UNIVERSITY OF WAIKATO

Hamilton
New Zealand

Luminosity and Local Economic Growth

Working Paper in Economics 8/24

October 2024

John Gibson
School of Accounting, Finance and Economics
University of Waikato
Private Bag 3105
Hamilton, 3240
New Zealand

Email: john.gibson@waikato.ac.nz

Bonggeun Kim
Seoul National University

Chao Li
University of Auckland

Abstract

We examine relationships between luminosity and local economic growth for counties in China and the US and districts in Indonesia. Many authors estimate treatment effects on local luminosity growth and transfer GDP-luminosity elasticities from elsewhere to calculate economic growth effects. Our insight is that these GDP-luminosity elasticities vary especially by spatial scale and metro status, and also by period and remote sensing source. The elasticities mainly capture extensive margins of luminosity. Measurement errors in popular DMSP data attenuate GDP-luminosity elasticities but aggregation-sensitivity persists even when using instrumental variables estimation. Consequently, claimed growth effects of various treatments may be quite inaccurate.

Keywords

GDP growth

Luminosity

measurement error

treatment effects

JEL Classification

C21

O40

R11

Acknowledgements

Initial work on this paper occurred while Gibson visited the Institute of Economic Research, Hitotsubashi University; he thanks them for their hospitality. Helpful comments from audiences at the University of Tokyo, the Asian Economic Development Conference, the EAEA and WEAI conferences, and the RSAI World Congress are gratefully acknowledged. We thank Geua Boe-Gibson and Xiaoxuan Zhang for assistance with the remote sensing data. Financial support is from the Marsden Fund project UOW1901.

I. Introduction

The seminal study by Henderson et al. (2012) introducing economists to night-time lights (NTL) data greatly expanded the range of what can be feasibly evaluated in data-poor environments. Whenever and wherever traditional economic activity data, such as GDP, are either absent or not trusted because of concerns about manipulation and error, NTL data might be used instead as a local economic growth proxy when estimating treatment impacts. For example, natural disasters (Heger and Neumayer, 2019; Kocornik-Mina et al., 2020; Joseph, 2022), public health crises (Roberts, 2021; Beyer et al., 2023) and monetary shocks (Chodorow-Reich et al., 2020; Chandra and Cook, 2022) have all recently been evaluated by examining associated changes in night-time lights (luminosity, for short).

To provide a more familiar metric, some authors estimate effects of these treatments on local luminosity growth and then transfer GDP-luminosity elasticities from elsewhere to show economic growth effects. For example, Chodorow-Reich et al. (2020) created quarterly lights variables for Indian districts, using monthly VIIRS data from April 2012 onwards, and estimated a 12 percentage point fall in luminosity with the 2016 demonetization shock.[1] They then apply a GDP-lights elasticity of 0.3, estimated by Henderson et al. (2012) with annual DMSP data on 188 countries over 1992-2008, to calculate an output decline of 3.6 percentage points $(= 0.3 \times 0.12)$ from India's demonetization.[2] Several questions are raised by this transferring of elasticities from one context to another: do the elasticities vary with the level of spatial aggregation? With the time period? With the remote sensing system providing the NTL data? In other words, is it meaningful to combine a GDP-luminosity elasticity that was estimated for annual country-level DMSP data with a treatment effect that is estimated at the district level using sub-annual VIIRS data for a much later period?

[1] VIIRS is the Visible Infrared Imaging Radiometer Suite of remote sensing instruments on the Suomi-NPP and NOAA-20 satellites providing data since 2012.

[2] DMSP is the Defense Meteorological Satellite Program, which is a series of satellites first launched in the 1960s by the US Department of Defense.

The next section discusses selected studies using this elasticity transfer approach, which relies on decomposing impacts that various treatments have on economic activity into two parts: $(\partial GDP/\partial luminosity) \times (\partial luminosity/\partial treatment)$. The range of treatments studied has no restriction; some use very finely-grained spatial data, like individual beaches in Barbados of average area 0.02 km^2 (Corral and Schling, 2017). In contrast, studies with the $\partial GDP/\partial luminosity$ values are necessarily more spatially aggregated because they need GDP data. So the elasticities are generally transferred from spatially aggregated sources to less aggregated contexts. Also, the GDP-luminosity elasticities are often for earlier periods; after Henderson et al. (2012) there is less novelty in reporting GDP-luminosity elasticities so not many updates were published.[3] This temporal effect makes GDP-luminosity elasticities more likely to be based on DMSP data (the only option prior to 2012) while the treatment effects on luminosity may use data from different remote sensing systems because both DMSP and VIIRS are available since 2012.

Our insight in this paper is that these GDP-luminosity elasticities are sensitive to the level of spatial aggregation and to whether they are for metro areas only rather than for samples also covering non-metro areas. In our two developing country examples (China and Indonesia) spatial aggregation inflates the GDP-luminosity elasticities, which also are higher in metro areas than in non-metro areas. In the United States, where luminosity is trending downwards even as GDP continues to grow, these patterns are reversed. The variation that we describe undermines the elasticity transfer approach to calculating economic growth effects of various treatments. We also show that measurement errors in popular DMSP data attenuate GDP-luminosity elasticities but the sensitivity to spatial aggregation persists if instrumental variables

[3] An early follower of Henderson et al. (2012) [hereafter, HSW (2012)], whose reported GDP-luminosity elasticities are used by others, is Hodler and Raschky (2014), who used 18 years of DMSP data for 1500 regions (mostly at the first sub-national level) to estimate a GDP-luminosity elasticity of 0.4, compared to 0.3 estimated by HSW with national level data. Some studies eschew published elasticity estimates, instead making an (internal) transfer by estimating their own GDP-luminosity elasticity from somewhere with the necessary GDP data and then applying that estimate to calculate growth effects in their treatment setting (Kim et al., 2023).

estimation is used to mitigate the effects of measurement error.

Our findings differ from influential prior studies that suggested linearity, with similar GDP-luminosity elasticities at country, region, and district level (Hodler and Raschky, 2014; Storeygard, 2016). Even in newer studies that find heterogeneous elasticities, with luminosity less responsive to GDP changes at higher baseline GDP, the relationships were seen as stable across alternative aggregation levels derived from artificial partitioning (Bluhm and McCord, 2022).[4] Hence, the following view, in a recent applied study, is fairly representative of how researchers likely consider the GDP-luminosity elasticities:

> "…although the relationship between lights and economic activity estimated in HSW (2012) could be different at the local level ..[a]... number of papers have estimated the lights-GDP relationship using subnational data, finding a similar elasticity of GDP to lights as HSW (2012)" (Kocornik-Mina et al, 2020, p.55).

Nevertheless, there have been some cautions: Asher et al (2021) highlight that the assumption of the elasticities being the same at very high and very low levels of geographic aggregation is largely untested, despite use of luminosity as a proxy for *local* economic growth.

Our results differ from the prior view because we go further down the spatial ladder, examining relationships between GDP and luminosity at county level to get closer to the granularity of many luminosity-treatment estimates. If elasticities are scale dependent, as our results suggest, the fact that prior studies were mostly at national and first sub-national level would tend to find elasticities more alike. We also use three different NTL data sources, of varying accuracy, to examine measurement error effects. Finally, we pay more attention to mechanisms linking changes in economic activity and luminosity, showing that the link is mostly on the extensive margin (changes in illuminated area), creating non-linearities that may get distorted when data are spatially aggregated.

[4] Nevertheless, Bluhm and McCord (2022) have a similar take-home message to ours; caution is needed when applying an elasticity from the literature to specific empirical contexts for wanting to indicate how a change in luminosity in a particular place might translate into GDP growth terms.

II. A Selective Review of Elasticity Transfer Studies

Appendix A provides details on 20 studies where GDP-luminosity elasticities are transferred from an external setting to combine with treatment effects on luminosity. We especially include papers in top journals like *AER*, *QJE*, *ReStud, RESTAT*, and *AEJ:Applied*, and in top field journals like *JDE*, *JUE*, *JIE*, and *JEEM*. Scores of other studies have similar features, but we want to show the broad acceptability of the elasticity transfer approach in economics, as seen in the decisions of reviewers and editors at top journals.

We use a 6-level geographic code, to compare spatial aggregation in the studies where GDP-luminosity elasticities originate with aggregation at the destination (where the treatment effect is actually estimated): national, region/province/state, district, city/county, village, and pixel/micro-grid (<5km). The elasticities originate from studies that, on average, use data at three levels more aggregated than the situation the elasticities are being 'imported' into. In some cases, origin and destination studies are at opposite ends of the spectrum, as when a GDP-luminosity elasticity estimated from national level data is applied to a treatment effect estimated with pixel-level data (Kocornik-Mina et al, 2020; Miranda et al, 2020). An inherent assumption of this procedure is that elasticities are invariant to spatial aggregation.

A persistent finding is that cross-sectional (or 'between' in panels) GDP-luminosity elasticities greatly exceed time-series ('within') ones.[5] Across three studies at county and state level in the US and China, cross-sectional elasticities are 10-28 times larger, at close to 1.0, yet time-series elasticities are usually below 0.1 (Chen and Nordhaus, 2019; Gibson and Boe-Gibson, 2021; Zhang and Gibson, 2022). This gap also shows up in predictive power, with R^2 values 50-times as large in cross-sections, on average, than for time-series changes.[6] Hence,

[5] Asher et al (2021) show elasticities of other indicators (e.g. non-farm employment) with respect to luminosity also have this pattern; for Indian villages the cross-sectional elasticity is 0.6 while the time-series elasticity is 38-times smaller, at 0.016.

[6] HSW (2012) inadvertently may have exaggerated predictive power of luminosity. They used *Stata* `xtreg` with year dummies, which counts the predictive power of the year dummies as part of the within-R^2. The `reghdfe` command lacks this flaw, and when used on their replication data the within-R^2 falls from 0.77 to just 0.21.

changes in luminosity only weakly predict changes in local economic activity, not just for GDP but for other indicators like employment and household expenditures, even when the cross-sectional relationships are far stronger (Goldblatt et al, 2020). Some elasticity transfer studies (e.g. Akter, 2023) apply the far higher cross-sectional elasticities to treatment effects estimated with difference-in-differences from time-series variation in luminosity. An upward bias in the calculated economic growth effects is therefore likely.

The other feature of the studies in Appendix A is the time gap: the GDP-luminosity elasticities are, on average, from samples with midpoints almost five years before midpoints of samples used to estimate treatment effects. Some gaps are 20 years (Gong et al, 2024). If elasticities evolve, as economic activity shifts from agriculture to services, as lighting types change (e.g. LED replacing halogen), and as satellite sensors improve, old elasticities may, again, lead to inaccurate calculations of economic growth effects.

III. Simulating aggregation bias in GDP-luminosity elasticities

The idea that luminosity changes as economic activity fluctuates seems so intuitive that the underlying mechanisms are hardly discussed by applied studies. The seminal HSW (2012, p.999) paper noted that:

> "consumption of nearly all goods in the evening requires lights. As income rises, so does lights usage per person, in both consumption activities and many investment activities."

Subsequent studies rarely note that luminosity is now observed at either 1.30am (VIIRS) or 3am (DMSP); well after the usual time of evening consumption.[7] Which activities generate light detectable from 830 km into space is also rarely discussed. Relevant to this second point is an experiment on DMSP spatial accuracy that temporarily illuminated wilderness areas (to ground-truth with background light ruled out) in winter when nights were darkest (Tuttle et al.,

[7] DMSP satellites have unstable orbits, observing Earth earlier as they age. For example, the local overpass time of satellite F18 was ca. 8pm in 2012; by 2019 it had shifted three hours earlier, to ca. 5pm. The 12-hour orbit yields corresponding pre-dawn observations so local overpass times of ca. 3am (for F15 and F16) are used by Ghosh et al (2021) to generate DMSP time-series post-2012.

2013). Detection from space required a bank of 1000-watt high pressure sodium lamps (large lamps of about 25 kg each, often used in big warehouses) with reflective shields to help direct light skyward. Even then, lights were detected on only half the nights.

Perhaps due to the timing of observation and type of lights needed to be detected from space, relationships between changes in economic outcomes and changes in luminosity come mostly from the extensive margin (illuminated area) not the intensive margin. This was first shown in India, where survey-measured poverty varied with changes in lit area but not with changes in brightness (Gibson et al., 2017) and the finding was repeated in Indonesia (Gibson et al., 2023). One caveat is these studies use DMSP data, which has many sources of temporal inconsistency (unstable orbits observing ever-earlier as satellites age, unrecorded changes in sensor amplification over the lunar cycle, lack of inter-satellite calibration), and so it is possible that such data are simply too crude to measure intensive margin effects (as suggested by Gibson et al., 2020). Also, India and Indonesia are early in their urban transition, so extensive margins may change in more dramatic ways than in already urbanized countries.

Yet already-urbanized economies also show this pattern. We use an 8-year panel (2012-19) of 3109 counties in the United States to estimate GDP-luminosity elasticities with the more accurate VIIRS data.[8] The within estimator of the elasticity is 0.18 (with a standard error of 0.03, clustered by county), as we show later in Table 3 below. If we partition the sum of lights into the extensive margin (lit area) and the intensive margin (average radiance of lit areas) the link with GDP changes is entirely on the extensive margin, with an elasticity that is 9-times larger than the intensive margin elasticity (standard errors clustered by county in ( )):

$$\ln(GDP)_{it} = 12.7 + \underset{(0.04)}{0.27} \ln(lit\ area)_{it} + \underset{(0.03)}{0.03} \ln(avg\ radiance)_{it} + \gamma_i + \delta_t + u_{it}$$

Thus, while annual changes in illuminated area predict annual changes in county-level GDP, changes in (conditional) average radiance do not. Extensive margin changes have ratchet

[8] Specifically, we use the masked median version 2.1 VIIRS nighttime lights (VNL) from Elvidge et al (2021).

effects; once an area becomes illuminated it rarely goes back to darkness.[9] Only the most dire long-run circumstances (e.g. Detroit) are likely to have street lights go dark. Intensive margin changes are more easily reversed (e.g. using lights for fewer or more hours).

