Sao Tome and Principe, and Seychelles), and panel C adds fixed effects for region-year (central, north, east, south, west) as well as excluding islands. All panels restrict the sample to every 5 years (starting from the sample base year). Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as the log sum of light in a country, the log sum of population, and fixed effects for year and country. Standard errors in parentheses are clustered at the country level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Africa National GDP Long Differences Estimates

	1992/93-2018/19	1992-2013		
	(1)	(2)	(3)	(4)
	DMSP ⁺ -VIIRS	DMSP-DMSP	DMSP ⁺ -DMSP ⁺	DMSP ⁺ -VIIRS
Panel A: All African countries				
$\Delta \ln(\text{NL})$	0.288*** (0.0610)	0.350*** (0.0764)	0.247*** (0.0482)	0.251*** (0.0521)
$\Delta \ln(\text{Pop.})$	0.423 (0.255)	0.433 (0.339)	0.574** (0.283)	0.577* (0.286)
Countries	48	48	48	48
R ²	0.416	0.332	0.349	0.363
Panel D: All African countries (median regression)				
$\Delta \ln(\text{NL})$	0.200*** (0.0559)	0.249** (0.123)	0.182*** (0.0540)	0.190*** (0.0395)
$\Delta \ln(\text{Pop.})$	0.379** (0.187)	0.633 (0.464)	0.724*** (0.222)	0.741** (0.279)
Countries	48	48	48	48
pseudo-R ²	0.313	0.223	0.250	0.255

Note: This table presents regressions of log national GDP from the World Bank on nightlights. Each panel is done for a different model: panel A reports OLS estimates, and panel B estimates median regressions. Column 1 takes the long difference 1992-2019 and columns 2-4 take the long difference from 1992-2013. Column 2 uses DMSP data before cleaning, and the remaining columns use the cleaned DMSP data (denoted DMSP⁺). All specifications are log GDP on log sum of nightlights and log population in a country controlling for region FEs. Standard errors are robust to heteroskedasticity. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.2 Local African Development (DHS Analysis). Further Evidence

This Appendix Section complements the regional analysis linking development outcomes from the Demographic and Health Surveys with the newly compiled harmonized VIIRS-DSMP luminosity series in Section 4 of the main paper.

Data Aggregation Appendix Figure B4 illustrates the aggregation of the DHS data to gridcells of 0.25×0.25 degrees, roughly $27.8km$ by $27.8km$ at the equator, zooming into central Mozambique. Circles report DHS enumeration areas or clusters, typically (large) villages, towns, and cities. The map also displays grid cells (black lines) to which we aggregate the underlying DHS data. The gridcell size choice (0.25×0.25 degrees) is a product of DHS displacement. The DHS cluster GPS points are displaced by up to $10km$ to maintain anonymity. So by choosing gridcells that are $28km$ wide and tall, we ensure that any DHS cluster located at the center of a gridcell will be assigned to that gridcell even after displacement. This is especially important when building the panel data, since grid cells that are too small will not have repeat observations due to random DHS displacement. Of course, even with our choice of gridcell size, it is always possible that we ‘miss’ DHS clusters that repeat over time but are displaced to adjacent gridcells.

Sample Appendix Table B7 gives the survey years for all counties in the DHS analysis. The 34 countries come from all parts of Africa; some are landlocked (e.g., Burkina Faso, Central African Republic, Burundi), some coastal (e.g., Mozambique, Sierra Leone), and a few are island nations (Madagascar, Comoros). The sample spans relatively richer and poorer countries and includes former British, French, Portuguese, and Belgian colonies.

Summary Statistics Appendix Table B8 gives summary statistics for the four outcomes from the DHS (years of schooling, composite wealth index, access to improved sanitation, and access to electricity); the luminosity series (with and without the adjustments in the DMSP for top-coding and blooming), and area.

Preliminary Cross-Sectional Patterns Appendix Figure B5 illustrates the significant, although far from perfect, cross-sectional association between luminosity and the four proxies of local development. The four panels plot the histogram of mean years of schooling, the composite wealth index, access to improved sanitation, and electricity, for lit and unlit small gridcells, netting out country-year fixed effects and conditioning on log gridcell area. The histogram for lit gridcells (red vertical bars) is evidently to the right of the analogous one for unlit gridcells (blue bars). The two distributions are different for all development outcomes, especially with education and household wealth. Nonetheless, there is also overlap, as the binary transformation of luminosity

cannot fully capture the wide spatial variation in well-being within African countries.

Baseline Regression Analysis Appendix Tables B9 and B10 report cross-sectional and panel estimates linking the four DHS outcome measures (education, composite wealth index, access to improved sanitation, and electricity) to log luminosity and the lit indicator. In these tables, we do not standardize the dependent variables (as we do in the graphical illustrations in the main paper). In both tables, panels *A* and *C* report estimates with the newly compiled, adjusted for top-coding and blooming DMSP series merged to the downgraded VIIRS. For comparability, panels *B* and *D* give the results with the merged VIIRS-DMSP series without adjusting the pre-2013 DMSP series for top-coding and blooming. The two tables, therefore, report the regression analogs to the coefficients reported in Figure 3. The cross-sectional specifications in Panel *C* of Appendix Table B9 suggest a significant increase of 1 (one standard deviation) in the DHS standardized composite wealth index and about 1.7 more schooling years between lit and unlit areas. The panel estimates, while yielding statistically significant correlations, imply smaller effects. The results in panel *C* of Appendix Table B10 imply an increase of 0.1 years of schooling and 0.04 in the composite wealth index for administrative units turning from unlit to lit.

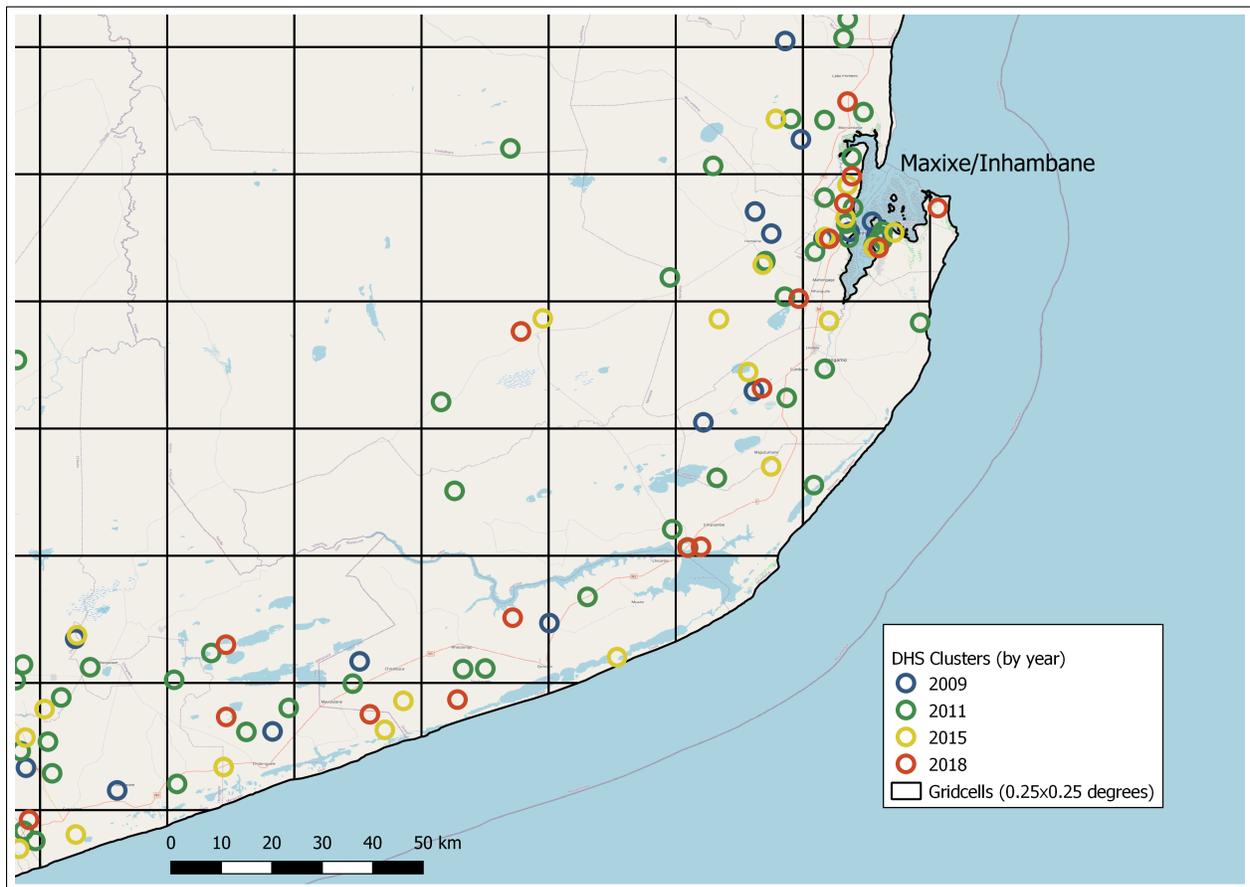
Further Evidence A. Varying Spatial Unit Appendix Figures B6, B7, and B8 panels (a) and (b) explore the association between luminosity and education, access to improved sanitation, and electricity varying the size of the spatial unit. The results, therefore, complement the analysis in Figure 4 - Panels (a)-(b), in the main paper with the composite wealth index. The patterns with the three development proxies are similar to the ones with the DHS wealth index, based on household assets. First, cross-sectionally, differences in education, access to improved sanitation and electricity correlate with log luminosity across both small, medium, and larger areas. Besides, the coefficients are similar. Second, changes in log luminosity within gridcells over time correlate with changes in the three development proxies, although the dynamic correlations are weaker and more noisy. Third, the use of the harmonized and adjusted merged VIIRS-DSMP data yield stronger and with smaller standard errors correlations, especially in the panel specifications.

Further Evidence B. Varying Localized Variation Panels (c)-(d) of Appendix Figures B6, B7, and B8 report coefficients on log luminosity with education, access to electrification, and to improved sanitation, varying the localized variation with fixed-effects of varying coarseness. These specifications are, thus, similar to the ones in panels (c) and (d) of Figure 4 the composite wealth index as the outcome variable. The qualitative takeaways are mostly similar as with the composite wealth index. Notably, the luminosity years of schooling correlation turns significant only when adding mid-size fixed effects, of blocks 8×8 cells or larger. Besides, the luminosity access to electrification correlation is highly significant with the newly harmonized series, even

when exploiting very granular variability with fine fixed-effects, illustrating again the reduction in measurement error from our ensemble method that fuses a downgraded vintage of VIIRS into the DMSP after adjusting them for top-coding, blooming, and sensor calibration (see Section 2).

Further Evidence C. Rural-Urban Figure 5 plots the coefficients on luminosity distinguishing across DHS respondents in rural and urban households (using the DHS classification). Panels (a) and (b) give cross-sectional estimates with log luminosity and the lit indicator (conditioning for log land area and country-survey-year constants). Panels (c) and (d) give panel estimates (with unit fixed effects and country-survey-year fixed effects). Red markers [diamonds] give the estimates with the harmonized and adjusted VIIRS-DMSP series, while blue markers [squares] report analogous estimates with the unadjusted for top-coding and blooming series. The cross-sectional analysis suggests that within-country across space) differences in luminosity correlate significantly with schooling, access to public goods, and household assets (as captured in the composite wealth index). The estimates appear similar in urban and rural locations. Besides, the adjustment for top-coding and blooming slightly improves the coefficient’s magnitude. The panel specifications yield somewhat different patterns. First, the coefficients are statistically indistinguishable from zero when one uses the ‘raw’ luminosity series. In contrast, the coefficients are higher and pass standard statistical significance thresholds with the newly compiled harmonized and adjusted for top-coding, blooming, and sensor calibration lights series. Second, the coefficients in the urban sample of survey respondents are always larger than the ones in the rural sample, telling that the local development-luminosity nexus is stronger in urban areas.

Figure B4: Aggregation Example. DHS Clusters and Gridcells



The figure maps the southern coast of Mozambique near the city of Maxixe/Inhambane. The circles represent DHS clusters (enumeration areas), colored by the survey year. The grid gives the cells at which we aggregate and analyze the DHS data. The background imagery is from OpenStreetMap.

Table B7: DHS Sample

	Country	N years	N cell-years	Sample years
1	Angola	3	486	2006, 2011, 2015
2	Benin	4	366	1996, 2001, 2012, 2017
3	Burkina Faso	7	1168	1993, 1999, 2003, 2010, 2014, 2017, 2021
4	Burundi	3	129	2010, 2012, 2016
5	Cameroon	4	836	2004, 2011, 2018, 2022
6	CAR	1	64	1994
7	Chad	1	330	2014
8	Comoros	1	10	2012
9	Cote d'Ivoire	4	658	1994, 1998, 2012, 2021
10	DRC	2	586	2007, 2013
11	Egypt	7	839	1992, 1995, 2000, 2003, 2005, 2008, 2014
12	Gabon	2	254	2012, 2019
13	Ghana	8	1303	1993, 1998, 2003, 2008, 2014, 2016, 2019, 2022
14	Guinea	5	730	1999, 2005, 2012, 2018, 2021
15	Kenya	6	1348	2003, 2008, 2014, 2015, 2020, 2022
16	Lesotho	3	157	2004, 2009, 2014
17	Liberia	7	609	2007, 2009, 2011, 2013, 2016, 2019, 2022
18	Madagascar	6	1453	1997, 2008, 2011, 2013, 2016, 2021
19	Malawi	7	809	2000, 2004, 2010, 2012, 2014, 2015, 2017
20	Mali	8	1487	1996, 2001, 2006, 2010, 2012, 2015, 2018, 2021
21	Morocco	1	223	2003
22	Mozambique	5	1173	2009, 2011, 2015, 2018, 2022
23	Namibia	3	572	2000, 2006, 2013
24	Niger	4	615	1992, 1998, 2012, 2021
25	Nigeria	7	2672	2003, 2008, 2010, 2013, 2015, 2018, 2021
26	Rwanda	5	208	2005, 2008, 2010, 2014, 2019
27	Senegal	13	1646	1993, 1997, 2005, 2008, 2010, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2023
28	Sierra Leone	4	392	2008, 2013, 2016, 2019
29	Swaziland	1	32	2006
30	Tanzania	8	2310	1999, 2003, 2007, 2010, 2012, 2015, 2017, 2022
31	Togo	3	220	1998, 2013, 2017
32	Uganda	8	1232	2000, 2006, 2009, 2010, 2011, 2014, 2016, 2018
33	Zambia	3	867	2007, 2013, 2018
34	Zimbabwe	4	814	1999, 2005, 2010, 2015

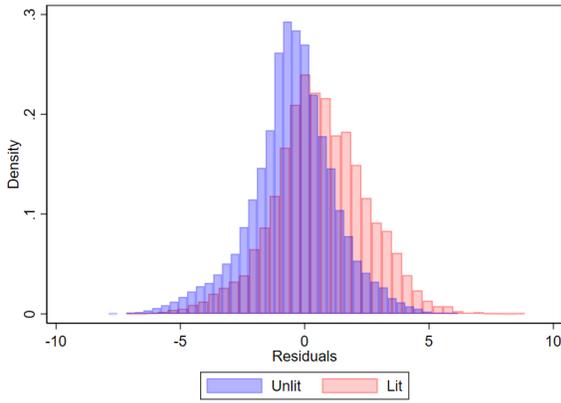
Table B8: Descriptive Statistics - Nightlights and DHS

	Min	p10	p50	p90	Max	Mean	SD	N
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	796.00	731,690.00	960.29	9,196.85	26598
Sensor calibration only	0.00	0.00	25.00	2,292.00	57,906.00	1,229.96	4,473.76	26598
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	6.68	13.50	1.75	3.19	26598
Log of sensor calibration only	-0.69	-0.69	3.24	7.74	10.97	2.94	3.56	26598
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.44	0.50	26598
Sensor calibration only (dummy)	0.00	0.00	1.00	1.00	1.00	0.56	0.50	26598
Gridcell area in km ²	4.00	693.96	754.60	768.67	769.31	727.77	102.72	26598
Log of gridcell area in km ²	1.39	6.54	6.63	6.64	6.65	6.56	0.31	26598
Years of adult (15-39) schooling	0.00	0.82	5.09	9.02	13.83	5.02	3.00	21681
Wealth Index	1.00	1.33	2.45	3.99	5.00	2.56	0.97	23224
Share of Households with Improved Sanitation	0.00	0.05	0.80	1.00	1.00	0.66	0.35	26515
Share of Households with electricity	0.00	0.00	0.04	0.88	1.00	0.25	0.33	26516

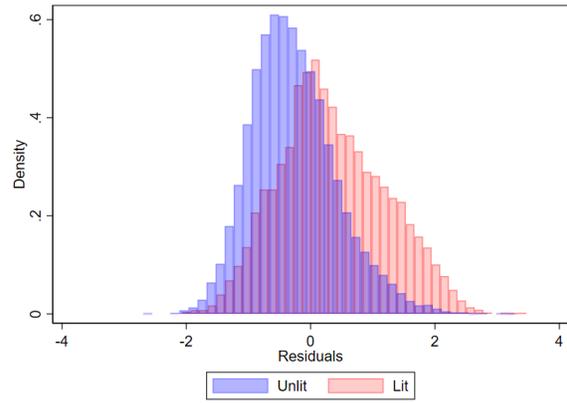
Note: This table presents summary statistics for the nightlights and DHS database. The observations are at the grid level. The total number of countries available is 34. For the log of nightlights we take ((half of the minimum value of positive NL) + NL) before taking the log.

Figure B5: DHS Development Outcomes across Lit and Unlit gridcells

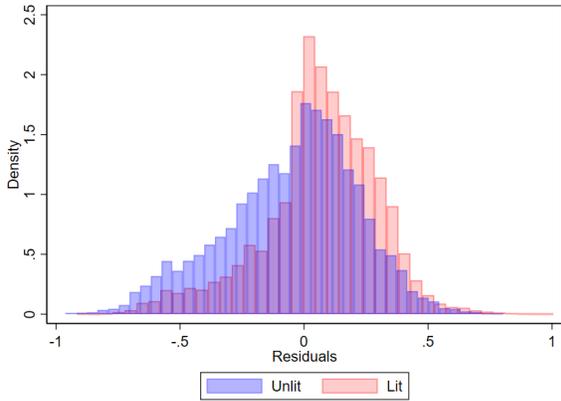
(a) Mean Adult (15-39) years schl.



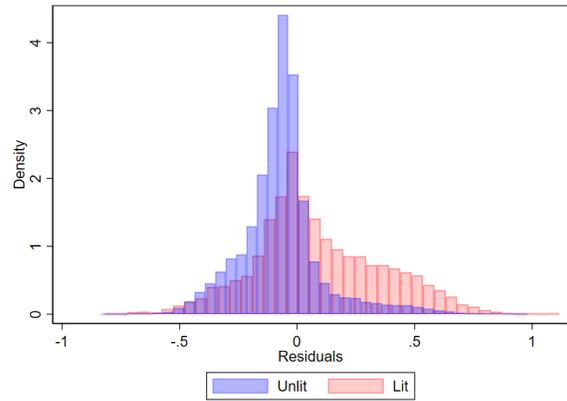
(b) Avg. Wealth Index



(c) Improved Sanitation



(d) HH has Elect.



The figure plots the histograms of four proxies of local development for lit and unlit 0.25×0.25 gridcells after netting country-survey-year fixed-effects and log land area. Panel (a) gives average years of schooling for individuals between 15 and 39 years old. Panel (b) gives the DHS composite wealth index, based on household assets. Panel (c) gives the tabulations on access to improved sanitation. Panel (d) gives the histograms for household access to electricity.

Table B9: DHS Cross-sectional Estimates

	(1)	(2)	(3)	(4)
	Mean Adult (15-39) years schl.	Avg. Wealth Index	Improved Sanitation	HH has Elect.
Panel A: Log sum of nightlights - sensor, blooming, & topcode fixes				
ln(minNL/2+NL)	0.365*** (0.00786)	0.196*** (0.00276)	0.0349*** (0.000840)	0.0516*** (0.000829)
Obs	21681	23224	26515	26516
Obs(NL=0)	12168	13102	14944	14945
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.581	0.396	0.430	0.578
Panel B: Log sum of nightlights - sensor calibration only				
ln(minNL/2+NL)	0.310*** (0.00694)	0.171*** (0.00272)	0.0321*** (0.000762)	0.0430*** (0.000763)
Obs	21681	23224	26515	26516
Obs(NL=0)	9630	10013	11594	11594
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.572	0.379	0.434	0.558
Panel C: lit dummy - sensor, blooming, & topcode fixes				
1(NL>0)	1.745*** (0.0454)	0.947*** (0.0189)	0.178*** (0.00512)	0.237*** (0.00525)
Obs	21681	23224	26515	26516
Obs(NL=0)	12168	13102	14944	14945
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.546	0.298	0.412	0.513
Panel D: lit dummy - sensor calibration only				
1(NL>0)	1.542*** (0.0433)	0.856*** (0.0181)	0.173*** (0.00513)	0.204*** (0.00471)
Obs	21681	23224	26515	26516
Obs(NL=0)	9630	10013	11594	11594
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.531	0.259	0.409	0.486