Sensitivity to extensive margin changes fits with established patterns. For example, cross-sectional ('between') elasticities greatly exceed time-series ('within') ones, aligning with the fact that extensive margin differences are a larger component of between variation (while intensive margin changes contribute more to within variation). Notably, the between estimator values of the GDP-luminosity elasticities for metro areas in developing countries are up to 10-times larger than the elasticities for non-metro/rural areas (Gibson et al., 2021) and these metro areas have rapidly changing extensive margins of luminosity. Likewise, the elasticities of agricultural GDP with respect to luminosity are just one-tenth the size of corresponding elasticities for urban activity (Gibson and Boe-Gibson, 2021) and there is very little extensive margin change in luminosity coming from the agricultural sector.

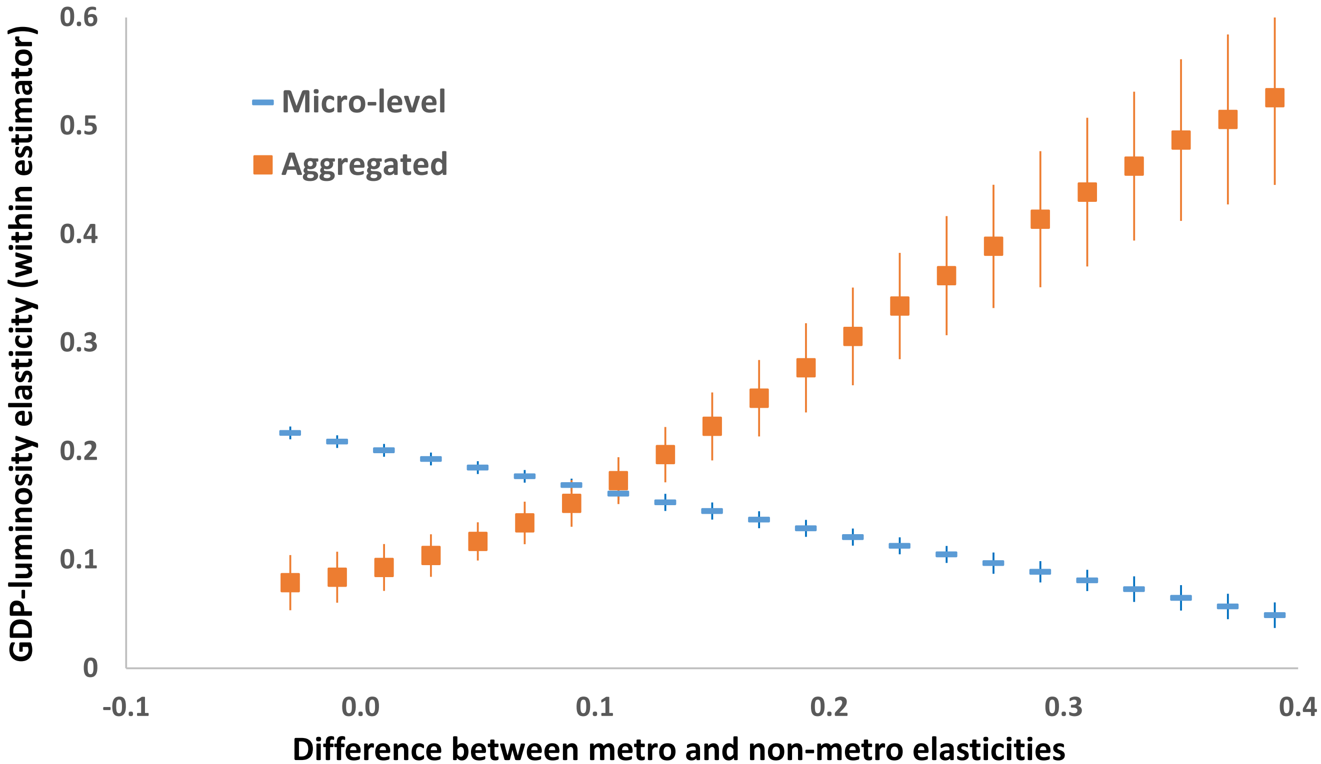
Consider an example: the Ukraine war raises Iowa farmer incomes (via grain prices) so farmers visit the county diner more often. The diner responds by extending closing time to 9pm and installing outdoor tables illuminated by fluorescent tubes. Satellites are unlikely to detect the extra activity; the change in opening hours is before the 1.30am observation time for VIIRS (or the ca. 3am time for DMSP post-2012) and the new lights will be too weak to detect from space. Richer farmers also buy winter homes in Arizona ('snowbirds') and that housing demand shock should be detectable; outdoor halogen lights usually stay on all night at construction sites for security reasons, and once built, street lights stay on all night in growing Phoenix suburbs. So the extensive margin change should be more easily detectable from space than is the intensive margin change.

[9] HSW considered ratchet effects and found none; elasticities were 0.27 in either direction. In contrast, we find elasticities of 0.14 for positive shocks and 0.22 for negative ones; significantly different at $p<0.01$.

In Appendix B we describe a simulation model with the features described here. The log-luminosity for 1000 micro units comes from Zipf's Law plus a random error; the largest one-tenth of these micro units are designated as metro areas, the remaining nine-tenths are designated as non-metro. The second period luminosity is from a random growth process, so ranks of the micro units hardly change over time. The log of GDP for each micro unit is predicted from linear functions of log-luminosity plus random errors in each period, with the sectoral elasticities allowed to vary between metro and non-metro areas. From the 1000 first differences that are estimated for the overall sample we obtain the micro-level elasticities as: $(\partial(ln(GDP_{t=2}) - ln(GDP_{t=1}))/\partial(ln(luminosity_{t=2}) - ln(luminosity_{t=1})))$. To show spatial aggregation effects, every ten units (1 metro and 9 non-metro) are grouped into one larger unit, akin to provinces (each having ten subordinate units). The aggregate elasticity is estimated from the first differences of these 100 groups. We use 1000 bootstrap replications.

Patterns of micro-level and aggregate elasticities of GDP growth with respect to lights growth are shown in Figure 1. If sectoral elasticities for metro areas happen to greatly exceed the elasticities for non-metro areas (as the empirical results for China show below) spatial aggregation inflates estimates of the first-differenced growth elasticities. For example, a 19-point metro-to-non-metro gap in sectoral elasticities (China's average gap over 2012-19) yields growth elasticities estimated from spatially aggregated units that are more than double the elasticities estimated from micro-level units. This aggregation-sensitivity follows from Jensen's inequality; logged convex combinations of growth rates exceed convex combinations of log-transformed values (see Appendix B).

Figure 1. Varying aggregation-sensitivity of the elasticities of GDP growth with respect to luminosity growth



Note: bars show 95% confidence intervals based on 1000 bootstrap replications

In contrast, if there are only small or even negative metro/non-metro gaps in sectoral elasticities (as we show below for the US) spatial aggregation reduces the growth elasticities. A gap of -3 points in the sectoral elasticities (which is the average for the US over 2012-19) yields growth elasticities estimated from spatially aggregated data that are less than one-half the size of the growth elasticities estimated with the micro-level data. This downward bias is due to the combined effects of the non-linear aggregation and of the random errors in the GDP-luminosity relationship, as we show in Appendix B. Overall, these simulation results suggest that within estimator values of the GDP-luminosity elasticities are not invariant to spatial aggregation, notwithstanding the assumption of invariance that is made by elasticity transfer studies such as those summarized in Appendix A. Moreover, spatial aggregation may raise the elasticities in some contexts and reduce them in others, so there is no rule-of-thumb that could be used as an adjustment when researchers use the elasticity transfer approach.

IV. Data

We create sub-national panel databases for the two most populous developing countries with official GDP data at the third (China) or second (Indonesia) sub-national level.[10] We also use data for the United States, as an urbanized country with high statistical capability (to address GDP measurement error concerns). Aggregating to province/state level from county (China and US) or district (Indonesia) level involves 16-76-fold reductions in observations (there are n=2342, n=3109 and n=497 cross-sectional units at our most spatially disaggregated level in China, the US and Indonesia). This spatial aggregation still leaves sufficient sample size for subsequent estimation, especially with time-series of roughly twenty years for each country.

[10] India has privately provided district-level data that are marketed as "GDP" (e.g. see http://www.indicus.net/) but these are not official government data. Few countries have county-level real GDP data; for example, while Europe reports real GDP data disaggregated to NUTS2 level, this involves combining adjacent counties in the UK into one unit and is at the level of provinces/states in several other countries.

Appendix C has full details on the sources of the GDP data, and on the spatial units and their classification into metro and non-metro areas.

Our three sources of NTL data are DMSP, and two flavors of VIIRS: VNL from the Earth Observation Group, who also worked extensively on DMSP (Elvidge et al, 2021); and, NASA Black Marble (BM) data (Román et al, 2018). The BM data differ from VNL in four ways: 16-bit precision (n=65,536 values) rather than 14-bit (while DMSP is just 6-bit); users can choose detection angles, with near-nadir, off-nadir, and all-angles composites; separate data for snow-free and snow-covered nights (as snow alters reflectance); and a stray-light correction was implemented from 2012 onwards (while for VNL it was from 2014). Prior comparisons using Vuong (1989) tests show that models using BM data to predict county GDP are closer to truth than models using VNL data (Zhang and Gibson, 2022), and models using VNL data are closer to truth than models using DMSP data (Gibson, 2021).

We provide full details on the NTL sources in Appendix D. The key contrasts are in terms of spatial precision and temporal consistency. At nadir, the VIIRS sensor is 45-times more spatially precise than DMSP (Elvidge et al., 2013). Off-nadir, DMSP does even worse due to angular viewing effects that enlarge ground footprints (Abrahams et al., 2018). DMSP data are top-coded, so densely populated and brightly lit areas get the same values as lower density, dimmer areas (Bluhm and Krause, 2022). Blurring and top-coding cause DMSP data to make local areas seem more economically alike (thus understating spatial inequality) versus what either BM, GDP, or VNL show (Zhang et al., 2023; Mathen et al., 2024). DMSP data are temporally inconsistent, from unrecorded changes in sensor amplification, orbits that observe earlier as satellites age, and no inter-satellite calibration. While VNL and BM can simultaneously handle both dim and bright lights, with dynamic range covering almost seven orders of magnitude, the range for DMSP is only two orders of magnitude. Finally, annual composites of BM (VNL) data use 3-times (2-times) as many nights per year as the DMSP composites, providing a better basis to measure changes in annual economic activity.

We provide descriptive statistics on our data in Appendix E. The inverse-hyperbolic sine is used to create logarithms of the sum-of-lights for each county-year (or district-year), rather than adding arbitrary constants for dealing with zeroes. The sum-of-lights naturally aggregates, facilitating our comparisons of the elasticities at different geographic levels.[11] Given that identification of within-estimator elasticities comes from time-series variation, we report time trends for each variable in addition to the means and standard deviations. While luminosity in the United States is trending downwards (for all sensors and time periods), even with ongoing real GDP growth of about two percent per annum, the trend rates of increase for China and Indonesia are consistent with their sum-of-lights doubling roughly every decade.

V. Results

The estimated GDP-luminosity elasticities at different aggregation levels, and for metro versus non-metro areas, and for different periods and NTL sources, are in Table 1 for China, Table 2 for Indonesia, and Table 3 for the United States. Each table has the same structure, where comparing panels A and D shows how elasticities change with spatial aggregation, comparing panels B and C shows differences by metro status, comparing the first three columns shows variation by time period for the same NTL source (DMSP), and the last three columns show differences by NTL data source for the same time period (2012-19).

a. Spatial aggregation

The GDP-luminosity elasticities are greatly inflated by spatial aggregation of the data for China. Across all five columns in Table 1, the province-level elasticities are, on average, six-times larger than the county-level ones. Using just the two most accurate NTL sources (VNL and BM), elasticities are four-times higher after spatial aggregation, rising from the 0.12-0.18

[11] Gibson et al. (2024) show the sum-of-lights outperforms lights/area when predicting China's county GDP.

range to 0.44-0.66.[12] The same pattern holds for Indonesia, where aggregation to provincial-level yields an almost eight-fold rise in elasticities for the 2012-19 period.

In contrast, spatial aggregation with US data reduces GDP-luminosity elasticities by one-half on average. The reduced elasticities if the county-level data are spatially aggregated are especially apparent for the two VIIRS data products; the state-level elasticities are just one-third the size of the county-level elasticities with these data. The VIIRS data should have less measurement error than DMSP data (see below) so this aggregation pattern is unlikely to be an artefact.

The aggregation-sensitivity of the GDP-luminosity elasticities is consistent with the simulation results in Appendix B. The simulations showed that if there was a more elastic relationship between luminosity and GDP growth in metro areas than in non-metro areas, spatial aggregation would raise the elasticities for the overall sample that pools both sectors. Conversely, if the sectoral gap is zero or negative (the GDP-luminosity elasticity is higher in non-metro areas) spatial aggregation reduces the elasticities estimated from the pooled sector sample. The research design used here has the *same* luminosity and GDP data either grouped together to create state- or province-level aggregates or else disaggregated to their finest level for counties or districts. Hence, this evidence is more compelling than in prior approaches that simply compared elasticities from different studies that happened to be at different levels of spatial aggregation (and that also had different samples and so on).

[12] These differences are at least 30-times larger than the (county-clustered) standard errors.