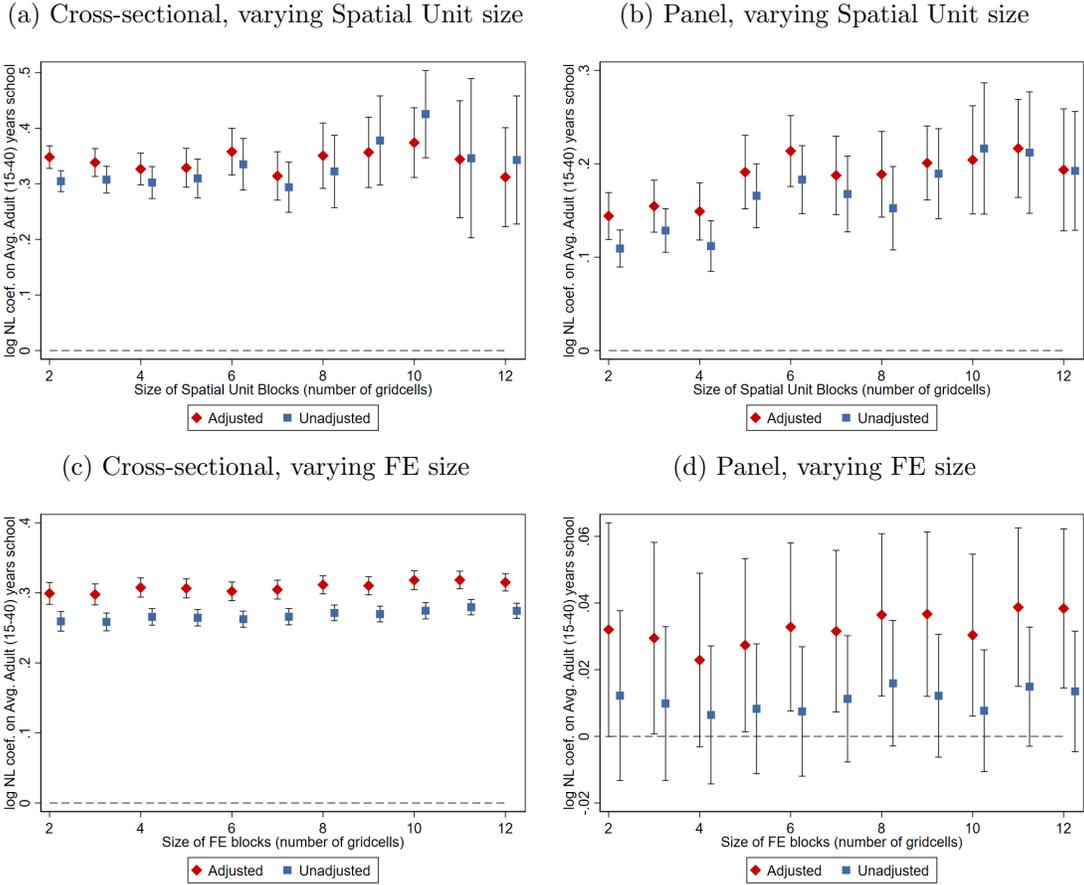
Note: This table presents regressions of economic indicators from the DHS on nightlights. Observations are 0.25 x 0.25 degree gridcell-years. Each panel is done with a different definition for nightlights: Panel A uses the log sum of nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, panel B uses the log of sum nightlights that have only been adjusted for cross-sensor calibration, Panel C uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, and Panel D uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration. All specifications include the log area of the cell, and fixed effects for country-year. Standard errors in parentheses are clustered at the gridcell level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B10: DHS Panel Estimates

	(1) Mean Adult (15-39) years schl.	(2) Avg. Wealth Index	(3) Improved Sanitation	(4) HH has Elect.
Panel A: Log sum of nightlights - sensor, blooming, & topcode fixes				
ln(minNL/2+NL)	0.0493*** (0.0117)	0.0176*** (0.00499)	0.000858 (0.00152)	0.0153*** (0.00167)
Obs	17454	19089	22472	22472
Obs(NL=0)	8777	9818	11674	11674
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.888	0.781	0.799	0.791
Panel B: Log sum of nightlights - sensor calibration only				
ln(minNL/2+NL)	0.0248*** (0.00882)	0.00918** (0.00376)	0.00266** (0.00118)	0.00640*** (0.00120)
Obs	17454	19089	22472	22472
Obs(NL=0)	6674	7167	8729	8729
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.888	0.781	0.799	0.790
Panel C: lit dummy - sensor, blooming, & topcode fixes				
1(NL>0)	0.0988** (0.0452)	0.0390** (0.0194)	0.00598 (0.00601)	0.0213*** (0.00646)
Obs	17454	19089	22472	22472
Obs(NL=0)	8777	9818	11674	11674
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.888	0.781	0.799	0.790
Panel D: lit dummy - sensor calibration only				
1(NL>0)	0.0133 (0.0406)	0.00974 (0.0173)	0.0119** (0.00544)	-0.000145 (0.00538)
Obs	17454	19089	22472	22472
Obs(NL=0)	6674	7167	8729	8729
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	gcell-yr	gcell-yr	gcell-yr	gcell-yr
R ²	0.888	0.781	0.799	0.790

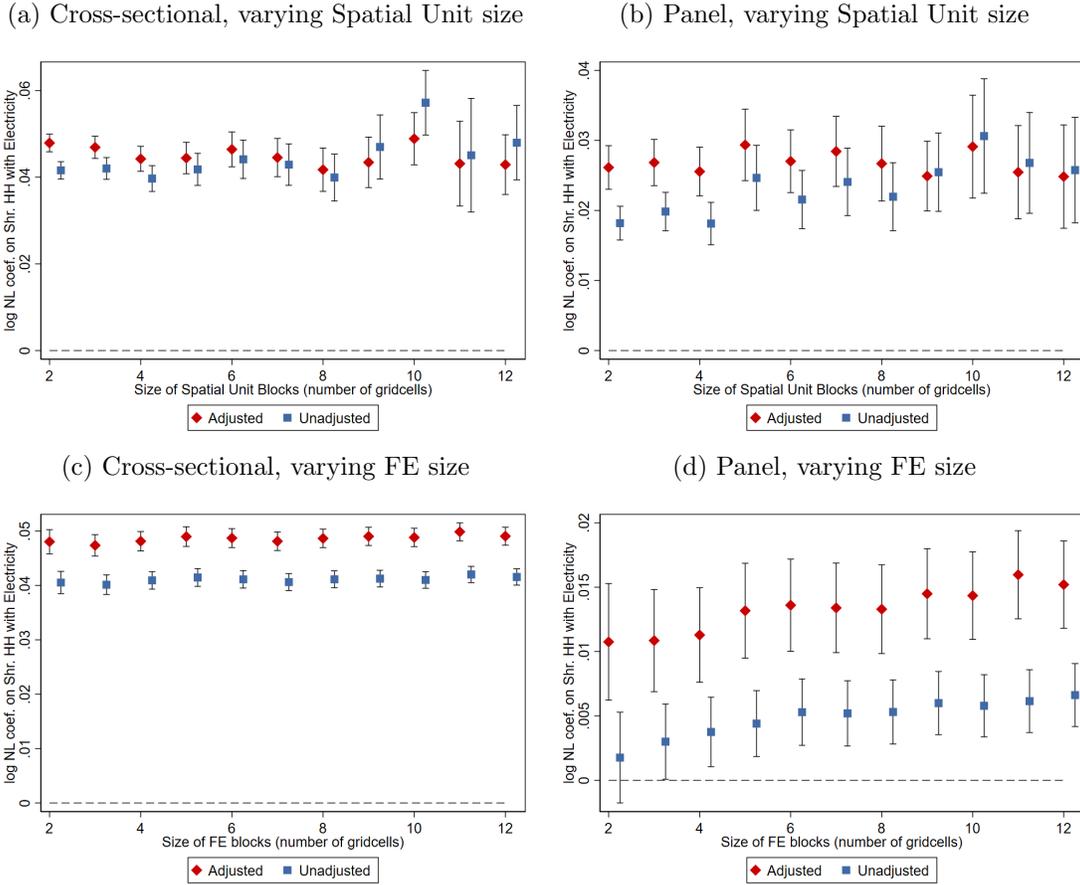
Note: This table presents regressions of economic indicators from the DHS on nightlights. Observations are 0.25 x 0.25 degree gridcell-years. Each panel is done with a different definition for nightlights: Panel A uses the log sum of nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, panel B uses the log of sum nightlights that have only been adjusted for cross-sensor calibration, Panel C uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration and to fix blooming and topcoding, and Panel D uses a dummy if the cell is lit based on nightlights that have been adjusted for cross-sensor calibration. All specifications include fixed effects for country-year, and gridcell. Standard errors in parentheses are clustered at the gridcells level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure B6: Schooling-Luminosity Correlation. Further Evidence



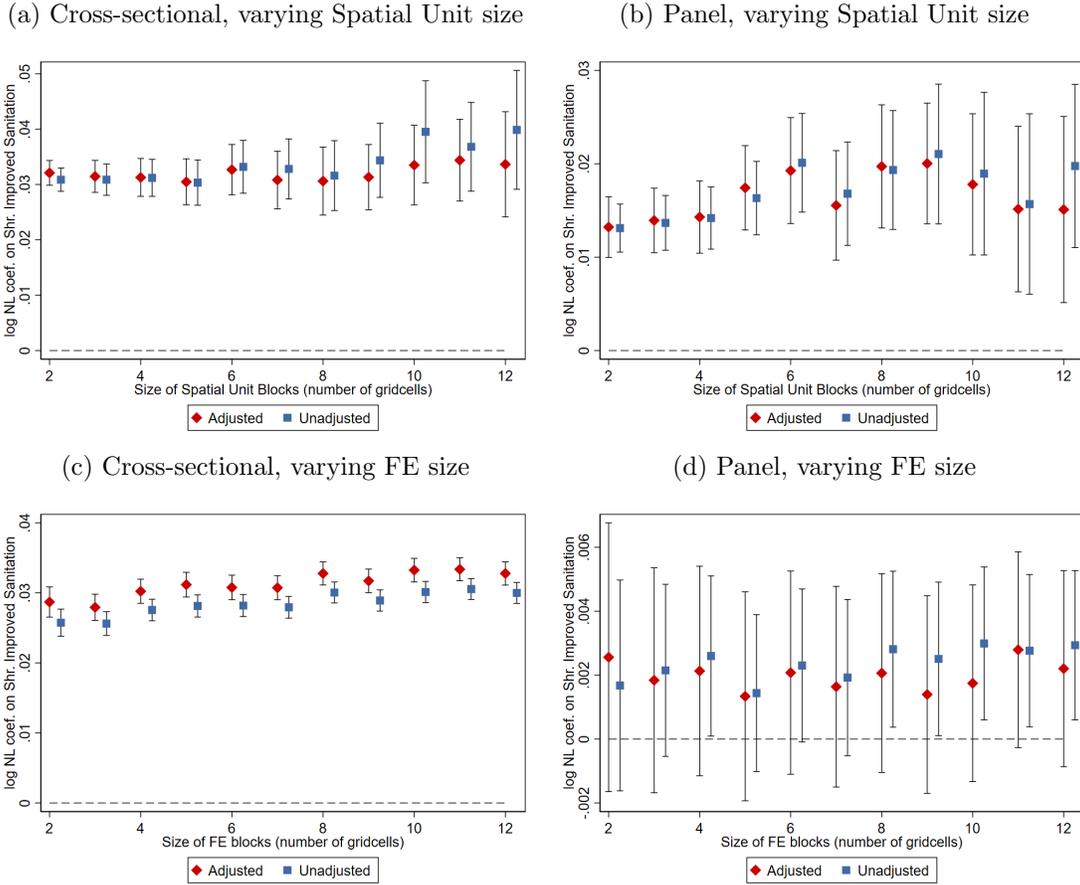
The figure plots coefficients from regressions associating mean years of schooling of individuals aged 15-39 on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity augmenting the specification with block fixed effects of various sizes. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of various sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the gridcell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

Figure B7: Household Access to Electricity-Luminosity Correlation. Further Evidence



The figure plots coefficients from regressions associating household access to electricity on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity augmenting the specification with block fixed effects of various sizes. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of various sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the gridcell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

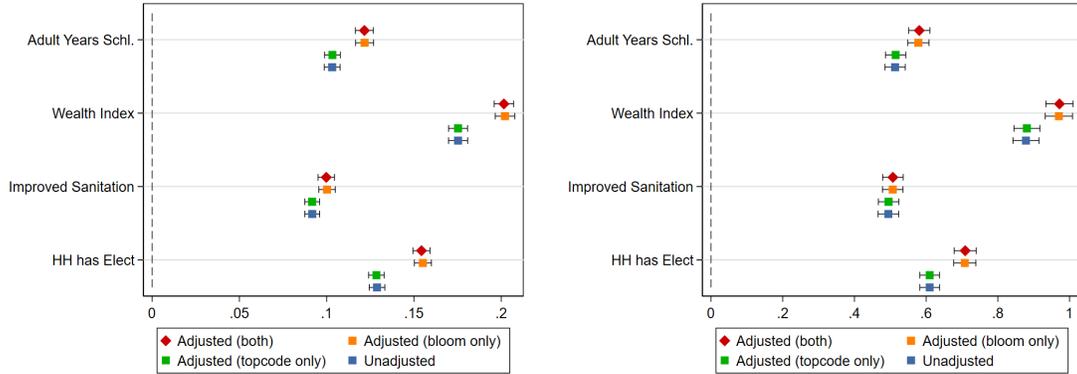
Figure B8: Household Access to Improved Sanitation-Luminosity Correlation. Further Evidence



The figure plots coefficients from regressions associating household access to improved sanitation facilities on Log Luminosity. Panels (a) and (c) give cross-sectional estimates with country-survey year constants, controlling also for the gridcell's log land area. Panels (b) and (d) give panel estimates that, besides the country-year constants, also include gridcell fixed effects. Panels (a)-(b) plot coefficients of log luminosity varying the (gridcell) size unit of the empirical analysis. Panel (c) plots cross-sectional coefficients of log luminosity augmenting the specification with block fixed effects of various sizes. Panel (d) plots panel coefficients of log luminosity augmenting the specification with interactions between country-survey-year constants with block fixed effects of various sizes. Red markers denote estimates using the harmonized VIIRS-DMSP luminosity series, adjusting the DMSP for top coding, sensor calibration, and blooming. Blue markers denote estimates using the unadjusted merged VIIRS-DMSP luminosity series. The bars denote 95% confidence intervals, based on standard errors clustered at the gridcell level for panels (c) and (d) and at the spatial unit level for panels (a) and (b).

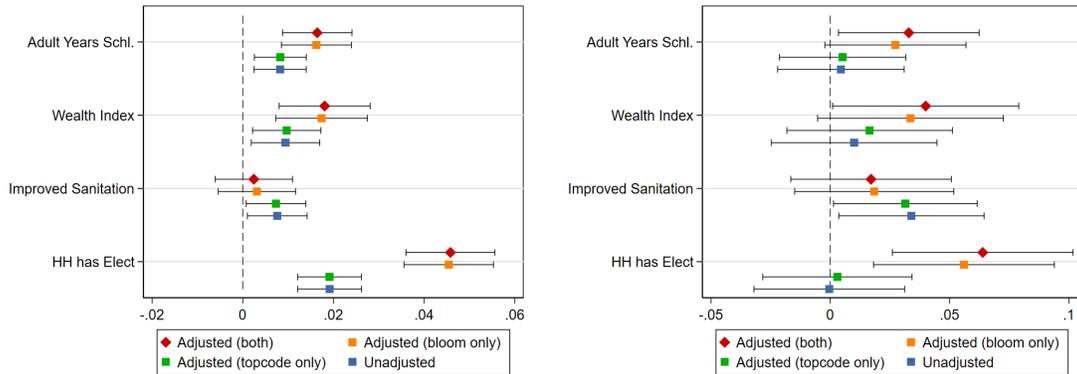
Figure B9: Local (standardized) Development-Luminosity Correlation. DHS comparing all fixes

(a) Cross-sectional Estimates - Log Nightlights (b) Cross-sectional Estimates - Lit Indicator



(c) Panel Estimates - Log Nightlights

(d) Panel Estimates - Lit Indicator



The figure plots coefficients from regressions associating proxies of development from the Demographic and Health Surveys (DHS) on night-time lights (luminosity). We break down the two major adjustments we make (blooming and topcoding corrections). Panels (a) and (c) use log luminosity; panels (b) and (d) use an indicator that equals one when the gridcell is lit and zero otherwise. Panels (a) and (b) control for country-survey-year fixed effects and log cell area. Panels (c) and (d) also include gridcell fixed effects. All outcome variables, mean years of schooling of the population aged 15-39, a composite wealth index, and access to improved sanitation and electricity are standardized to have a mean of zero and a standard deviation of one. The bars give 95% confidence intervals based on standard errors clustered at the gridcell level.

B.3 Country Case Studies. Further Evidence

This Section provides summary statistics, details, and additional evidence of the country case studies reported in Section 5 of the main paper.

B.3.1 Mozambique Census

Summary Statistics Appendix Table B11 gives summary statistics of local development measures and nighttime luminosity across Mozambican localities (admin-4 units). We proxy local development with the mean years of schooling of the population aged 15-39 years, as recorded in the Censuses of 1997, 2007, and 2017; and with the share of employment outside agriculture for 15-24 years old, using information from the 1997 and 2007 Censuses (as the data is missing for the 2017 Censuses). We take the mean values among the young, as in the panel estimates, since we want to capture the “flow” of these variables more accurately. We have information for the full census for 1997, 2007, and 2017. The table gives summary statistics of various luminosity transformations using the newly-compiled harmonized and adjusted VIIRS-DMSP series and with the merged VIIRS-DMSP series without adjusting the later for top-coding and blooming. Panel *A* gives the statistics pooling across all census years. Panels *B*, *C*, and *D* give the statistics for the 1997, the 2007, and the 2017 Census, respectively.

Cross-Sectional Estimates Appendix Tables B12 and B13 report cross-sectional estimates associating mean years of schooling of the population aged 15-39 years and the share of youth employment outside agriculture with log luminosity and a lit indicator, respectively across the census years. Panel *A* gives estimates across admin-2 units (*distritos*). Panel *B* gives estimates across admin-3 units (*postos*). Panel *C* gives estimates across admin-4 units (*localidades*). All specifications condition on log land area and census-year constants. Panel *A* conditions also on census-year specific admin-1 (province) fixed-effects. Panel *B* conditions also on census-year specific admin-2 (district) fixed-effects. Panel *C* conditions also on census-year specific admin-3 (posto) fixed-effects. The estimates suggest that employment outside of agriculture is about 10 percentage points and mean schooling about 0.6 years higher in lit admin-2, admin-3, or admin-4 units than unlit ones.

Panel Estimates Appendix Tables B14 and B15 report panel estimates that explore the dynamic correlation between local development and luminosity across Mozambican administrative units. All specifications include administrative unit fixed effects and census year fixed effects. Panel *A* gives estimates across admin-2 units (districts). Panel *B* gives estimates across admin-3 units (postos). Panel *C* gives estimates across admin-4 units (localities). Panel *A* conditions also on census-year specific admin-1 (province) fixed-effects. Panel *B* conditions also on census-year specific admin-2 (district) fixed-effects. Panel *C* conditions also on census-year specific admin-3 (posto) fixed-effects.

The estimates suggest an increase in mean years of schooling of about 0.25 – 0.37 and an increase of non-agriculture employment of about three percentage points for localities that turn lit than those that stay unlit.

Visual Illustration. Dynamic Correlation Appendix Figure B10 illustrates the within-locality co-movement of luminosity and mean years of schooling of 15-39 year-old Mozambicans over 1997 – 2007 (panels (a)-(b)) and 1997 – 2017 (panels (c)-(d)). The green bars plot the increase in schooling across 1,028 unlit in 1997 localities. Dark green bars in panel (a) reveal an increase in average schooling of 2.3 years in the 89 localities that turned lit by 2007, much higher than in the 939 localities that remained unlit by 2007 (1.76). The difference in schooling between initially unlit locations that either stay unlit or turn lit over the twenty years (2017 – 1997) in panel (c) is 0.5 years (4.26 vs 3.85). Blue bars plot the increase in mean schooling for the 98 localities lit in 1997. Schooling increased by 2.44 years for the 85 that remained lit in 1997, while schooling increased by 2 years in the 13 localities that turned unlit by 2007. Panels (b) and (d) plot changes in schooling years for the four categories of localities [unlit - unlit (light green), unlit - lit (dark green), lit - unlit (light blue) and lit - lit (dark blue)], conditional on admin-3 fixed effects, to control for the considerable differences in local development across Mozambique and compare nearby localities. Differences in schooling correlate with differences in nighttime lights.

Table B11: Descriptive Statistics - Mozambique census and nightlights

	Min	p10	p50	p90	Max	Mean	SD	N
Panel A: 1997, 2007, and 2017								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	26.00	18,981.00	50.85	560.15	3378
Sensor calibration only	0.00	0.00	0.00	160.00	20,307.00	104.36	709.31	3378
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	3.28	9.85	0.09	1.81	3378
Log of sensor calibration only	-0.69	-0.69	-0.69	5.08	9.92	0.75	2.47	3378
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.19	0.39	3378
Sensor calibration only (dummy)	0.00	0.00	0.00	1.00	1.00	0.28	0.45	3378
Gridcell area in km ²	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.70	3378
Log of gridcell area in km ²	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	3378
Years of adult (15-40) schooling	0.04	0.47	2.45	5.53	8.70	2.82	1.95	3377
Share of youth (15-24) emp. out agriculture	0.00	0.02	0.09	0.34	0.99	0.15	0.16	2252
Panel B: 1997								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	0.00	5,378.00	15.09	204.16	1126
Sensor calibration only	0.00	0.00	0.00	24.00	7,898.00	36.55	339.22	1126
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	-0.69	8.59	-0.37	1.17	1126
Log of sensor calibration only	-0.69	-0.69	-0.69	3.20	8.97	-0.02	1.78	1126
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	0.00	1.00	0.09	0.28	1126
Sensor calibration only (dummy)	0.00	0.00	0.00	1.00	1.00	0.14	0.34	1126
Gridcell area in km ²	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.96	1126
Log of gridcell area in km ²	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	1126
Years of adult (15-40) schooling	0.04	0.27	0.68	1.78	4.66	0.89	0.68	1126
Share of youth (15-24) emp. out agriculture	0.00	0.02	0.07	0.29	0.99	0.12	0.14	1126
Panel C: 2007								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	16.00	9,338.00	36.61	417.80	1126
Sensor calibration only	0.00	0.00	0.00	113.00	12,824.00	83.00	617.70	1126
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	2.80	9.14	-0.05	1.65	1126
Log of sensor calibration only	-0.69	-0.69	-0.69	4.73	9.46	0.48	2.31	1126
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.15	0.36	1126
Sensor calibration only (dummy)	0.00	0.00	0.00	1.00	1.00	0.22	0.41	1126
Gridcell area in km ²	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.96	1126
Log of gridcell area in km ²	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	1126
Years of adult (15-40) schooling	0.32	1.53	2.46	4.48	7.25	2.75	1.19	1126
Share of youth (15-24) emp. out agriculture	0.01	0.03	0.11	0.38	0.97	0.17	0.17	1126
Panel D: 2017								
Sensor, blooming, & topcode fixes	0.00	0.00	0.00	81.00	18,981.00	100.86	849.49	1126
Sensor calibration only	0.00	0.00	0.00	326.00	20,307.00	193.53	1,000.32	1126
Log of sensor, blooming, & topcode fixes	-0.69	-0.69	-0.69	4.40	9.85	0.69	2.27	1126
Log of sensor calibration only	-0.69	-0.69	-0.69	5.79	9.92	1.78	2.84	1126
Sensor, blooming, & topcode fixes (dummy)	0.00	0.00	0.00	1.00	1.00	0.32	0.47	1126
Sensor calibration only (dummy)	0.00	0.00	0.00	1.00	1.00	0.49	0.50	1126
Gridcell area in km ²	0.96	86.17	416.78	1,458.47	8,349.15	668.26	854.96	1126
Log of gridcell area in km ²	-0.04	4.46	6.03	7.29	9.03	5.91	1.19	1126
Years of adult (15-40) schooling	0.66	3.14	4.81	6.48	8.70	4.83	1.34	1125

Note: This table presents summary statistics for the Mozambique census and nightlights database. The observations are at the admin 4 level. The total number of localities available is 1126. For the log of nightlights we take ((half of the minimum value of positive NL) + NL) before taking the log.