Table 1. Within estimator results for the GDP-lights elasticity for county-level units and for provinces in China, 2000-2019

	DMSP Stable Lights Annual Composites			VIIRS Annual Composites, 2012-2019	
	2000-2019	**2000-2011**	**2012-2019**	**VNL V2.1**	**Black Marble**
			A. All county-level and district-level spatial units		
ln(sum of lights)	0.069	0.037	0.050	0.184	0.117
	(0.007)	(0.007)	(0.005)	(0.011)	(0.011)
Number of observations	77286	46840	30446	18736	18736
R-squared (Within)	0.025	0.008	0.017	0.067	0.028
			B. Urban cores (districts)		
ln(sum of lights) – metro	0.376	0.251	0.277	0.324	0.255
	(0.041)	(0.044)	(0.038)	(0.049)	(0.057)
Number of observations	9801	5940	3861	2376	2376
R-squared (Within)	0.115	0.061	0.063	0.105	0.054
			C. Non-metro areas (counties and similar county-level units)		
ln(sum of lights) – non-metro	0.066	0.035	0.049	0.179	0.112
	(0.007)	(0.007)	(0.005)	(0.012)	(0.011)
Number of observations	67485	40900	26585	16360	16360
R-squared (Within)	0.025	0.008	0.018	0.066	0.028
			D. Province-level		
ln(sum of lights)	0.410	0.248	0.422	0.660	0.436
	(0.073)	(0.094)	(0.118)	(0.121)	(0.283)
Number of observations	1023	620	403	248	248
R-squared (Within)	0.222	0.113	0.165	0.285	0.079

Notes: Results are based on strongly balanced panels, and all models include fixed effects for each satellite and year and for each spatial unit (either n=2342 county-level and district-level units in panels A-C or n=31 provinces in panel D), estimated using the `reghdfe` *Stata* command. Standard errors are in parentheses and are clustered by county-level unit for panels A-C and by province for panel D. The DMSP annual composites are based on Baugh et al (2010) and Ghosh et al (2021), VNL V2.1 annual composites are the masked medians series based on Elvidge et al (2021) and the Black Marble composites are the all-angle, weighted average of snow-free and snow-covered nights based on series described in Román et al (2018). The dependent variable is ln(GDP).

Table 2. Within estimator results for the GDP-lights elasticity for district-level units and for provinces in Indonesia, 2004-2019

	DMSP Stable Lights Annual Composites			VIIRS Annual Composites, 2012-2019	
	2004-2019	**2004-2011**	**2012-2019**	**VNL V2.1**	**Black Marble**
			A. All district-level spatial units		
ln(sum of lights)	0.022	0.020	0.006	0.012	0.022
	(0.007)	(0.009)	(0.002)	(0.004)	(0.008)
Number of observations	11972	5511	6461	3976	3976
R-squared (Within)	0.007	0.007	0.007	0.014	0.024
			B. Urban cores (kota)		
ln(sum of lights) – metro	0.080	0.068	0.004	-0.009	0.011
	(0.040)	(0.035)	(0.017)	(0.024)	(0.073)
Number of observations	2401	1127	1274	784	784
R-squared (Within)	0.013	0.013	0.000	0.002	0.001
			C. Non-metro areas (kabupaten)		
ln(sum of lights) – non-metro	0.022	0.020	0.006	0.014	0.022
	(0.007)	(0.009)	(0.002)	(0.004)	(0.008)
Number of observations	9571	4384	5187	3192	3192
R-squared (Within)	0.007	0.008	0.007	0.022	0.028
			D. Province-level		
ln(sum of lights)	-0.014	-0.018	0.048	0.093	0.162
	(0.078)	(0.032)	(0.031)	(0.039)	(0.048)
Number of observations	850	408	442	272	272
R-squared (Within)	0.001	0.002	0.035	0.117	0.218

Notes: Results for the 2012-19 sub-samples are based on strongly balanced panels. There are some missing values of GDP prior to 2010. All models have fixed effects for each satellite and year and for each spatial unit (either n=497 districts in panels A-C or n=34 provinces in panel D), estimated using the `reghdfe` *Stata* command. Standard errors are in parentheses and are clustered at district level for panels A-C and by province for panel D. The DMSP annual composites are based on Baugh et al (2010) and Ghosh et al (2021), VNL V2.1 annual composites are the masked medians series based on Elvidge et al (2021) and the Black Marble composites are the all-angle snow-free observations based on series described in Román et al (2018). The dependent variable is ln(GDP).

Table 3. Within estimator results for the GDP-lights elasticity for county-level units and States, United States, 2001-2019

	DMSP Stable Lights Annual Composites			**VIIRS Annual Composites, 2012-2019**	
	2001-2019	**2001-2011**	**2012-2019**	**VNL V2.1**	**Black Marble**
			A. All county-level spatial units		
ln(sum of lights)	0.156	0.055	0.051	0.176	0.201
	(0.029)	(0.010)	(0.013)	(0.030)	(0.026)
Number of observations	96362	55945	40417	24872	24872
R-squared (Within)	0.075	0.018	0.013	0.047	0.070
			B. Counties in urban cores		
ln(sum of lights) – metro	0.160	0.129	0.041	0.120	0.170
	(0.046)	(0.020)	(0.013)	(0.027)	(0.034)
Number of observations	13204	7666	5538	3408	3408
R-squared (Within)	0.030	0.019	0.007	0.025	0.046
			C. Non-metropolitan counties		
ln(sum of lights) – non-metro	0.156	0.054	0.050	0.180	0.198
	(0.030)	(0.010)	(0.014)	(0.031)	(0.028)
Number of observations	83158	48279	34879	21464	21464
R-squared (Within)	0.078	0.019	0.013	0.049	0.068
			D. States (including District of Columbia)		
ln(sum of lights)	0.099	0.041	0.029	0.066	0.045
	(0.068)	(0.030)	(0.016)	(0.052)	(0.076)
Number of observations	1581	918	663	408	408
R-squared (Within)	0.088	0.034	0.008	0.023	0.006

Notes: Results are based on strongly balanced panels, and all models include fixed effects for each satellite and year and for each spatial unit (either n=3109 county-level units in panels A-C or n=51 States (including the District of Columbia) in panel D), estimated using the `reghdfe` *Stata* command. Standard errors are in parentheses and are clustered by county-level unit for panels A-C and by State for panel D. The DMSP annual composites are based on Baugh et al (2010) and Ghosh et al (2021), VNL V2.1 annual composites are the masked medians series based on Elvidge et al (2021) and the Black Marble composites are the all-angle, weighted average of snow-free and snow-covered nights based on series described in Román et al (2018). The dependent variable is ln(GDP).

b. Metro versus non-metro

The GDP-luminosity elasticities differ greatly between metro and non-metro areas in the two developing countries. This pattern is especially clear in Table 1 for China, where the metro elasticities are 6-times larger than the non-metro ones using the DMSP data. With the two VIIRS data products, the metro elasticities are twice as large. The metro/non-metro gaps in predictive power are of similar scale. Over the period that we study, GDP growth rates in metro and non-metro areas were the same but luminosity grew faster in non-metro areas.[13] Hence, with faster luminosity growth but similar GDP growth the non-metro elasticities are lower.[14] Likewise, the pre-2012 period in Indonesia had GDP growth in non-metro areas in the absence of any trend growth in DMSP luminosity, while the metro areas had fast growth in both GDP and luminosity, so the metro elasticity is higher. Since then, the metro and non-metro elasticities have been more alike (and very close to zero).

For the United States, elasticities are more alike between metro and non-metro areas. This is especially for the full period, and also since 2012 with DMSP and BM data.[15] This similarity across sectors may reflect luminosity reaching some saturation level (and is now falling in both metro and non-metro areas and for all three NTL sources) even with ongoing GDP growth. This goes beyond the non-linearities shown by Bluhm and McCord (2022), where luminosity responded less to changes in GDP at higher baseline level of GDP (noting that they estimated luminosity-GDP elasticities rather than GDP-luminosity ones). Several mechanisms may underlie divergent GDP and luminosity trends, such as responses to rising concern about

[13] These luminosity trends may reflect China's recent adoption of a more dispersed form of urbanization, with migration to big cities like Beijing and Tianjin increasingly restricted while smaller non-metro urbanized parts of counties grew more rapidly due to relaxed restrictions on in-coming migrants (Li et al., 2024).
[14] Local officials in county-level cities have incentives to convert agricultural land into urban use (Lichtenberg and Ding, 2009) and so faster luminosity growth in these non-metro units (partly from land conversion) with no faster GDP growth may imply that there are efficiency costs of this dispersed urbanization.
[15] In the pre-2012 period, trend GDP growth rates in metro areas exceeded those for non-metro areas and metro areas also had less negative rates of luminosity growth.

light pollution (Falchi et al, 2019), and changes in lighting types (e.g. LEDs replacing halogen) that affect satellite detection of luminosity.

c. Time period

Comparisons by time period are possible with DMSP data but not with VIIRS due to the short VIIRS time-series. For the most spatially disaggregated units, only slight changes in the elasticities are seen from the pre-2012 subperiod to the 2012-19 subperiod. Aggregating to the province/state level makes temporal changes in the elasticities seem larger; the provincial-level elasticity for China is 70% larger in the second subperiod, while for the US the state-level elasticity is 30% smaller than in the first subperiod. The other apparent pattern is that elasticities based on DMSP are higher in the two-decade full time series in the first column of the tables, than in the sub-period shorter time-series. Given the various sources of measurement error in the DMSP data, the hypothesis of elasticities varying over time (or varying with the length of the time-series) is a question to examine in future when the newer generation sensors have longer time-series.

d. Remote sensing system

For the 2012-19 period we can compare elasticities from the three NTL sources. The DMSP data yield smaller elasticities, especially with the spatially disaggregated data. For the US and China, county-level elasticities when using DMSP data are only 30% the size of elasticities estimated with the more accurate BM and VNL data but after aggregation to province/state level the DMSP elasticities average 67% of the size of the elasticities estimated from BM and VNL data. Measurement error is a plausible explanation for this pattern because blurring and top-coding in DMSP data create spatially mean-reverting measurement errors (Gibson, 2021; Kim et al, 2024). Spatial aggregation is inherently mean-reverting so effects of these errors become less apparent in spatially aggregated data.

In Appendix F we examine biases from measurement errors in luminosity data. We especially aim to assess effects of measurement errors on estimated aggregation sensitivity of GDP-luminosity elasticities. A prior study artificially aggregated US county-level GDP and DMSP data into different configurations of subnational regions, finding that the elasticities were stable as the number of subnational regions decreased (Bluhm and McCord, 2022). This is contrary to the aggregation-sensitivity that we find and so understanding how measurement errors may affect results using DMSP data is important.

There are two pathways for measurement errors in luminosity data to bias estimates of the GDP-luminosity elasticity—an attenuation effect and by mean reversion. Following HSW (2012), let $x, y,$ and z denote logs of observed luminosity, true GDP, and observed GDP, and

$$z_{jt} = y_{jt} + \varepsilon_{z,jt} \tag{1}$$

$$x_{jt} = \beta y_{jt} + \varepsilon_{x,jt}, \tag{2}$$

$\varepsilon_{z,jt}$ and $\varepsilon_{x,jt}$ are measurement errors in GDP and luminosity, with variances σ_z^2 and σ_x^2 (for area j, year *t*). True GDP variance is σ_y^2. Regressing observed GDP on observed luminosity:

$$z_{jt} = \psi x_{jt} + u_{jt} \tag{3}$$

yields an estimated GDP-luminosity elasticity that is attenuated with respect to the true $1/\beta$:

$$plim\ \hat{\psi}_{x,LS} = \frac{cov(x,z)}{var(\mathrm{x})} = \frac{\beta\sigma_y^2}{\beta^2\sigma_y^2+\sigma_x^2} = \frac{1}{\beta}\left(\frac{\beta^2\sigma_y^2}{\beta^2\sigma_y^2+\sigma_x^2}\right) \tag{4}$$

if observed luminosity has any measurement error, $\sigma_x^2 > 0$. To overcome this issue, HSW (2012) use country statistical capacity ratings to impose parametric assumptions on GDP 'reliability ratios' $\left(\sigma_y^2/(\sigma_y^2+\sigma_z^2)\right)$ when estimating β. Recently, Kim et al. (2023) and Chor and Li (2024) attempt to mitigate attenuation bias by using Instrumental Variables (IV) for equation (3), with lagged luminosity as the instrument.[16]

[16] This IV strategy assumes no serial correlation in measurement errors for annual luminosity. Yet errors within the same areas may correlate across years (thus, not random). For example, North-South missing data patterns due to summer glare (Gibson, 2021) or errors from not adjusting for snow-cover (Zhang and Gibson, 2022) tend

This approach is incomplete for DMSP data because IV is only consistent for random measurement error, not non-classical error (Black et al, 2000). DMSP spatial mean-reversion, where $x_{DMSP,jt} = \theta + \lambda x_{VIIRS,jt} + v_{jt}$ for $\hat{\lambda} < 1$, creates non-classical errors; prior estimates of $\hat{\lambda}$ range from 0.4 (Kim et al, 2024) to 0.7 (Gibson, 2021). With mean reversion, the right-hand side of equation (4) becomes: $\frac{1}{\lambda\beta}\left(\frac{\lambda^2\beta^2\sigma_y^2}{\lambda^2\beta^2\sigma_y^2+\sigma_x^2}\right)$ and so a two-step procedure is needed to recover the true $1/\beta$. First, equation (3) is estimated by IV, using more accurate BM data to instrument for potentially error-ridden DMSP data. Second, $\hat{\psi}_{x,IV}$ is adjusted to allow for effects of mean-reverting errors, using $1/\hat{\lambda}$, to give a mean-reverting-adjusted IV (MRA-IV).