Table B12: Mozambique Cross-Sectional Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Share	Mean	Mean	Share	Mean
	Adult	Youth	Adult	Adult	Youth	Adult
	(15-39)	(15-24)	(15-39)	(15-39)	(15-24)	(15-39)
	years schl.	Emp. out Ag.	years schl.	years schl.	Emp. out Ag.	years schl.
	(97&07)	(97&07)	(17 only)	(97&07)	(97&07)	(17 only)
Panel A: Admin Level 2						
ln(minNL/2+NL)	0.193*** (0.0252)	0.0252*** (0.00436)	0.0461 (0.0597)	0.209*** (0.0263)	0.0270*** (0.00478)	0.0475 (0.0926)
Obs	282	282	141	282	282	141
Obs(NL=0)	157	157	19	132	132	3
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R ²	0.842	0.721	0.183	0.840	0.717	0.180
Panel B: Admin Level 3						
ln(minNL/2+NL)	0.181*** (0.0181)	0.0325*** (0.00327)	0.212*** (0.0244)	0.132*** (0.0136)	0.0213*** (0.00233)	0.165*** (0.0193)
Obs	774	774	387	774	774	387
Obs(NL=0)	627	627	201	560	560	125
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R ²	0.916	0.775	0.583	0.916	0.759	0.569
Panel C: Admin Level 4						
ln(minNL/2+NL)	0.215*** (0.0265)	0.0360*** (0.00458)	0.170*** (0.0325)	0.138*** (0.0157)	0.0218*** (0.00287)	0.146*** (0.0248)
Obs	2124	2124	1061	2124	2124	1061
Obs(NL=0)	1901	1901	734	1778	1778	548
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R ²	0.894	0.782	0.614	0.894	0.778	0.622

Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, and panel C admin 4. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as the log sum of light in a district, the log area of the district, and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Standard errors in parentheses are clustered at the admin 2 level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B13: Mozambique Cross-Sectional Estimates - Lit Indicator

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Share	Mean	Mean	Share	Mean
	Adult	Youth	Adult	Adult	Youth	Adult
	(15-39)	(15-24)	(15-39)	(15-39)	(15-24)	(15-39)
	years schl.	Emp. out Ag.	years schl.	years schl.	Emp. out Ag.	years schl.
	(97&07)	(97&07)	(17 only)	(97&07)	(97&07)	(17 only)
Panel A: Admin Level 2						
1(NL>0)	0.663*** (0.122)	0.0819*** (0.0213)	0.829*** (0.293)	0.648*** (0.113)	0.0868*** (0.0209)	0.314 (0.356)
Obs	282	282	141	282	282	141
Obs(NL=0)	157	157	19	132	132	3
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R ²	0.787	0.664	0.214	0.786	0.669	0.178
Panel B: Admin Level 3						
1(NL>0)	0.655*** (0.0800)	0.111*** (0.0149)	0.865*** (0.118)	0.576*** (0.0775)	0.0869*** (0.0124)	0.671*** (0.127)
Obs	774	774	387	774	774	387
Obs(NL=0)	627	627	201	560	560	125
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R ²	0.906	0.744	0.549	0.903	0.730	0.516
Panel C: Admin Level 4						
1(NL>0)	0.610*** (0.0799)	0.0954*** (0.0131)	0.657*** (0.128)	0.551*** (0.0724)	0.0840*** (0.0138)	0.627*** (0.119)
Obs	2124	2124	1061	2124	2124	1061
Obs(NL=0)	1901	1901	734	1778	1778	548
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R ²	0.888	0.766	0.608	0.889	0.767	0.610

Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, and panel C admin 4. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as an indicator for positive values of luminosity, the log area of the district, and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Standard errors in parentheses are clustered at the admin 2 level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B14: Mozambique Panel Estimates

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Share	Mean	Mean	Share	Mean
	Adult	Youth	Adult	Adult	Youth	Adult
	(15-39)	(15-24)	(15-39)	(15-39)	(15-24)	(15-39)
	years schl.	Emp. out Ag.	years schl.	years schl.	Emp. out Ag.	years schl.
	(97&07)	(97&07)	(97,07&17)	(97&07)	(97&07)	(97,07&17)
Panel A: Admin Level 2						
ln(minNL/2+NL)	0.0347** (0.0156)	0.00395* (0.00211)	0.115*** (0.0384)	0.0276* (0.0160)	0.00343 (0.00218)	0.145*** (0.0537)
Obs	282	282	423	282	282	423
Obs(NL=0)	157	157	176	132	132	135
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R ²	0.983	0.982	0.869	0.983	0.982	0.869
Panel B: Admin Level 3						
ln(minNL/2+NL)	0.0555*** (0.0178)	0.0101*** (0.00281)	0.0957*** (0.0196)	0.0268* (0.0142)	0.00396* (0.00202)	0.0504*** (0.0167)
Obs	774	774	1161	774	774	1161
Obs(NL=0)	627	627	828	560	560	685
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R ²	0.987	0.971	0.962	0.987	0.970	0.961
Panel C: Admin Level 4						
ln(minNL/2+NL)	0.102*** (0.0243)	0.0148*** (0.00405)	0.0778*** (0.0245)	0.0743*** (0.0136)	0.00707** (0.00275)	0.0571*** (0.0178)
Obs	2124	2124	3185	2124	2124	3185
Obs(NL=0)	1901	1901	2635	1778	1778	2326
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R ²	0.978	0.966	0.963	0.978	0.965	0.963
Panel D: Admin Level 4 - No FEs						
ln(minNL/2+NL)	0.163*** (0.0238)	0.0161*** (0.00371)	0.0753*** (0.0169)	0.125*** (0.0167)	0.0100*** (0.00180)	0.0801*** (0.0119)
Obs	2250	2250	3374	2250	2250	3374
Obs(NL=0)	1976	1976	2744	1848	1848	2420
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R ²	0.942	0.937	0.907	0.944	0.936	0.908

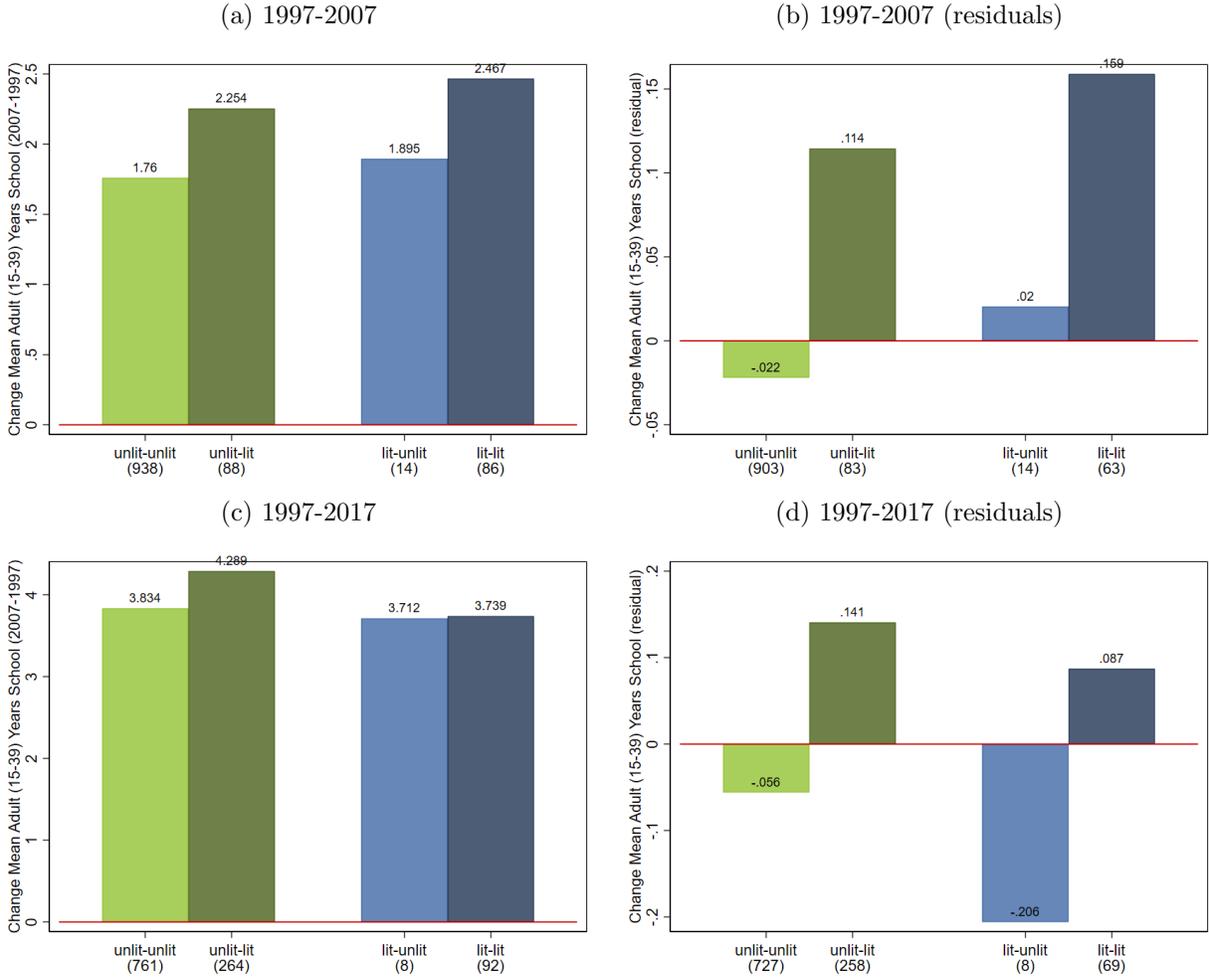
Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, panel C admin 4, panel D also uses admin 4 units but does not add fixed effects for admin 3 by year. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as the log sum of light in a district and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Also included are fixed effects for the admin level denoted in the panel title, and therefore these coefficients reflect changes. Standard errors in parentheses are clustered at the admin 2 level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B15: Mozambique Panel Estimates - Lit Indicator