Table 4 has county-level and province/state-level GDP-luminosity elasticities for China and the US from OLS and instrumental variables. The ratios of province/state-level estimates to county-level estimates are used to indicate aggregation bias (see, also, Figure 1). China's province-to-county ratio is 8.4 using OLS on equation (3) with DMSP data. The ratio falls to 3.7 using our MRA-IV procedure. Thus, aggregation-sensitivity of GDP-luminosity elasticities persists, even allowing for effects of DMSP measurement errors. If we, instead, use VNL data as the potentially error-ridden luminosity measure (with BM data, again, as the instrument) the province-to-county ratio is 3.6 (using OLS) or 3.2 (using IV). A lesser effect of mitigating measurement error, compared to the results using DMPS, implies that VNL data are less error-ridden than are DMSP data.

to affect the same areas each year. Our strategy of using luminosity estimates from a more accurate source (BM) as the IV for the estimates from the less accurate source (DMSP) does not depend on this assumption.

Table 4. Aggregation-sensitivity of the GDP-lights elasticity, with and without adjustment for measurement error

A. China (2012-19, 2342 county-level units and 31 provinces)

	DMSP			VNL V2.1		
Measurement error	**County (1)**	**Province (2)**	**Ratio (2)/(1)**	**County (4)**	**Province (5)**	**Ratio (5)/(4)**
Ignored	0.050	0.422	8.4	0.184	0.660	3.6
Mitigated	0.117	0.437	3.7	0.173	0.549	3.2

B. United States (2012-19, 3109 county-level units and 51 states (including DC))

	DMSP			VNL V2.1		
Measurement error	**County (1)**	**State (2)**	**Ratio (2)/(1)**	**County (4)**	**State (5)**	**Ratio (5)/(4)**
Ignored	0.051	0.029	0.6	0.176	0.066	0.4
Mitigated	0.202	0.045	0.2	0.442	0.068	0.2

Notes: Elasticity values for "measurement error ignored" rows are from Table 1 (China) and Table 3 (United States). The "measurement error mitigated" results are estimated using MRA-IV for DMSP data, and IV for VNL data, where in both cases the Black Marble luminosity estimates are used as the instrumental variables. Details on the IV results are in Appendix F. Other notes are as in Tables 1 and 3.

The US results show little aggregation bias if measurement errors in DMSP data are ignored, with a state-to-county ratio of 0.6. This ratio falls to 0.2 if the DMSP measurement errors are dealt with using our MRA-IV approach. Therefore, findings from Bluhm and McCord (2020) of stable elasticities across different (artificial) aggregation levels may have been affected by DMSP measurement errors. The same state-to-county ratio, of 0.2, comes from using VNL data (with IV) for the US. Thus, the pattern where spatial aggregation reduces estimates of the GDP-luminosity elasticity for the US (and increases them for China) seems to be robust to the possible presence of measurement errors in luminosity data.

VI. Discussion and Conclusions

Applied economists increasingly use satellite-detected night lights data to estimate treatment effects in settings where conventional economic indicators, like GDP, are unavailable. These estimates are often converted into economic growth terms by transferring GDP-luminosity elasticities from elsewhere. Our results caution against this procedure. The elasticities differ by level of spatial aggregation, between metro and non-metro areas, and between the various remote sensing systems providing the data—especially due to measurement errors in DMSP data. The spatial aggregation issue particularly matters. GDP-luminosity elasticity estimates are typically from spatially aggregated data (as they need GDP data) but are often applied to treatment effects on luminosity from spatially disaggregated data. Using the elasticities from aggregated data to proxy for relationships between local economic growth and luminosity will distort calculated local growth effects if estimated elasticities vary with aggregation.

Our simulations and empirical results show that spatial aggregation reduces estimated GDP-luminosity elasticities in settings like the United States, and raises them in settings like China and Indonesia. The gap in elasticities for metro versus non-metro areas determines the direction of bias. This gap should be larger when luminosity grows rapidly (as in China and Indonesia, where trends imply a doubling every decade) because illuminating unlit areas (extensive margin changes) is most predictive of changes in local GDP, and metro areas in such countries rapidly expand on extensive margins. GDP-luminosity elasticities estimated from aggregated data will be too large for such places, as a proxy for elasticities that apply at local levels. For example, the 20 elasticity-transfer studies we reviewed used GDP-luminosity elasticities averaging 0.4, when calculating local economic growth effects of treatments (Appendix A). Yet our most spatially disaggregated data for China and Indonesia shows elasticities of about 0.1, so the elasticity transfer approach will exaggerate calculated growth effects. The opposite bias for the US should matter less because US analyses are less reliant on luminosity data, given the other data sources available to measure local economic growth. Thus, biased GDP-luminosity elasticities from using aggregated data are most likely to distort evaluations for developing countries.

Our results also pose a challenge for studies that do not calculate local growth effects by transferring GDP-luminosity elasticities and, instead, simply report treatment effects on luminosity as if this is a sufficient measure of economic activity (e.g. Elliot et al., 2015). An assumed proportionality between GDP and luminosity (e.g. as in equation (2)) may not hold. The luminosity data available for the past decade are for readings in the very early hours of the morning, between 1.30am and ca. 3am, and so these data cannot measure usual evening activities of households or of many firms. Also, the lights observable from 830 km away are not ordinary household lights or the lights that small enterprises might use. Perhaps because of this, changes in local GDP vary far more with local changes in extensive margins of satellite-detected luminosity than with intensive margin changes. Hence, changes in annual NTL data will only reflect changes in certain types of economic activity (such as converting unlit areas to illuminated areas, especially for lights that stay on all night, like street lights). Many spatially disaggregated units, such as villages or pixels, may experience little change in extensive margins of luminosity from year-to-year even as local economic activity fluctuates.

Aggregating over many such local areas will yield stronger relationships, especially in countries where extensive margins of luminosity expand rapidly, even though those stronger relationships may not hold at local levels.

References

Abrahams, Alexei, Christopher Oram and Nancy Lozano-Gracia (2018) Deblurring DMSP night-time lights: A new method using Gaussian filters and frequencies of illumination. *Remote Sensing of Environment* 210(1): 242-258.

Akter, Sonia (2023) Australia's Black Summer wildfires recovery: A difference-in-differences analysis using nightlights. *Global Environmental Change* 83(1): 102743.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad (2021) Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the SHRUG open data platform. *The World Bank Economic Review* 35(4): 845-871.

Beyer, Robert, Tarun Jain and Sonalika Sinha (2023) Lights out? COVID-19 containment policies and economic activity. *Journal of Asian Economics* 85(1): 101589.

Black, Dan, Mark Berger, and Frank Scott (2000) Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association* 95(451): 739-748.

Bluhm, Richard, and Melanie Krause (2022). Top lights: Bright cities and their contribution to economic development. *Journal of Development Economics* 157: 102880.

Bluhm, Richard, and Gordon McCord (2022) What can we learn from nighttime lights for small geographies? Measurement errors and heterogeneous elasticities. *Remote Sensing* 14(5): 1190.

Chanda, Areendam, and Justin Cook (2022) Was India's demonetization redistributive? Insights from satellites and surveys. *Journal of Macroeconomics* 73(1): 103438.

Chen, Xi, and William Nordhaus (2019) VIIRS night time lights in the estimation of cross-sectional and time-series GDP. *Remote Sensing* 11(9): 1057-1068.

Chodorow-Reich, Gabriel, Gita Gopinath, Prachi Mishra and Abhinav Narayanan (2020) Cash and the economy: Evidence from India's demonetization. *The Quarterly Journal of Economics* 135(1): 57-103.

Chor, Davin, and Bingjing Li (2024) Illuminating the effects of the US-China tariff war on China's economy. *Journal of International Economics* 150: 103926.

Corral, Leonardo, and Maja Schling (2017) The impact of shoreline stabilization on economic growth in small island developing states. *Journal of Environmental Economics and Management* 86(1): 210-228.

Elliott, Robert, Eric Strobl, and Puyang Sun (2015) The local impact of typhoons on economic activity in China: A view from outer space. *Journal of Urban Economics* 88(1): 50-66.

Elvidge, Christopher, Kimberly Baugh, Mikhail Zhizhin, and Feng-Chi Hsu (2013) Why VIIRS data are superior to DMSP for mapping night time lights. *Proceedings of the Asia-Pacific Advanced Network* 35(1): 62-69.

Elvidge, Christopher, Mikhail Zhizhin, Tilottama Ghosh, Feng-Chi Hsu, and Jay Taneja (2021) Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing* 13(5): 922.

Falchi, Fabio, Riccardo Furgoni, Terrel Gallaway, Nataliya Rybnikova, Boris Portnov, Kimberly Baugh, Pierantonio Cinzano, and Christopher Elvidge (2019) Light pollution in USA and Europe: The good, the bad and the ugly. *Journal of Environmental Management* 248: 109227.

Ghosh, Tilottama, Kimberly Baugh, Christopher Elvidge, Mikhail Zhizhin, Alexey Poyda, and Feng-Chi Hsu (2021) Extending the DMSP nighttime lights time series beyond 2013. *Remote Sensing* 13(24): 5004.

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

Gibson, John, Gaurav Datt, Rinku Murgai, and Martin Ravallion (2017) For India's rural poor, growing towns matter more than growing cities. *World Development* 98: 413-429.

Gibson, John, Susan Olivia, and Geua Boe-Gibson (2020) Night lights in economics: Sources and uses. *Journal of Economic Surveys* 34(5): 955-980.

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

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

Gibson, John, Yi Jiang, and Bambang Susantono (2023) Revisiting the role of secondary towns: How different types of urban growth relate to poverty in Indonesia. *World Development* 169: 106281.

Gibson, John, Xiaoxuan Zhang, Albert Park, Yi Jiang and Xi Li (2024) Remotely measuring rural economic activity and poverty: Do we just need better sensors? *Working Paper* No 2023-08, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University, https://cei.ier.hit-u.ac.jp/Japanese/2024/03/22/WP2023-8.pdf.

Goldblatt, Ran, Kilian Heilmann and Yonatan Vaizman (2020) Can medium-resolution satellite imagery measure economic activity at small geographies? Evidence from Landsat in Vietnam. *The World Bank Economic Review* 34(3): 635-653.

Gong, Da, Zhuocheng Shang, Yaqin Su, Andong Yan, and Qi Zhang (2024) Economic impacts of China's zero-COVID policies. *China Economic Review* 83(C): 102101.

Heger, Martin Philipp, and Eric Neumayer (2019) The impact of the Indian Ocean tsunami on Aceh's long-term economic growth. *Journal of Development Economics* 141(1), 102365.

Henderson, Vernon, Adam Storeygard, and David Weil (2012) Measuring economic growth from outer space. *American Economic Review* 102(2): 994-1028.

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

Joseph, Iverson-Love (2022) The effect of natural disaster on economic growth: Evidence from a major earthquake in Haiti. *World Development* 159(1): 106053.

Kim, Jihee, Kyoochul Kim, Sangyoon Park, and Chang Sun (2023) The economic costs of trade sanctions: Evidence from North Korea. *Journal of International Economics* 145: 103813.

Kim, Bonggeun, John Gibson, and Geua Boe-Gibson (2024) Measurement errors in popular night lights data may bias estimated impacts of economic sanctions: Evidence from closing the Kaesong Industrial Zone. *Economic Inquiry* 62(1): 375-389.

Kocornik-Mina, Adriana, Thomas McDermott, Guy Michaels, and Ferdinand Rauch (2020) Flooded cities. *American Economic Journal: Applied Economics* 12(2): 35-66.

Li, Zhihui, Chao Li, John Gibson, and Xiangzheng Deng (2024) Two decades of inter-city migration in China: The role of economic, natural and social amenities. *Working Paper* No. 24/05 Department of Economics, University of Waikato.

Lichtenberg, Erik, and Chengri Ding (2009) Local officials as land developers: Urban spatial expansion in China. *Journal of Urban Economics* 66(1): 57-64.

Mathen, Christopher, Siddhartha Chattopadhyay, Sohini Sahu, and Abhijit Mukherjee (2024) Which nighttime lights data better represent India's economic activities and regional inequality? *Asian Development Review* 41(2): 193-217.

Miranda, Juan, Oscar Ishizawa and Hongrui Zhang (2020) Understanding the impact dynamics of windstorms on short-term economic activity from night lights in Central America. *Economics of Disasters and Climate Change* 4(3): 657-698.

Roberts, Mark (2021) Tracking economic activity in response to the COVID-19 crisis using nighttime lights–The case of Morocco. *Development Engineering* 6(1): 100067.

Román, Miguel, Zhuosen Wang, Qingsong Sun, Virginia Kalb, Steven Miller, Andrew Molthan, Lori Schultz et al (2018) NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment* 210(1): 113-143.

Storeygard, Adam (2016) Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *The Review of Economic Studies* 83(3): 1263-1295.

Tuttle, Benjamin, Sharolyn Anderson, Paul Sutton, Christopher Elvidge, and Kim Baugh (2013) It used to be dark here. *Photogrammetric Engineering & Remote Sensing* 79(3): 287-297.

Vuong, Quang (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57(2): 307-333.

Zhang, Xiaoxuan, and John Gibson (2022) Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. *Remote Sensing* 14(5): 1282.

Zhang, Xiaoxuan, John Gibson, and Xiangzheng Deng (2023) Remotely too equal: Popular DMSP night-time lights data understate spatial inequality. *Regional Science Policy & Practice* 15(9): 2106-25.

Appendices

A. Selected Studies Using Transferred GDP-Luminosity Elasticities

We reviewed papers that especially were published in top general interest and top field journals in economics to find examples of studies that rely on transferring GDP-luminosity elasticities from one context to another. The details are in Table A.1. Across the 20 studies that we include, the mean value of the GDP-luminosity elasticity transferred from elsewhere is 0.41 (standard deviation 0.17, with a range from 0.2 to 0.9). The modal value of the transferred elasticities is 0.3, reflecting the high reliance on the GDP-luminosity elasticity estimate originally made by HSW (2012).

To provide quantitative evidence on the spatial mismatch between the "origin" studies where the GDP-luminosity elasticities had originally been estimated and the "destination" studies that these elasticities are transferred into, to enable calculation of the local economic growth effects of various treatments we categorized each study according to a 6-level geographic code with the following categories: 1) national, 2) region/province/state, 3) district, 4) city/county, 5) village, and 6) pixel/micro-grid (<5km). The average origin study for the GDP-luminosity elasticities has a geographic aggregation level of 1.7 on our 6-point scale (where "1" is most aggregated and "6" is least aggregated) and the modal value is 1 (in other words, national or cross-country data are the most common type of data used to provide the elasticities, while the average across all the studies is somewhere between the national level and the first sub-national level). The destination studies where the elasticities are "imported" into are far more spatially disaggregated, having an average value of 4.4 on our 6-point scale, which corresponds to the city/county level. Thus, the typical elasticity transfer study involves taking a GDP-luminosity from an origin setting that is about three steps more aggregated on our 6-point scale than is the destination context.

There is also a temporal gap, between the time-series for the samples used in the origin studies and the samples used in the destination studies. On average, this gap is about five years, at the midpoint of each of the time-series. Once there is a longer time-series of luminosity data from VIIRS the issue of temporal evolution of the GDP-luminosity elasticities could be examined with these more accurate data. Currently, however, only the time-series for DMSP is sufficiently long enough to study changes in elasticities over time, and the various sources of temporal instability in DMSP data (such as lack of calibration and shifts in orbit as satellites age) limit the usefulness of these data for studying temporal changes in the GDP-luminosity elasticities.

Table A.1: Selected Elasticity Transfer Studies

Details on the Publication			Origin elasticity (∂GDP/∂luminosity)					Destination setting (∂luminosity/∂treatment)			
Authors	Journal	Date	Source	NTL	Period	Unit	Value	NTL	Period	Unit	Treatment
Hodler & Raschky	*QJE*	2014	Self	DMSP	A: 1992-2008	Region (2)	0.38	DMSP	A: 1992-2009	District (3)	Regional favoritism
Storeygard	*ReStud*	2016	HSW	DMSP	A: 1992-2008	National (1)	0.30	DMSP	M: 2016-2017	City (4)	Transport costs
Bazzi et al	*AER*	2016	O&G	DMSP	A: 2000	District (3)	0.40	DMSP	A: 2000	Village (5)	Skills transfer
Villa	*EDCC*	2016	Self	DMSP	A: 2001-2004	Region (2)	0.58	DMSP	A: 2001-2004	Municipio (3)	Social transfers
Corral & Schling	*JEEM*	2017	Self	DMSP	A: 1992-2012	National (1)	0.53	DMSP	A: 1992-2012	Pixel (6)	Beach repairs
De Luca et al	*JDE*	2018	Self	DMSP	A: 1992-2013	Region (2)	0.39	DMSP	A: 1992-2013	Sub-dist (4)	Nepotism
Lee	*J.Urb.Econ*	2018	HSW	DMSP	A: 1992-2008	National (1)	0.30	DMSP	A: 1992-2013	Microgrid (6)	Sanctions
Mamo et al	*JDE*	2019	HSW	DMSP	A: 1992-2008	National (1)	0.30	DMSP	A: 1992-2012	District (3)	Mining
Eberhard-Ruiz et al	*JDE*	2019	H&R	DMSP	A: 1992-2013	Region (2)	0.38	DMSP	A: 1992-2013	City (4)	Market integration
Chodorow-Reich et al	*QJE*	2020	HSW	DMSP	A: 1992-2008	National (1)	0.30	VIIRS	Q: 2016-2017	District (3)	Demonetization
Kocornik-Mina et al	*AEJ: Appl*	2020	HSW	DSMP	A: 1992-2008	National (1)	0.30	DSMP	A: 2003-2008	Pixel (6)	Floods
Miranda et al	*EconDisCC*	2020	Self	DSMP	A: 1992-2013	National (1)	0.33	DSMP	A: 1992-2011	Pixel (6)	Hurricanes
Dreher et al	*World Dev*	2021	HSW	DMSP	A: 1992-2008	National (1)	0.30	DMSP	A: 2001-2012	District (3)	Chinese aid
Roberts	*Dev Eng*	2021	Self	VIIRS	Q: 2012-2020	National (1)	0.30	VIIRS	M: 2012-2020	City (4)	COVID-19
Cook & Shah[a]	*RESTAT*	2022	Self	DMSP	A: 2005	State (2)	0.32	DMSP	M: 2016-2017	District (3)	Public works
Felbermayr et al	*World Dev*	2022	Self	DMSP	A: 1992-2013	National (1)	0.37	DMSP	A: 1992-2013	Grid (5)	Weather anomalies
Kim et al	*J. Int Econ*	2023	Self	VIIRS	A: 2013-2018	County (4)	0.42	VIIRS	Q: 2014-2019	County (4)	Sanctions
Linsenmeier	*JEEM*	2023	GOBL	VIIRS	A: 2015-2016	District (3)	0.20	VIIRS	A: 2015	Grid (5)	Climate variability
Akter	*GlobEnvCh*	2023	Gibs.	VIIRS	A: 2015	District (3)	0.84	VIIRS	M: 2017-2021	Block (6)	Wildfires
Gong et al	*ChinaEcRev*	2024	Mart.	DMSP	A: 1992-2008	National (1)	0.86	VIIRS	M: 2019-2022	County (4)	Zero COVID

Notes : Source Self is estimated by the authors in another context, HSW is Henderson et al (2012), O&G is Olivia and Gibson (2015), H&R is Hodler and Raschky (2014), Mart. is Martinez (2022), Gibs. is Gibson (2021), GOBL is Gibson et al (2021). Numbers in ( ) in the Unit columns correspond to the six levels of the geographic aggregation coding that we use.

[a]Cook & Shah report that a 1 unit increase in average state level luminosity is associated with an increase of INR 1,785 in per capita net state domestic product. Using values they report, we calculate a linear elasticity at the mean.

References for studies cited in Table A.1

Akter, Sonia (2023) Australia's Black Summer wildfires recovery: A difference-in-differences analysis using nightlights. *Global Environmental Change* 83(1): 102743.

Bazzi, Samuel, Arya Gaduh, Alexander Rothenberg, and Maisy Wong (2016) Skill transferability, migration, and development: Evidence from population resettlement in Indonesia. *American Economic Review* 106(9): 2658-2698.

Chodorow-Reich, Gabriel, Gita Gopinath, Prachi Mishra and Abhinav Narayanan (2020) Cash and the economy: Evidence from India's demonetization. *The Quarterly Journal of Economics* 135(1): 57-103.

Cook, Justin, and Manisha Shah (2022) Aggregate effects from public works: Evidence from India. *Review of Economics and Statistics* 104(4): 797-806.

Corral, Leonardo, and Maja Schling (2017) The impact of shoreline stabilization on economic growth in small island developing states. *Journal of Environmental Economics and Management* 86(1): 210-228.

De Luca, Giacomo, Roland Hodler, Paul Raschky, and Michele Valsecchi (2018) Ethnic favoritism: An axiom of politics? *Journal of Development Economics* 132(1): 115-129.

Dreher, Axel, Andreas Fuchs, Roland Hodler, Bradley Parks, Paul Raschky and Michael Tierney (2021) Is favoritism a threat to Chinese aid effectiveness? A subnational analysis of Chinese development projects. *World Development* 139(1): 105291.

Eberhard-Ruiz, Andreas, and Alexander Moradi (2019) Regional market integration in East Africa: Local but no regional effects? *Journal of Development Economics* 140(1): 255-268.

Felbermayr, Gabriel, Jasmin Gröschl, Mark Sanders, Vincent Schippers, and Thomas Steinwachs (2022) The economic impact of weather anomalies. *World Development* 151(1): 105745.

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

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

Gong, Da, Zhuocheng Shang, Yaqin Su, Andong Yan, and Qi Zhang (2024) Economic impacts of China's zero-COVID policies. *China Economic Review* 83(C): 102101.

Henderson, Vernon, Adam Storeygard, and David Weil (2012) Measuring economic growth from outer space. *American Economic Review* 102(2): 994-1028.

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

Kim, Jihee, Kyoochul Kim, Sangyoon Park, and Chang Sun (2023) The economic costs of trade sanctions: Evidence from North Korea. *Journal of International Economics* 145: 103813.

Kocornik-Mina, Adriana, Thomas McDermott, Guy Michaels, and Ferdinand Rauch (2020) Flooded cities. *American Economic Journal: Applied Economics* 12(2): 35-66.

Lee, Yong Suk (2018) International isolation and regional inequality: Evidence from sanctions on North Korea. *Journal of Urban Economics* 103(1): 34-51.

Linsenmeier, Manuel (2023) Temperature variability and long-run economic development. *Journal of Environmental Economics and Management* 121(1): 102840.

Mamo, Nemera, Sambit Bhattacharyya, and Alexander Moradi (2019) Intensive and extensive margins of mining and development: Evidence from Sub-Saharan Africa. *Journal of Development Economics* 139(1): 28-49.

Martinez, Luis (2022) How much should we trust the dictator's GDP growth estimates? *Journal of Political Economy* 130(10): 2731-2769.

Miranda, Juan, Oscar Ishizawa and Hongrui Zhang (2020) Understanding the impact dynamics of windstorms on short-term economic activity from night lights in Central America. *Economics of Disasters and Climate Change* 4(3): 657-698.

Olivia, Susan, and John Gibson (2015) Value added from space? Empirical evidence from Indonesia. *Unpublished Manuscript*.

Roberts, Mark (2021) Tracking economic activity in response to the COVID-19 crisis using nighttime lights–The case of Morocco. *Development Engineering* 6(1): 100067.

Storeygard, Adam (2016) Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *The Review of Economic Studies* 83(3): 1263-1295.

Villa, Juan (2016) Social transfers and growth: Evidence from luminosity data. *Economic Development and Cultural Change* 65(1): 39-61.

B. Simulating Aggregation Bias in GDP-luminosity Elasticities

We use simulation models to study how spatial aggregation alters the 'within estimator' values of GDP-luminosity elasticities. Our simulations estimate the elasticities using first-differenced models, as a simple way to estimate relationships between GDP growth and luminosity growth. We also develop analytical expressions to explain the patterns coming from the simulations.

Our main simulations allow for size differences between two types of micro-level units – 'metro' and 'non-metro' (where the average metro unit is n-times larger, in terms of GDP or luminosity, than the average non-metro unit). These simulations show the most detectable effects of spatial aggregation; for example, for a situation like China where metro elasticities exceed non-metro ones, aggregation causes a clear upward bias in the first-differenced elasticity. Conversely, in a situation like the United States, where metro elasticities are slightly smaller than non-metro ones, the simulations show that spatial aggregation causes the first differenced elasticity to be biased downwards. We describe the gap between the micro-level and aggregated elasticities as 'bias' because in the elasticity transfer studies described in Appendix A, the elasticities estimated from more spatially aggregated data (the "origin" studies) are being used as a proxy for unavailable GDP-luminosity elasticities at the more micro level (of the "destination" studies) corresponding to the aggregation level of the spatial units where the effects of treatment on luminosity are estimated. Thus, any distortion to micro-level elasticities is treated as an aggregation bias in our framework because the micro-level GDP-luminosity elasticity would be the one used to estimate local economic growth effects of the treatment if these micro-level elasticities were available.

The specifics of the simulation are as follows (Stata code is reported below). Log-luminosity for 1000 micro units in the first period comes from Zipf's Law plus a random error; the largest one-tenth of these micro units are designated as metro areas, the remaining nine-tenths are designated as non-metro areas. Second period luminosity is generated from a random growth process, so the ranks of the micro units hardly change over time. The log-GDP for each micro unit is predicted from linear functions of log-luminosity plus random errors in each period, the sectoral elasticities differ between the metro units and the non-metro units (the predictive functions are the inverse of production function relationships, for reasons set out by HSW (2012)). We vary these metro and non-metro elasticities in our simulations. The 1000 first differences yield micro-level growth elasticities: $(\partial(ln(GDP_{t=2}) - ln(GDP_{t=1}))/\partial(ln(luminosity_{t=2}) - ln(luminosity_{t=1})))$. In other words, these are the 'within estimator' elasticities based on time-series variation, which are the correct ones for comparing with our empirical results for China, Indonesia and the US, and for informing the vast majority of elasticity transfer studies. To show the effects of spatial aggregation every ten units (1 metro and 9 non-metro) are grouped into one larger unit, that can be thought of as akin to a province or state. The aggregate elasticity is then estimated from the first differences of these 100 groups. We use 1000 bootstrap replications.

To link to our subsequent analytical expressions, we can express the predicting equations as:

$$\text{Metro: } ln(GDP_{it}) = \gamma + \alpha\ ln(luminosity_{it}) + u_{it}$$

$$\text{Non-metro: } ln(GDP_{it}) = \gamma + \beta\ ln(luminosity_{it}) + u_{it}$$

In the simulations we vary the values of α and β, with particular interest in how the micro-level and aggregate elasticities for the first-differenced data evolve as the gap between α and β varies, or reverses. We impose the constraint that $\alpha + \beta = 0.41$, to reflect the range of differences we see in the metro and non-metro elasticities in the empirical results from our three samples. This also serves to put a feasible limit on the number of different simulations conducted.

Table B.1 provides a summary from the main set of simulations, where different values of α and β are used, and the micro level within estimator elasticity is estimated using all 1000 units, and then the aggregate level elasticity is estimated from the 100 aggregated units. The estimates for each row of the table are each repeated 1000 times (with standard deviations of these replicates providing the standard errors reported in the table).

Table B.1: Within estimator GDP-luminosity elasticities for micro-level units and for spatially aggregated units, based on differences in the metro/non-metro elasticity gap.