	Sensor, Blooming, & Topcode Fixes			Sensor Calibration Only		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Share	Mean	Mean	Share	Mean
	Adult	Youth	Adult	Adult	Youth	Adult
	(15-39)	(15-24)	(15-39)	(15-39)	(15-24)	(15-39)
	years schl.	Emp. out Ag.	years schl.	years schl.	Emp. out Ag.	years schl.
	(97&07)	(97&07)	(97,07&17)	(97&07)	(97&07)	(97,07&17)
Panel A: Admin Level 2						
1(NL>0)	0.0787 (0.0620)	0.00661 (0.00821)	0.518*** (0.166)	0.0565 (0.0646)	0.00914 (0.00881)	0.407** (0.187)
Obs	282	282	423	282	282	423
Obs(NL=0)	157	157	176	132	132	135
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
R ²	0.982	0.982	0.870	0.982	0.982	0.868
Panel B: Admin Level 3						
1(NL>0)	0.0805 (0.0786)	0.0177 (0.0108)	0.232** (0.0949)	0.0539 (0.0720)	0.00754 (0.0109)	0.0728 (0.0913)
Obs	774	774	1161	774	774	1161
Obs(NL=0)	627	627	828	560	560	685
FEs	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr	adm2-yr
units	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
R ²	0.987	0.970	0.961	0.987	0.970	0.961
Panel C: Admin Level 4						
1(NL>0)	0.186** (0.0780)	0.0257** (0.0124)	0.228*** (0.0830)	0.271*** (0.0574)	0.0241** (0.0117)	0.226*** (0.0806)
Obs	2124	2124	3185	2124	2124	3185
Obs(NL=0)	1901	1901	2635	1778	1778	2326
FEs	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr	adm3-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R ²	0.978	0.964	0.963	0.978	0.964	0.963
Panel D: Admin Level 4 - No FEs						
1(NL>0)	0.356*** (0.0712)	0.0276** (0.0109)	0.309*** (0.0581)	0.459*** (0.0730)	0.0344*** (0.00813)	0.345*** (0.0565)
Obs	2250	2250	3374	2250	2250	3374
Obs(NL=0)	1976	1976	2744	1848	1848	2420
FEs	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr	cntry-yr
units	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr	adm4-yr
R ²	0.940	0.934	0.907	0.942	0.935	0.908

Note: This table presents regressions of economic indicators from the Mozambique census on nightlights. Each panel is done at a different administrative level: panel A used admin 2 units, panel B admin 3, panel C admin 4, panel D also uses admin 4 units but does not add fixed effects for admin 3 by year. Columns 1-3 use nightlights that have been adjusted for cross-sensor calibration including the downgrading of VIIRS. Columns 4-6 use nightlights that have also been adjusted to fix blooming and topcoding. All specifications include nightlights as an indicator for positive values of luminosity and fixed effects for year interacted with the admin unit one level above (e.g. in panel C, units are admin 4 and so we include admin 3 by year fixed effects). Also included are fixed effects for the admin level denoted in the panel title, and therefore these coefficients reflect changes. Standard errors in parentheses are clustered at the admin 2 level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure B10: Mozambique Δ Mean Years Schooling by Changes in Lit/Unlit



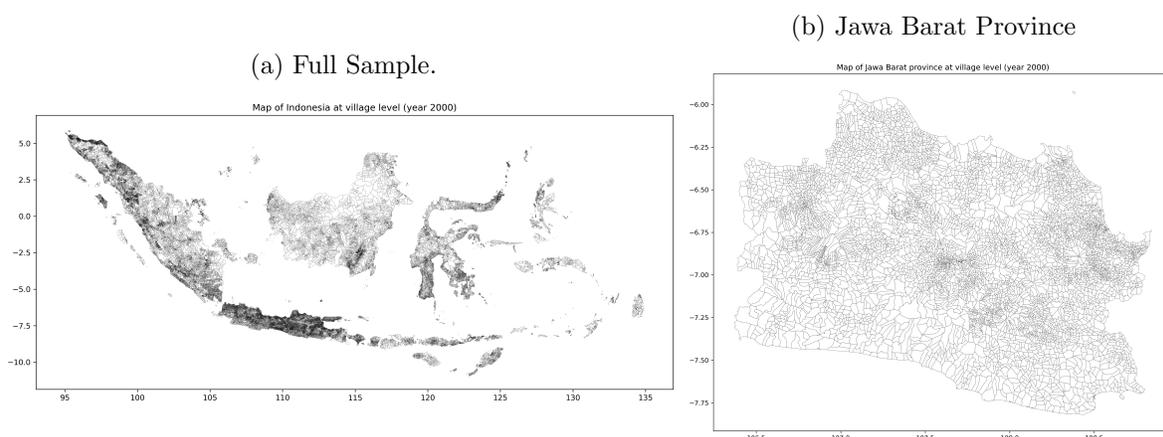
The figure plots the change in average years of schooling for 15-39 year-olds by changes in the extensive margin of luminosity across Mozambican localities (admin-4 units). Green bars plot the mean years of schooling change for initially unlit localities. Dark green bars plot the change in schooling for localities that are turning lit, while light green bars plot the change in schooling for localities remaining unlit. Blue bars plot the mean years of schooling change for initially lit localities. Dark blue bars plot the change in schooling for localities that remain lit, while light blue bars plot the change for localities that turn from lit to unlit. Panels (a) and (c) plot unconditional changes in mean schooling years over 1997 and 2007 and over 1997 and 2017, respectively; panels (b) and (d) plot changes in mean years of schooling over 1997-2007 and 1997-2017, conditional on admin-3 fixed effects.

B.3.2 Indonesia

Below, we report summary statistics and additional evidence of the analysis that explores the cross-sectional and panel association between luminosity and local development across more than 60,000 Indonesian villages using the PODES data. Appendix Figure B11 gives a visual illustration.

Summary Statistics Appendix Table B16 presents summary statistics of various public goods provision measures, as well as the first principal components, which aggregate the variation, and nighttime lights across Indonesian villages, encompassing all PODES surveys from the mid-1990s to 2018. It also reports the summary statistics for the same variables at the admin-3 and admin-2 levels.

Figure B11: Spatial Distribution of Indonesian Villages (DESA) Level.



Notes: The Figure reports the spatial distribution of villages (DESA) across Indonesia (Panel (a)). Panel (b) zooms into the province of Jawa Barat, West of Java and south of Jakarta.

Further Evidence. Localized Variation Appendix Figure B12 reproduces the analysis in Figure 7 with the first principal component (PC1) of the numerous public goods measures from the PODES. In our sample, we have 3,737 admin-3 units and the average number of village per admin-3 unit is 25.28; we also count 311 (28) admin-2 (admin-1) units, containing on average 329.58 (4540.56) villages. The panel labeled PC1 – level 3 reports coefficients for the adjusted and unadjusted night-lights series estimated with village fixed effects and admin-3-by-year fixed effects. The specifications labeled PC1-admin-level-2 include village constants and admin-2-by-year fixed effects, while the models labeled PC - level 1 include village fixed effects and admin-1-by-year fixed effects. The adjusted for top-coding, blooming, and sensor quality DMSP series fused to a downward VIIRS post 2013 are significantly more important correlates of the local public goods proxy (the principal component). As in the analysis with DHS data across dozens of African countries and the Mozambique-based analysis, the adjusted series perform way better in panel estimation, which

often magnifies error-in-variables.

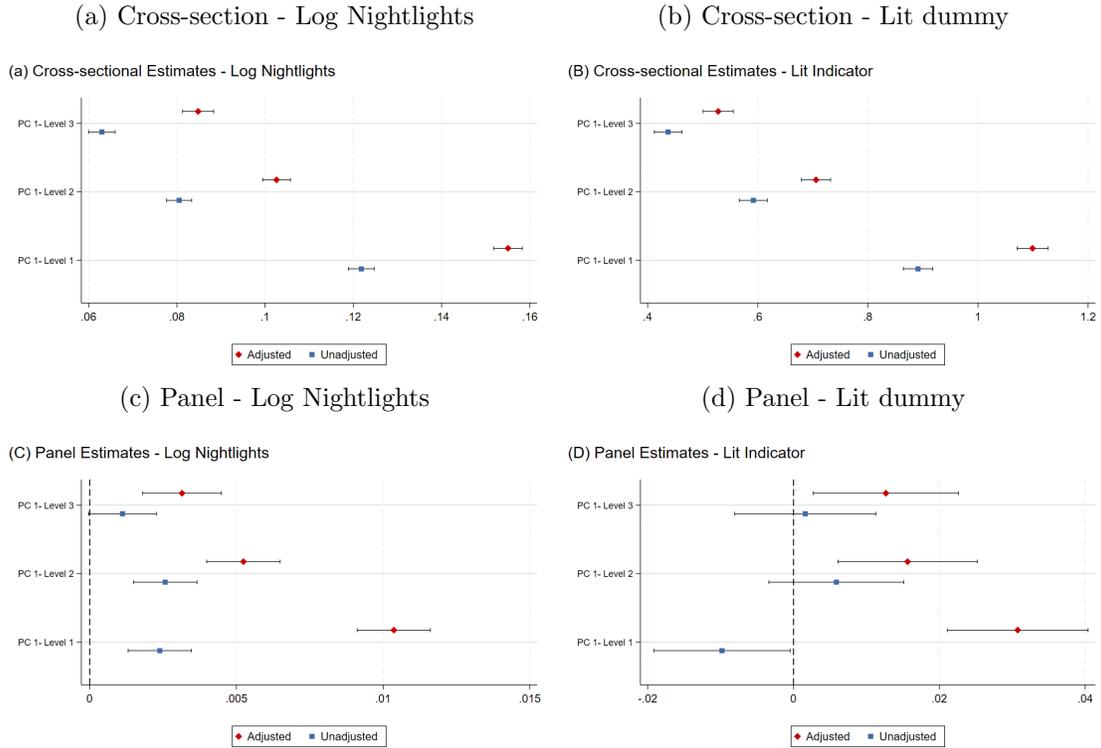
Further Evidence. Spatial Aggregation Appendix Figures B13 and B14 report cross-sectional (panels (a)-(b)) and panel (panels (c)-(d)) estimates associating the various public goods measures to the log luminosity and an indicator that takes the value of one for lit villages across the 3,737 admin-3 and the 311 admin-2 units, respectively. Panels (a) and (b) control for log area of the corresponding unit of observation and the survey-year fixed effects. Panels (c) and (d) control for the admin-level fixed-effects are interacted with survey-year fixed-effects. Two results stand out. First, a significantly positive correlation between the adjusted and harmonized luminosity series and most public goods measures emerges with both transformations of luminosity, both when exploring within-admin-2 and within-admin-3 variation. Second, in line with the DHS evidence in Section 4.4.1, the adjusted and harmonized luminosity series correlate more strongly with local development at more granular units of observation compared to the unadjusted DMSP nighttime data.