			Micro-level		Aggregated	
Metro elasticity (100 units)	Non-metro elasticity (900 units)	Metro to non-metro gap	Within estimator elasticity	Std Error	Within estimator elasticity	Std Error
0.40	0.01	0.39	0.049	0.006	0.526	0.041
0.39	0.02	0.37	0.057	0.006	0.506	0.040
0.38	0.03	0.35	0.065	0.006	0.487	0.038
0.37	0.04	0.33	0.073	0.006	0.463	0.035
0.36	0.05	0.31	0.081	0.005	0.439	0.035
0.35	0.06	0.29	0.089	0.005	0.414	0.032
0.34	0.07	0.27	0.097	0.005	0.389	0.029
0.33	0.08	0.25	0.105	0.004	0.362	0.028
0.32	0.09	0.23	0.113	0.004	0.334	0.025
0.31	0.10	0.21	0.121	0.004	0.306	0.023
0.30	0.11	0.19	0.129	0.004	0.277	0.021
0.29	0.12	0.17	0.137	0.004	0.249	0.018
0.28	0.13	0.15	0.145	0.004	0.223	0.016
0.27	0.14	0.13	0.153	0.004	0.197	0.013
0.26	0.15	0.11	0.161	0.003	0.173	0.011
0.25	0.16	0.09	0.169	0.003	0.152	0.011
0.24	0.17	0.07	0.177	0.003	0.134	0.010
0.23	0.18	0.05	0.185	0.003	0.117	0.009
0.22	0.19	0.03	0.193	0.003	0.104	0.010
0.21	0.20	0.01	0.201	0.003	0.093	0.011
0.20	0.21	-0.01	0.209	0.003	0.084	0.012
0.19	0.22	-0.03	0.217	0.003	0.079	0.013

Note: Results are based on 1000 bootstrap replications for each set of metro/non-metro gaps.

When there is a large (positive) gap between the metro and non-metro elasticities, spatial aggregation increases the within estimator values of the growth elasticities. For example, with an elasticity gap of 0.19, which is about what is reported below for China, the elasticity estimated with the spatially aggregated data is about double the magnitude of what is estimated with the micro-level data (0.28 versus 0.13). This aggregation bias has deterministic and random components. For the deterministic part, consider the two groups (metro and non-metro) with θ proportion having an elasticity of α and $(1 - \theta)$ proportion having an elasticity of β (in our simulation, $\theta = 0.1$). It is then instructive to see what happens when luminosity ($\ln x$) eventually doubles in each area and relatedly the predicted GDP ($\ln y$) increases by α and β for each group respectively. We first consider the less realistic case where there is no size differential for the two groups (that is, the GDP and luminosity of each metro area is, on average, no larger than for each non-metro area, so n=1). We then consider the more realistic case, where the metro areas are, on average, larger than the non-metro areas (in terms of luminosity or GDP, by some factor n>1). With these assumptions we have three within estimator values of φ, the GDP-luminosity elasticity, to consider:

$$\varphi_{\text{micro}} = \theta \frac{ln(1+\alpha)x - lnx}{\ln 2x - \ln x} + (1-\theta)\frac{ln(1+\beta)x - lnx}{\ln 2x - \ln x} = \frac{1}{\ln 2}[\theta \ln(1+\alpha) + (1-\theta)\ln(1+\beta)] \quad (1)$$

$$\varphi_{\text{aggregated}} = \frac{ln[\theta(1+\alpha)x + (1-\theta)(1+\beta)x] - ln2x}{\ln 4x - \ln 2x} = \frac{1}{\ln 2}\ln[\theta(1+\alpha) + (1-\theta)(1+\beta)] \quad (2)$$

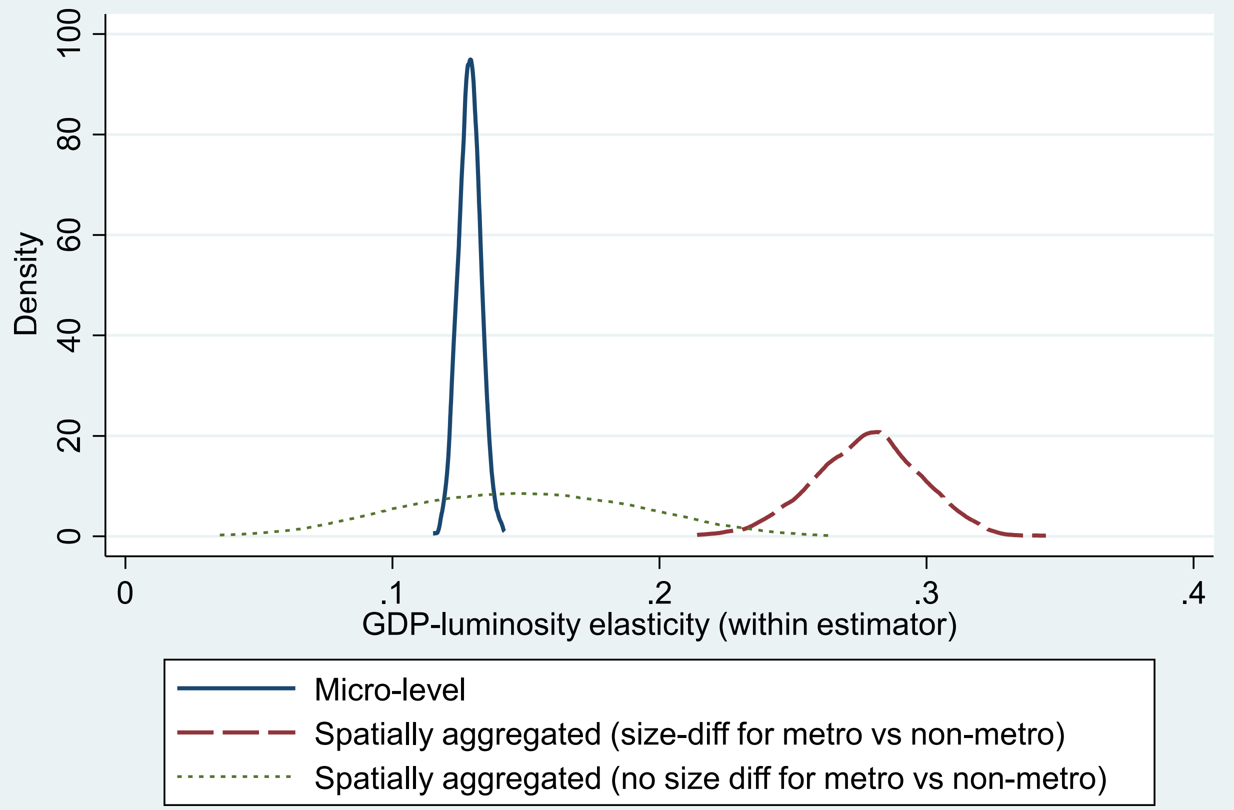
$$\varphi_{\text{aggregated, size diff}} = \frac{ln[\theta(1+\alpha)nx + (1-\theta)(1+\beta)x] - ln(n+1)x}{\ln 2(n+1)x - \ln(n+1)x} = \frac{1}{\ln 2}\ln[\frac{n\theta(1+\alpha) + (1-\theta)(1+\beta)}{n\theta + (1-\theta)}] \quad (3)$$

These assumptions yield the following rankings: $\varphi_{\text{micro}} < \varphi_{\text{aggregated}}$ if $\alpha > \beta$. This inequality shows that spatial aggregation will inflate the GDP-luminosity elasticity under these conditions, where this follows from Jensen's inequality; the log transformed value of a convex combination of two growth rates is greater than the convex combination of the log-transformed values. Once we allow for size differences between the metro and non-metro areas, we have a further ranking, as follows: $\varphi_{\text{micro}} < \varphi_{\text{aggregated}} < \varphi_{\text{aggregated, size diff}}$ if $\alpha > \beta$ and $n > 1$.

The density charts in Figure B.1 from the simulations (for a metro/non-metro gap of 0.19, which is similar to the values that we estimate below for China) show that the sharpest evidence on aggregation bias comes from the size-differentiated case. Specifically, with these parameter values the simulations do not produce any overlap between the distribution of the micro-level elasticities, which are tightly bunched around 0.13, and the elasticities from the spatially aggregated data, which are less tightly bunched but are centered on a value of 0.28. In contrast, if we use the less realistic assumption, of equal average size of the metro units and the non-metro units, there is still a slight upward bias due to spatial aggregation that is apparent, with the elasticities from the spatially aggregated data centered around 0.15 (compared to elasticities centered on 0.13 coming from the micro-level data). However, there is a much less clear distinction between the micro-level elasticities and the aggregated elasticities because the distribution of the simulations for the (size-undifferentiated) aggregated units is so wide (spanning a range from 0.05 to 0.25) that it overlaps the entire distribution of the smoothed density function for the micro-level elasticities.

A driving force for these inequalities in the growth elasticities is the differences in the sectoral elasticities, between the metro and the non-metro areas. In contrast, if $\alpha = \beta$ equations (1) to (3) would not imply differences between the growth elasticities estimated at the micro-level and those estimated from the spatially aggregated data (irrespective of the value of n). This may initially seem to be a puzzle when considering empirical results for the United States reported below, where the sectoral elasticities are close to equal (being slightly larger for non-metro areas than for metro areas), yet the spatial aggregation is associated with a downward bias in the growth elasticities. This apparent puzzle can be resolved once the contribution from the random part of the simulations is taken account of, which we explain using equations (4) to (6).

Figure B.1: Simulation results for metro/non-metro gap of 0.19 (matching the China case)



If we use the previous expressions from equations (1) to (3), which considered the doubling over time of luminosity in each micro-level unit, but we now also add the random terms from each of the prediction equations, we obtain:

$$\varphi_{\text{micro}} = \frac{1}{\ln 2}[\theta \ln(1+\alpha) + (1-\theta)\ln(1+\beta)] + \mathrm{u} \quad \text{where} \quad \mathrm{E(u)} = 0 \tag{4}$$

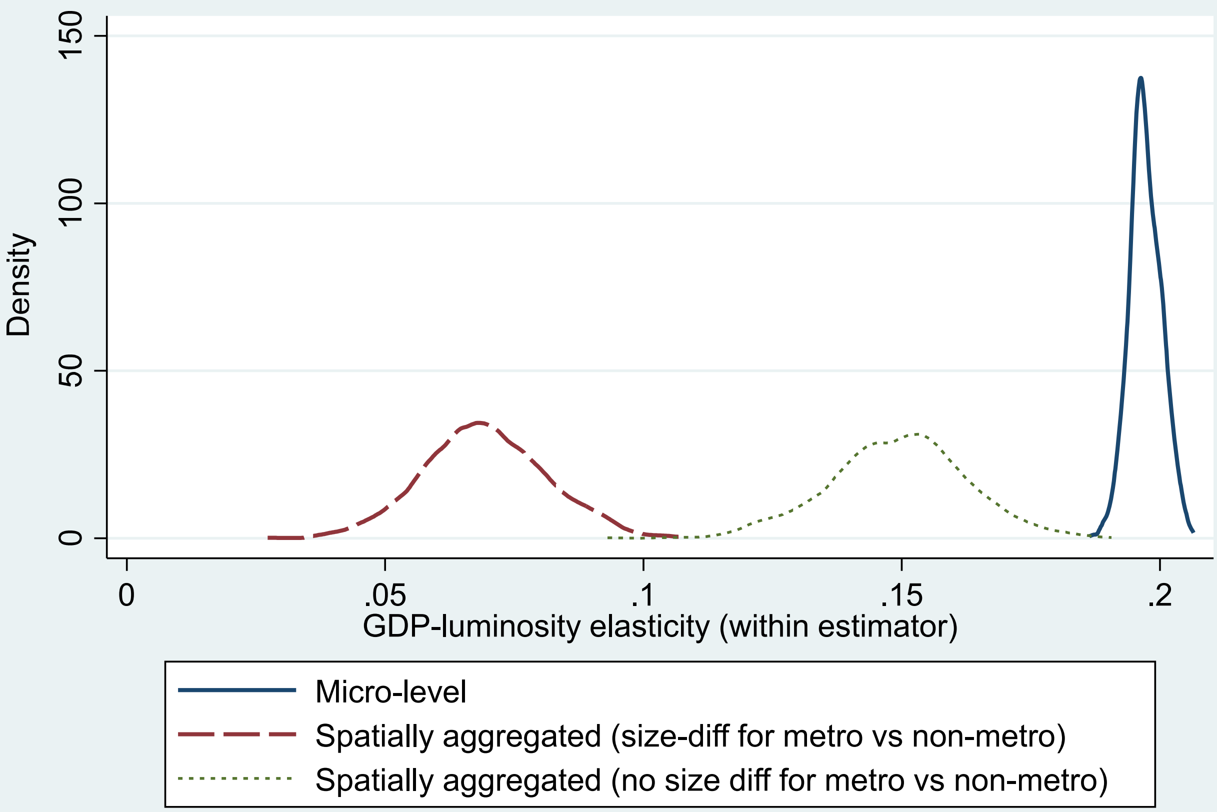
$$\varphi_{\text{aggregated}} = \frac{1}{\ln 2}\ln[\theta(1+\alpha) + (1-\theta)(1+\beta) + \mathrm{u}] \tag{5}$$

$$\varphi_{\text{aggregated, size diff}} = \frac{1}{\ln 2}\ln\left[\frac{\mathrm{n}\theta(1+\alpha)+(1-\theta)(1+\beta)}{n\theta+(1-\theta)} + \mathrm{u}\right] \tag{6}$$

In equation (4), based on the within estimator of the elasticity estimated from micro-level units, the random error is added to the weighted average of the logged terms. In contrast, the errors are spatially aggregated and then the logarithm is taken in equations (5) and (6). Jensen's inequality comes into play once again to create a wedge between the micro-level and spatially aggregated elasticities. This gives the ordering: $\varphi_{\text{micro}} > \varphi_{\text{aggregated}} = \varphi_{\text{aggregated, size diff}}$, even if $\alpha = \beta$ for any n. In other words, when the sectoral elasticities are the same, the spatial aggregation reduces the value of the GDP-luminosity growth elasticity due to the combined effect of the random terms and the non-linear aggregation.