Table B16: Summary Statistics by Administrative Level

Panel A: Village (Desa) Level					
	Obs	Mean	SD	Min	Max
Log of Sensor, blooming, & topcode fixes	492808	-4.28	5.48	-10.13	5.88
Log of Sensor calibration only	492808	-2.09	5.52	-10.09	4.14
Lit (dummy) Sensor, blooming, & topcode fixes	492808	0.55	0.50	0.00	1.00
Lit (dummy) Sensor calibration only	492808	0.69	0.46	0.00	1.00
Garbage Disposal	479041	0.00	1.00	-0.37	2.72
Use Toilet	479041	0.00	1.00	-1.50	0.67
Access drinking water	479041	-0.00	1.00	-0.42	2.40
Gas/electricity for cooking	479041	-0.00	1.00	-0.56	1.81
Any paved road	479040	-0.00	1.00	-2.85	0.36
Any doctor	479041	0.00	1.00	-0.49	2.09
Any modern public health facility	479041	-0.00	1.00	-0.38	2.73
Number of kindergartens	492808	0.00	1.00	-0.60	47.59
Number of primary schools	492808	-0.00	1.00	-1.03	37.91
Number of middle schools	492808	0.00	1.00	-0.59	22.46
Number of high schools	492808	0.00	1.00	-0.35	59.44
Panel B: Admin-3 (Kecamatan) Level.					
	Obs	Mean	SD	Min	Max
Log of Sensor, blooming, & topcode fixes	29896	-2.27	5.15	-11.65	5.69
Log of Sensor calibration only	29896	-0.01	4.06	-10.87	4.14
Lit (dummy) Sensor, blooming, & topcode fixes	29896	0.81	0.39	0.00	1.00
Lit (dummy) Sensor calibration only	29896	0.91	0.29	0.00	1.00
Garbage Disposal	29677	0.00	1.00	-0.56	2.80
Use Toilet	29677	0.00	1.00	-2.11	0.82
Access drinking water	29677	-0.00	1.00	-0.64	2.91
Gas/electricity for cooking	29677	0.00	1.00	-0.64	1.89
Any paved road	29677	0.00	1.00	-4.27	0.47
Any doctor	29677	0.00	1.00	-0.95	2.87
Any modern public health facility	29677	0.00	1.00	-1.07	5.73
Number of kindergartens	29896	0.00	1.00	-0.86	18.62
Number of primary schools	29896	-0.00	1.00	-1.41	10.94
Number of middle schools	29896	0.00	1.00	-0.98	14.51
Number of high schools	29896	-0.00	1.00	-0.64	15.19
Panel C: Admin-2 (Kabupaten) Level.					
	Obs	Mean	SD	Min	Max
Log of Sensor, blooming, & topcode fixes	2488	-0.22	2.66	-9.73	5.45
Log of Sensor calibration only	2488	1.08	2.07	-9.05	4.14
Lit (dummy) Sensor, blooming, & topcode fixes	2488	0.99	0.08	0.00	1.00
Lit (dummy) Sensor calibration only	2488	1.00	0.07	0.00	1.00
Garbage Disposal	2478	0.00	1.00	-0.75	2.60
Use Toilet	2478	-0.00	1.00	-2.87	1.02
Access drinking water	2478	-0.00	1.00	-0.87	3.01
Gas/electricity for cooking	2478	-0.00	1.00	-0.66	2.00
Any paved road	2478	0.00	1.00	-5.26	0.67
Any doctor	2478	0.00	1.00	-1.13	2.83
Any modern public health facility	2478	-0.00	1.00	-1.18	5.58
Number of kindergartens	2488	-0.00	1.00	-0.96	9.28
Number of primary schools	2488	-0.00	1.00	-1.49	6.42
Number of middle schools	2488	-0.00	1.00	-1.07	7.78
Number of high schools	2488	0.00	1.00	-0.79	7.85

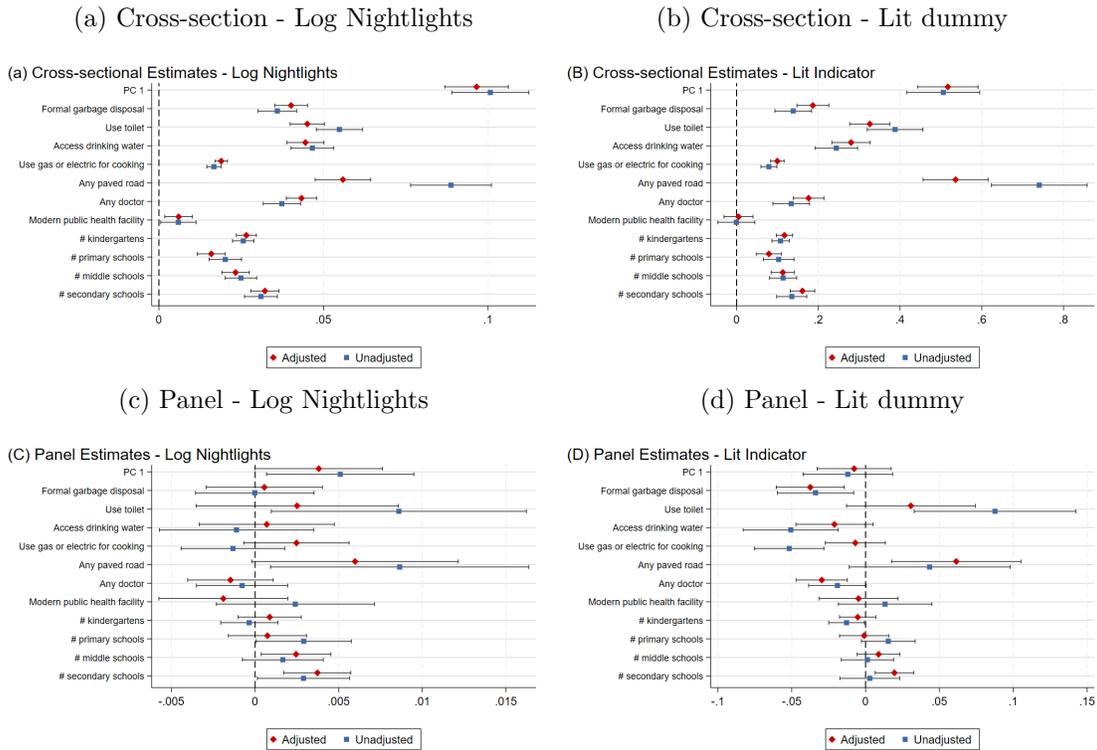
Notes: This table presents summary statistics for the nightlights and PODES database. Panel A reports the statistics at the Village (DESA) level; Panel B provides information at the Admin-3 (Kecamatan) level; Panel C presents the descriptive statistics at the Admin-2 (Kabupaten) level. All PODES variables are standardized

Figure B12: Local (standardized) Development - Luminosity Association. Indonesia PODES dataset. Alternative Admin \times Year FEs.



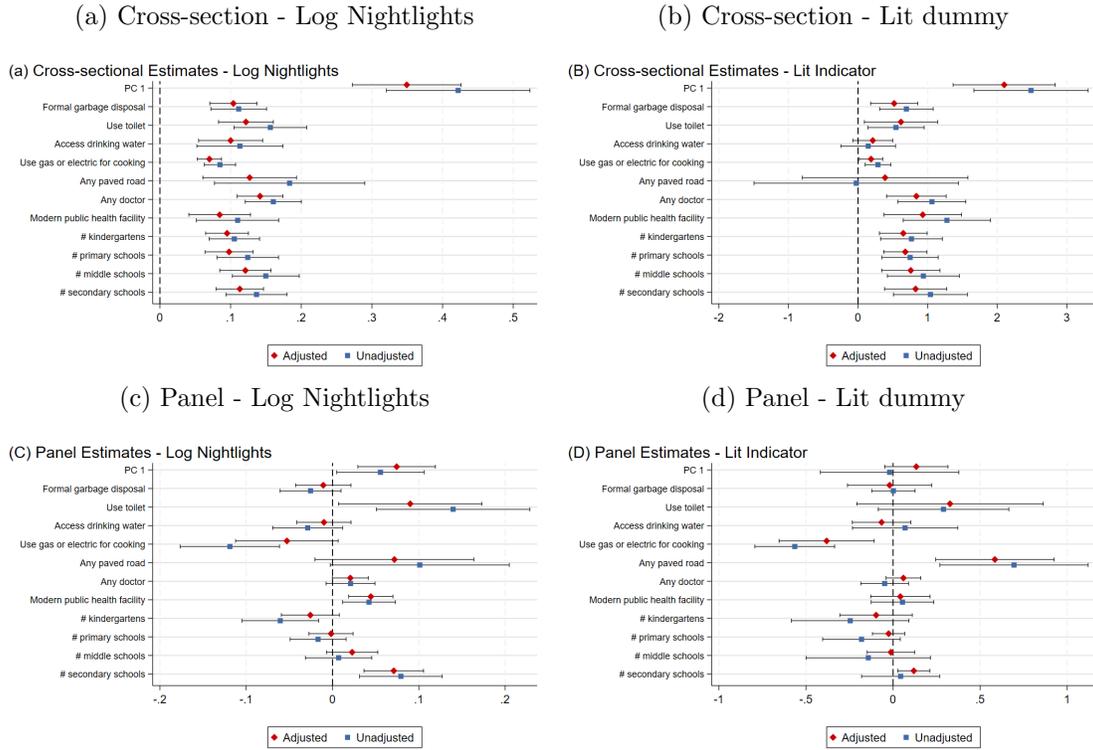
Notes: This figure plots coefficients from regressions of the First Principal Component of the different PODES measures on nightlights at the village (DESA) level. For the luminosity variables, panels (a) and (c) use log nightlights and panels (b) and (d) use an indicator equal to one for positive lights and zero otherwise. Panels (a) and (b) control for log village area, while panels (b) and (d) control for village. On top of these controls, we also control for admin-3 \times period FEs (PC 1 - level 3), admin-2 \times period FEs (PC 1 - level 2), and admin-1 \times period FEs (PC 1 - level 1). The red diamonds denote estimates using our corrected nightlight series, and blue squares denote estimates using the unadjusted series. For these figures, all PODES outcomes are standardized to have a mean of zero and a standard deviation of one. The bars represent 95% confidence intervals, and standard errors are clustered at the village level.

Figure B13: Local (standardized) Development - Luminosity Association. Indonesia PODES dataset. Admin-3 level.



Notes: This figure plots coefficients from regressions of PODES measures on nightlights at the admin-3 level. For the luminosity variables, panels (a) and (c) use log nightlights, and panels (b) and (d) use an indicator equal to one for positive lights and zero otherwise. Panels (a) and (b) control for log village area and admin-2 \times survey-year fixed effects, while panels (c) and (d) control for village and admin-3 \times survey-year fixed effects. The red diamonds denote estimates using our corrected nightlight series, and blue squares denote estimates using the unadjusted series. For these figures, all PODES outcomes are standardized to have a mean of zero and a standard deviation of one. The bars represent 95% confidence intervals, and standard errors are clustered at the village level.

Figure B14: Local (standardized) Development - Luminosity Association. Indonesia PODES dataset. Admin-2 level.



Notes: This figure plots coefficients from regressions of PODES measures on nightlights at the admin-2 level. For the luminosity variables, panels (a) and (c) use log nightlights and panels (b) and (d) use an indicator equal to one for positive lights and zero otherwise. Panels (a) and (b) control for log village area and admin-1 \times survey-year fixed effects, while panels (c) and (d) control for village and admin-1 \times survey-year fixed effects. The red diamonds denote estimates using our corrected nightlight series, and blue squares denote estimates using the unadjusted series. For these figures, all PODES outcomes are standardized to have a mean of zero and a standard deviation of one. The bars represent 95% confidence intervals, and standard errors are clustered at the village level.

B.3.3 India

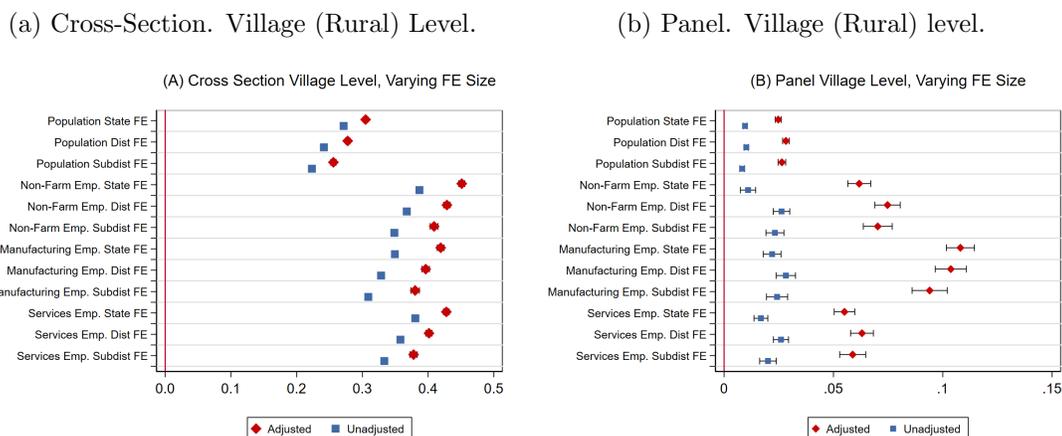
Below, we report summary statistics and additional evidence from the analysis that explores the correlation between luminosity and local development across approximately 550,000 rural and more than 7,000 urban Indian municipalities, using data from the SHRUG portal and following the analysis in Asher *et al.* (2021).

Summary Statistics Appendix Table B17 provides descriptive statistics for all variables used in the Indian analysis. At the village level (Panel A), it lists the log population (from the Population Censuses of 1991, 2001, 2011); the log total non-farm, manufacturing, and services employment (from the Economic Censuses of 1990, 1998, 2005, 2013); and the logarithm of nighttime lights. Panels B and C also report the same statistics after aggregating villages to the sub-district (admin-

3) and district (admin-2) levels, respectively.

Further Evidence. Localized Variation Appendix Figure B15 reproduces the analysis in Figure 8, exploiting different batteries of admin \times year FEs. On top of the more than 550,000 villages and towns, the dataset counts 5,867 subdistricts (admin-3) units and the average number of village per admin-3 unit is 93.58; we also count 623 (33) admin-2 (admin-1) units, containing on average 881.32 (16638.42) villages. The panels labeled “State FE” report coefficients for the adjusted and unadjusted night-lights series estimated with village fixed effects and admin-1-by-year fixed effects. The specifications labeled “Dist FE” include village constants and admin-2-by-year fixed effects, while the models labeled “Subdist FE” include village fixed effects and admin-3-by-year fixed effects [our baseline specification in Figure 8]. Our adjusted series, which merges the corrected for top-coding, blooming, and sensor quality DMSP series with a downgraded VIIRS post 2013, correlates very strongly with the different proxies of development across the different permutations. As in the analysis with DHS data across dozens of African countries, the Mozambique-based analysis, and the Indonesia-based validation, the adjusted series performs way better in panel estimation, which often magnifies error-in-variables.

Figure B15: Local Development - Luminosity Association. India (SHRUG). Village (Rural) Level. Alternative Admin \times Year FEs.



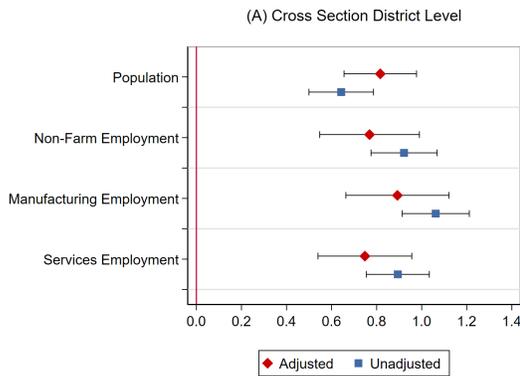
Notes: This Figure plots coefficients from regressions of SHRUG measures on nightlights at the village (rural) level. For the luminosity variables, all panels use log nightlights. Panel (a) controls for log town (village) area; while panel (b) controls for village fixed effects. Specifications labeled “State FE” control for admin-1 \times year FEs. Specifications labeled “Dist FE” control for admin-2 \times year FEs. Specifications labeled “Subdist FE” control for admin-3 \times year FEs. The red diamonds denote estimates using our corrected nightlight series, and blue squares denote estimates using the unadjusted series. The bars represent 95% confidence intervals, and standard errors are clustered at the village level.

Further Evidence. Spatial Aggregation. Appendix Figure B16 compares cross-sectional (panels a–b) and panel (panels c–d) regressions that link log night-time lights to development

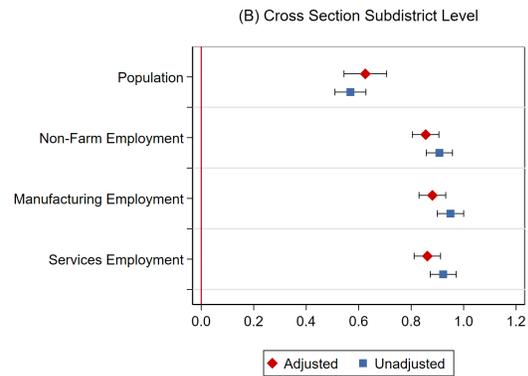
measures from the Population and Economic Censuses. The left column covers 623 districts (admin-2); the right column covers 5,867 sub-districts (admin-3). Cross-sectional models (a, b) control for the log of land area, while panel models (c, d) absorb unit fixed effects. To account for local common shocks, district regressions (a, c) include state-year dummies (admin-1 \times year), and sub-district regressions (b, d) include district-year dummies (admin-2 \times year). The findings mirror those documented for the DHS waves in Africa and Indonesia. Across specifications, coefficients derived from the newly adjusted luminosity series (red diamonds) are consistently larger than those based on the unadjusted data (blue squares), especially for panel estimates. This divergence is most pronounced at finer spatial scales: estimates at the village and town level (Figure 8) are both larger in magnitude and more precisely measured than those at the sub-district and district level (Appendix Figure B16).

Figure B16: Local Development - Luminosity Association. India (SHRUG). District and Subdistrict level.

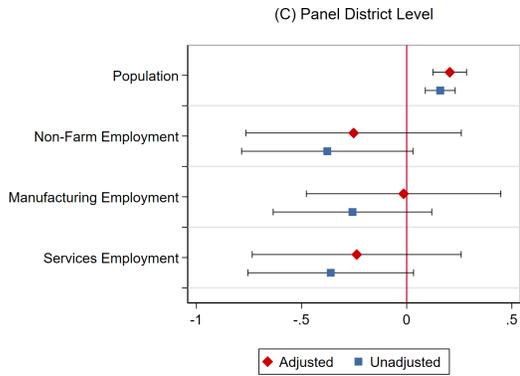
(a) Cross-Section. District level



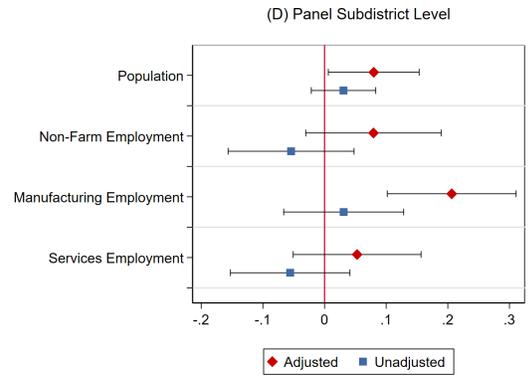
(b) Cross-Section. Subdistrict level



(c) Panel. District Level.



(d) Panel. Subdistrict Level.



Notes: This figure plots coefficients from regressions of SHRUG measures on nightlights at the district and subdistrict level. For the luminosity variables, all panels use log nightlights. Panels (a) controls for log district area and admin-1 (state) \times period fixed effects; panel (b) controls for log district area and admin-2 (district) \times period fixed effects; while panels (b) and (d) control for district and subdistrict FEs as well as admin-1 (2) \times period fixed effects. The red diamonds denote estimates using our corrected nightlight series, and blue squares denote estimates using the unadjusted series. The bars represent 95% confidence intervals, and standard errors are clustered at the town (village level).

Table B17: India. Summary Statistics

Panel A: District	obs	mean	sd	min	max
Population Census Variables					
Log of Sensor, blooming, & topcode fixes	1769	0.274	0.360	0.000	2.708
Log of Sensor calibration only	1769	0.578	0.497	0.000	2.892
Log Population	1769	13.845	1.127	8.931	16.636
Economics Census Variables					
Log of Sensor, blooming, & topcode fixes	2524	0.804	0.812	0.000	5.325
Log of Sensor calibration only	2524	1.393	0.852	0.000	4.159
Log Non-Farm Employment	2524	10.559	2.341	0.000	16.062
Log Manufacturing Employment	2524	9.116	2.298	0.000	15.237
Log Services Employment	2524	10.233	2.271	0.000	15.485
Panel B: Subdistrict					
Population Census Variables					
Log of Sensor, blooming, & topcode fixes	16343	0.231	0.392	0.000	3.549
Log of Sensor calibration only	16343	0.523	0.537	0.000	3.865
Log Population	16343	11.473	1.142	1.386	16.636
Economics Census Variables					
Log of Sensor, blooming, & topcode fixes	23608	0.663	0.802	0.000	5.365
Log of Sensor calibration only	23608	1.323	0.913	0.000	4.159
Log Non-Farm Employment	23608	8.033	2.304	0.000	16.062
Log Manufacturing Employment	23608	6.587	2.227	0.000	15.237
Log Services Employment	23608	7.680	2.231	0.000	15.485
Panel C: Town					
Population Census Variables					
Log of Sensor, blooming, & topcode fixes	12385	2.040	1.084	0.000	5.346
Log of Sensor calibration only	12385	2.640	0.814	0.000	4.159
Log Population	12385	10.007	1.057	4.710	15.962
Economics Census Variables					
Log of Sensor, blooming, & topcode fixes	22928	0.943	0.730	0.000	4.670
Log of Sensor calibration only	22928	1.739	0.679	0.000	3.778
Log Non-Farm Employment	22928	7.346	1.508	0.000	14.227
Log Manufacturing Employment	22928	5.738	1.691	0.000	12.847
Log Services Employment	22928	6.981	1.561	0.000	14.010
Panel D: Village					
Population Census Variables					
Log of Sensor, blooming, & topcode fixes	1517197	0.411	0.707	0.000	5.130
Log of Sensor calibration only	1517197	1.021	0.981	0.000	4.159
Log Population	1517197	6.664	1.016	0.693	12.545
Economics Census Variables					
Log of Sensor, blooming, & topcode fixes	1779715	0.358	0.689	0.000	5.059
Log of Sensor calibration only	1779715	0.960	0.973	0.000	4.159
Log Non-Farm Employment	1779715	3.081	1.484	0.000	12.206
Log Manufacturing Employment	1779715	1.596	1.497	0.000	11.708
Log Services Employment	1779715	2.737	1.429	0.000	12.206

Notes: This table presents summary statistics for the nightlights and SHRUG database. Panel A reports the statistics at the District level; Panel B provides information at the Subdistrict level; Panel C and Panel D present the descriptive statistics at the Town (urban) and Village (rural) level, respectively. Population and employment variables are expressed in natural logarithms. Light variables represent night-time luminosity, either Sensor, blooming, & topcode fixes (adjusted) or Sensor calibration only (unadjusted).