This ordering is shown in Figure B.2, with the simulation results for a setting like the United States where the metro elasticity is slightly smaller than the non-metro one (a gap of -0.03). The elasticity estimated from micro-level units is tightly bunched around a value of 0.2. In contrast, the elasticity estimated from the spatially aggregated data is centered on 0.15 (if there is no size differentiation between metro and non-metro – that is, $n = 1$), or around 0.08 if we allow the metro units to be larger, on average, than the non-metro units.

Figure B.2: Simulation results for metro/non-metro gap of -0.03 (matching the US case)



To summarize these simulation results, the within estimator values of the GDP-luminosity elasticities are not invariant to spatial aggregation, contrary to the underlying assumption made by elasticity transfer studies such as those summarized in Appendix A. When micro-level data are spatially aggregated there can be significant shifts in the distribution of the elasticities – biased upwards when there is a large gap in sectoral elasticities (being greater for metro units than for non-metro units), and biased downwards when there is little (or even a negative) sectoral gap in the elasticities. Consequently, a GDP-luminosity elasticity that is estimated from spatially aggregated data will not be the appropriate elasticity to use when calculating the local economic growth impact of some treatment whose effects on luminosity have been estimated with spatially disaggregated data.

Stata Code for the Main Simulations

```
log using "C:\Current\Lights\Growth\Simulation\Table_B1.smcl"
* Luminosity and local economic growth -- aggregation simulation
* Chao Li
* 21/February/2024
* adding outer loop to Bonggeun's coding
******************************************************
** for micro unit simulation, Col (4) of Table B.1  **
******************************************************
program define p4, eclass
drop _all
set obs 1000
gen rank=_n
gen u=0.1*invnorm(uniform())
gen lrank=ln(rank)
gen size=10000000/(rank)
gen x1=ln(size)+0.1*invnorm(uniform())
**dx is independent of size
gen x2=x1+invnorm(uniform())
gen dx=x2-x1
** two groups (urban vs non-urban)
gen y1=10+(0.4-`0'/100)*x1 if rank<=100
replace y1=10+(0.01+`0'/100)*x1 if rank>100
gen y2=10+(0.4-`0'/100)*x2+u if rank<=100
replace y2=10+(0.01+`0'/100)*x2+u if rank>100
gen dy=y2-y1
reg dy dx
end
forval i = 0(1)21 {
clear
di `i'
simulate _b, reps(1000): p4 `i'
sum _b_dx
tempfile b4`i'
keep _b_dx
rename _b_dx b4`i'
gen id=_n
save `b4`i'', replace
}
* merge saved temps
use `b40', clear
forval i = 1(1)21 {
merge 1:1 id using `b4`i''
drop _merge
}
* final results
sum b4*, sep(0)
clear
***********************************************************
** for aggregated unit simulation, Col (6) of Table B.1  **
***********************************************************
program define p5, eclass
drop _all
set obs 1000
gen rank=_n
gen u=0.1*invnorm(uniform())
```

```
gen lrank=ln(rank)
gen size=10000000/(rank)
gen x1=ln(size)+0.1*invnorm(uniform())
**dx is independent of size
gen x2=x1+invnorm(uniform())
gen dx=x2-x1
** two groups (urban vs non-urban)
gen y1=10+(0.4-`0'/100)*x1 if rank<=100
replace y1=10+(0.01+`0'/100)*x1 if rank>100
gen y2=10+(0.4-`0'/100)*x2+u if rank<=100
replace y2=10+(0.01+`0'/100)*x2+u if rank>100
gen dy=y2-y1
*** for aggregation
for var y1 y2 x1 x2: gen X0=exp(X)
gen agg=1 if rank==1
forval i = 1(100)1000 {
replace agg=1 if rank==`i'
}
forval i = 2(1)100 {
replace agg=`i' if rank==`i' | rank==(100+`i')| rank==(200+`i')|
rank==(300+`i')| rank==(400+`i')| rank==(500+`i')| rank==(600+`i')|
rank==(700+`i')| rank==(800+`i')| rank==(900+`i')
}
collapse y10 y20 x10 x20, by(agg)
for var y10 y20 x10 x20: gen lX=ln(X)
gen dy=ly20-ly10
gen dx=lx20-lx10
reg dy dx
end
forval i = 0(1)21 {
clear
di `i'
simulate _b, reps(1000): p5 `i'
sum _b_dx
tempfile b5`i'
keep _b_dx
rename _b_dx b5`i'
gen id=_n
save `b5`i'', replace
}
* merge saved temps
use `b50', clear
forval i = 1(1)21 {
merge 1:1 id using `b5`i''
drop _merge
}
* final results
sum b5*, sep(0)

clear
log close
```

C. GDP Database Construction

C.1 China

China's general administrative hierarchy is province then prefecture; most prefectures further divide into counties for more rural areas or into districts for more urban areas. There are also county-level cities, that are distinct from the districts that make up the urban core of a prefecture, and some county-like units, such as banners. Prefecture names include the word 'City' even if most of the area is rural. For example, China's Household Responsibility System that permitted own-account family farming to resume from 1978 reputedly originated in Xiaogang village of Fengyang County, within Chuzhou City of Anhui Province. Xiaogang remains a decidedly rural area today yet is administered as part of a 'City'. In other words, 'City' in China's sub-national statistics is an administrative designation not a functional economic unit like a metropolitan statistical area in other countries. Moreover, parts of western China and a few other areas are not organized into prefectures underneath provinces and instead their 2nd level spatial units are either Leagues, Autonomous regions or Provincially Administered areas. Some of the non-prefectural areas eventually get upgraded to prefectural status so this division between the great majority of the country that is organized into prefectures and the remainder changes over time.[1]

These administrative details matter because there is no single unified source of data on China's sub-national GDP. Instead, there are different sources of data for the various spatial levels, with at least three types of publications from the National Bureau of Statistics (NBS) needed to build our database: the annual editions of the China Statistical Yearbook (county-level) (in Chinese it is *Zhongguo Xianyu Tongji Nianjian (Xianshi Juan)*), annual editions of the China City Statistical Yearbook (known in Chinese as *Zhongguo Chengshi Tongji Nianjian*), and annual editions of the Statistical Yearbook for each city or province (for example the Beijing Statistical Yearbook) (NBS, various dates). Each edition reports on GDP the previous year, so we use the 2001 to 2020 editions to obtain annual GDP data from 2000 to 2019. In addition we also needed to use the district-level local government economic communiques to get GDP data in some years for some of the areas that had recently been upgraded to district status.

The parts of China not organized into prefectures are not covered by the China City Statistical Yearbooks (given that they are not considered as a City) and are not always covered in the other sources like county yearbooks and the yearbooks issued by each province. In the 2020 census, 96.7% of China's resident population was included in the areas covered by the spatial units that we use. Areas not covered are shown in Figure C.1. Given that these non-prefectural areas are typically poorer than average the share of sub-national GDP covered by our sample is likely to be higher than the share of population covered; thus, our results can be thought of as pertaining to at least 97% of China's sub-national GDP.

[1] Chung and Lam (2004) provide some details on the variety of urban administrative units in China, with multiple examples of units being merged, split, reformed, upgraded and sometimes downgraded.

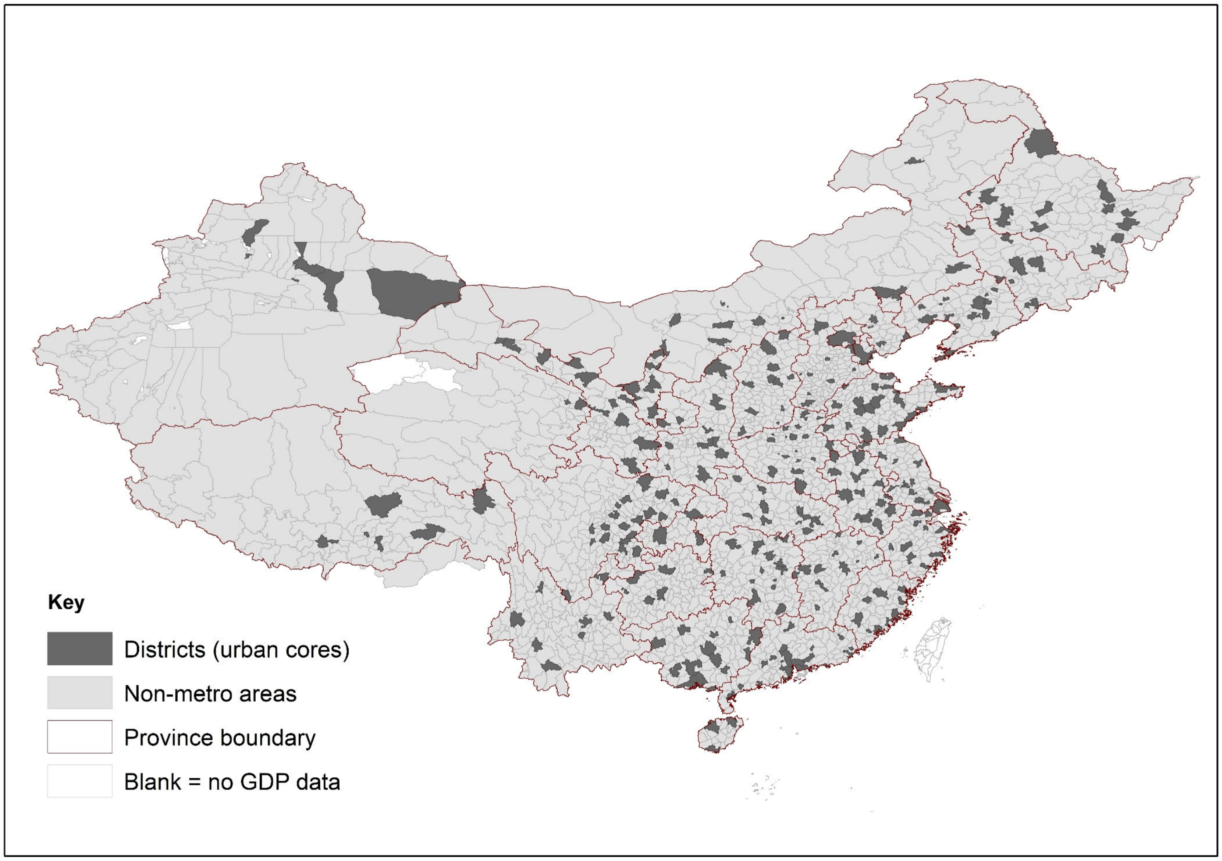


Fig C.1: China's sub-national spatial units used for the analysis

The second complication with China's sub-national data is that the administrative units can be reorganized according to the changing aims of the national government (the State Council). Usually the reorganization is for counties or county-level cities to be upgraded (and occasionally downgraded); since 2000, over 200 counties and county-level cities were upgraded to district status. This change in status can affect where the GDP data are reported (e.g. it may no longer be reported in the provincial yearbooks that provide county-by-county statistics). A related issue with these organizational changes is merging units; for example, Nanhui County was upgraded to district status and then in 2009 merged with Pudong District to form the Pudong New Area. Consequently, any GDP time-series that we could construct for Nanhui would end in 2008. In order to have a balanced panel, when any of the administrative areas are subsequently merged we also enforce the same aggregation on the earlier years to have a consistent 2000-19 time-series for each spatial unit. Relatedly, the established districts (i.e. not the ones that had been recently upgraded from county or county-level city status) that are adjacent to each other in the urban core of each prefecture are treated as one unit that is functionally the same city. For example, the night lights data show no gap between Xicheng and Dongcheng, the two innermost districts of Beijing that are almost entirely inside the 3rd Ring Road. In contrast, Beijing's five districts that were recently upgraded from county status are separately distinguished because each of these former counties has its own distinct urban core surrounded by a largely rural hinterland.

Overall, we have annual GDP data for each of n=2342 units at the 3rd level of the sub-national administrative hierarchy, where these units maintain a consistent spatial definition from 2000 to 2019. We additionally classify these units into either the urban core of the 2nd level unit that administers them or into the remaining areas. For established prefectures, the urban core is usually the merged area of districts in existence prior to 2000 (as opposed to counties recently upgraded to district status) while for recently established prefectures the urban core had typically been either a county-level city or a county prior to being upgraded to district status. Given the diversity of administrative units, we refer to urban cores (n=297) and the other 3rd level units (n=2045) are called non-metro areas. In the 2020 census there were 485 million people residing in the urban core areas, and about 880 million in the remaining areas, giving population densities of almost 700 persons per square kilometer in the urban cores and 100 per square kilometer in the non-metro areas (the national average is 150 per square kilometer). Some urban cores in provinces such as Xinjiang, Heilongjiang, and Tibet cover large areas (see Figure C.1), so population density of the urban cores in other more populous provinces is far higher than 700 per square kilometer.

C.2 Indonesia

The GDP data are from the Indonesian government's Central Bureau of Statistics (*Badan Pusat Statistik* or BPS for short) which provides annual estimates of Gross Regional Domestic Product (GRDP). The data that we use are in spatially and temporally real terms and have a 2010 base. The BPS calculate and report these GRDP figures at both the provincial level, and at the next level down in the administrative hierarchy, where this second level in the sub-national geography is either Regencies (also known as *Kabupaten*) or else Cities (also known as *Kota*). All parts of Indonesia are classified as either part of a Regency or part of a City (compared to the more complicated situation of only partial and time-varying coverage that affected the data for China). Previous comparisons of these GRDP data (we will often refer to them as GDP data for simplicity) with household survey estimates of consumption and labour force survey estimates of total employment and total wages at the second sub-national level show high degrees of agreement (Gibson et al, 2021).

For the decomposition into metropolitan or non-metropolitan areas we rely on the existing classification of second-level units as either City or Regency. In the 2010 census the n=98 units designated as City covered just under 30,000 square kilometers and had a population of 52 million, giving an average density of 1900 per square kilometer. The 399 unites that were classified as Regencies had a population of 185 million, distributed over about 1.9 million square kilometers, with an average population density of 95 per square kilometer. In other words, the non-metro areas have a population density similar to the non-metro areas of China but the metropolitan areas are more densely populated. The second sub-national level units, along with their designation as urban core or not, and the province boundaries are shown in Figure C.2.

To control for the splitting of regencies (or of cities), going forward from 2010 we re-aggregate back into the original spatial unit. In a few cases prior to 2010 we had to do the reverse, where one spatial unit had subsequently split we use the post-2010 fractions of GDP for the combined unit to disaggregate, in order to have a strongly balanced panel that extended from 2004 to 2019.

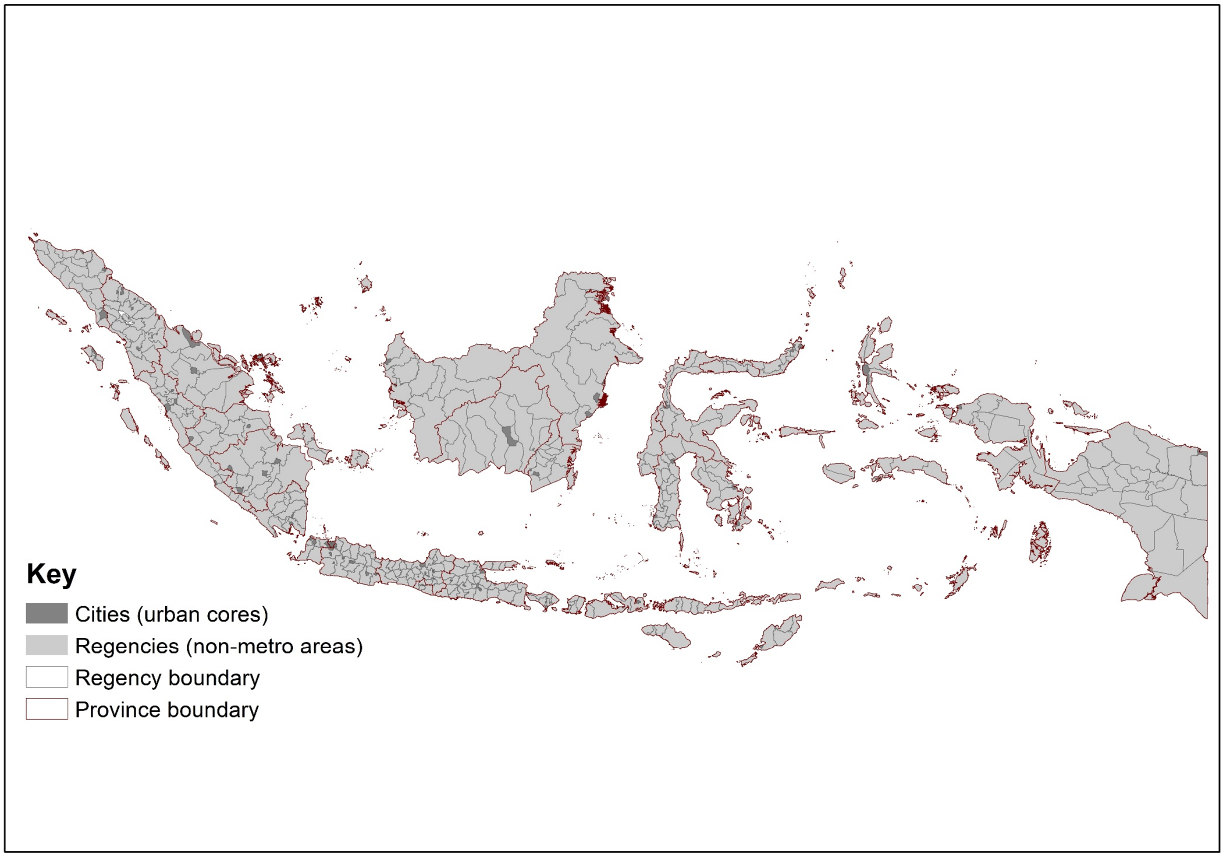


Fig C.2: Indonesia's sub-national spatial units used for the analysis

C.3 United States

The county-level GDP data for the United States are from the Bureau of Economic Analysis (BEA) available at the following link: https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas. We use the chained series in 2012 dollars. The annual estimates are provided separately for each county for the 2001 to 2019 period, with three broad exceptions. In Alaska the BEA combine some census areas in their reporting, and in Hawaii they combine Maui and Kalawao counties. It is for Virginia that the greatest amount of adjustment occurs; the BEA create 23 combination areas where one or two independent cities whose population in the 1980 census was less than 100,000 are combined with an adjacent county. The dissolve function in ArcGIS was used to modify a county-level shapefile so as to match these combination areas. There are n=3109 counties and combination areas (we refer to all of these as county-level units) with data available in each year. In comparison to China and Indonesia, the spatial definitions of counties in the United States are far more stable over time and so we do not need to do any merging of the county-level units to create a consistent time-series.

To classify county-level units as either urban cores or non-metropolitan we use the National Center for Health Statistics (NCHS) urban-rural classification scheme for counties (2013 edition) available at the following link: https://www.cdc.gov/nchs/data_access/urban_rural.htm. This has six groups of counties based on the population of the metropolitan statistical area (MSA) that the county may be associated with, and the centrality of that county to the MSA. We classify the core urban counties as those that are central metro counties in an MSA of 1 million population or greater that either contain the entire population of the largest principal city of the MSA or are completely contained within the largest principal city of the MSA, and those that are large fringe metro counties in an MSA of 1 million or more population. There are 426 of these urban cores, and the remaining 2683 county-level units are classified as non-metropolitan. In the 2010 census the urban cores had a combined population of 170 million (occupying 0.8 million square kilometers) and the non-metropolitan areas had a population of 139 million spread over 8.5 million square kilometers. The average population density in the urban cores was 480 persons per square kilometer and in the non-metropolitan areas was 30 persons per square kilometer. The county-level units, along with their designation as urban core or not, and the state boundaries are shown in Figure C.3.

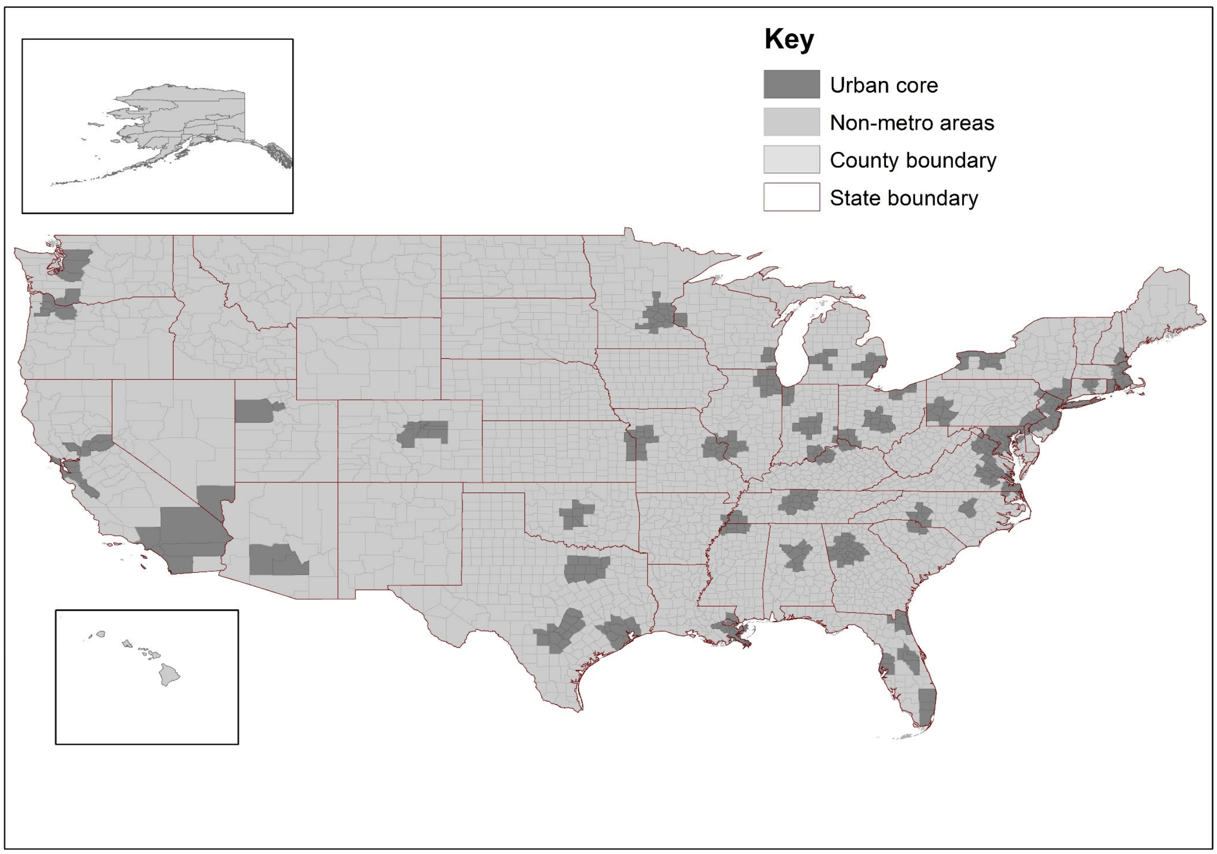


Fig C.3: United States spatial units used for the analysis (Alaska and Hawaii not to scale)

D. Night-time Lights Database Construction

Our database of night-time lights (NTL) is assembled from three sources. The first is annual composites from Defense Meteorological Satellite Program (DMSP) satellites F14, F15, F16 and F18 which cover each year from 2000 to 2019 (some years have two satellites providing data). The stable lights product provides 6-bit digital numbers (DN) ranging from 0–63 (with higher values indicting greater luminosity) for each 30 arc-second output pixel. Ephemeral lights such as from fires and flaring are removed, and processing excludes (at pixel level) images for nights affected by clouds, moonlight, sunlight, and other glare. The most widely used stable lights product described by Baugh et al (2010) had a time-series that ended in 2013 but that has now been extended with images from F15 and F16 from 2013 onwards (Ghosh et al, 2021). The extended series observes earth in the early hours of the morning, exploiting an unstable DMSP orbit whereby what started out as early evening observation became instead a late afternoon observation as the satellites aged; while late afternoon was not useful for studying night-time lights, the 12 hour revisit time provides a pre-dawn observation time which Ghosh et al (2021) used to extend the DMSP time-series. This switch in observation time, along with other inter-satellite differences such as due to sensor degradation over time, are controlled for by using satellite and year fixed effects (along with spatial fixed effects in all estimation results).

The second set of night-time lights data is version 2.1 VIIRS nighttime lights (VNL V2.1) annual composites from 2012 to 2019. The VNL V2.1 data are produced by Elvidge et al (2021) by using monthly cloud-free radiance averages coming from the Suomi/NPP satellite. These data undergo an initial filtering to remove extraneous features such as fires and aurora before the resulting rough annual composites are subjected to further outlier removal procedures. The lit grid cells are isolated from back-ground noise using thresholds that apply across years, making these data better for change detection than earlier vintages of VIIRS data (Elvidge et al, 2017) that used year-specific thresholds (which made it harder to see if different average radiance values between years were due to on-the-ground changes or to differences in the thresholds). The data are in units of nano Watts per square centimeter per steradian (nW/cm^2/sr) presented on a 15 arc-second output grid.

The third source of NTL data is the NASA Black Marble annual composites (Román et al, 2018), which are derived from the same satellite as the VNL data but are processed in a different way. The Black Marble data use a bi-directional reflectance distributional function (BRDF) to remove the effects of extraneous artefacts, and the processing steps also remove cloud-contaminated pixels. The data products are corrected for atmospheric, terrain, vegetation, snow, lunar, and stray light effects on the radiance values, which are calibrated across time and also validated against ground measurements. The data are in units of nano Watts per square centimeter per steradian (nW/cm^2/sr) and are reported with 16-bit precision on a 15 arc-second output grid. We use the all-angle composites, which use the greatest number of nights per year, compared to either the near-nadir (view zenith angle of 0-20 degrees) composites or the off-nadir (view zenith angle of 40-60 degrees) composites which are composed from fewer nights per year. A feature of the Black Marble data is that separate estimates for snow-free and snow-covered periods are provided, noting that snow has difference reflectance properties to usual land cover. This does not matter for Indonesia, given its equatorial location, but parts of China and the United States are snow covered on more than one-tenth of cloud-free nights. For these two countries we therefore use a weighted-average of the snow-free and snow-covered composites, where the weights are the number of snow-free nights and snow-covered nights per year for each pixel. This weighted average should better capture variation across the entire year, in the same manner that annual GDP aggregates over economic activity in all of the seasons of the year. This weighted average has been shown to better predict sub-national GDP compared to just using the snow-free composites (Zhang and Gibson, 2022).

The VNL V2.1 and Black Marble annual composites are expected to be more accurate measures of true luminosity than are the DMSP annual composites. At nadir the sensor on the Suomi/NPP satellite has 45-times greater spatial precision than the DMSP sensor. This advantage in terms of spatial precision is illustrated in Figure D.1, which is reproduced from Elvidge et al (2013). The spatial precision advantage of data based on the Suomi/NPP satellite, as opposed to data coming from the DMSP platform, should be even greater away from the nadir because of the problem of an expanded ground footprint for the parts of the earth viewed at an angle. This issue, along with pixel aggregation due to limited data storage on DMSP satellites, and various geo-location errors (Tuttle et al, 2013) contribute to the well-known blurring problem in DMSP data (Abrahams et al, 2018). In contrast, there are no known blurring problems in the VNL V2.1 and Black Marble data.

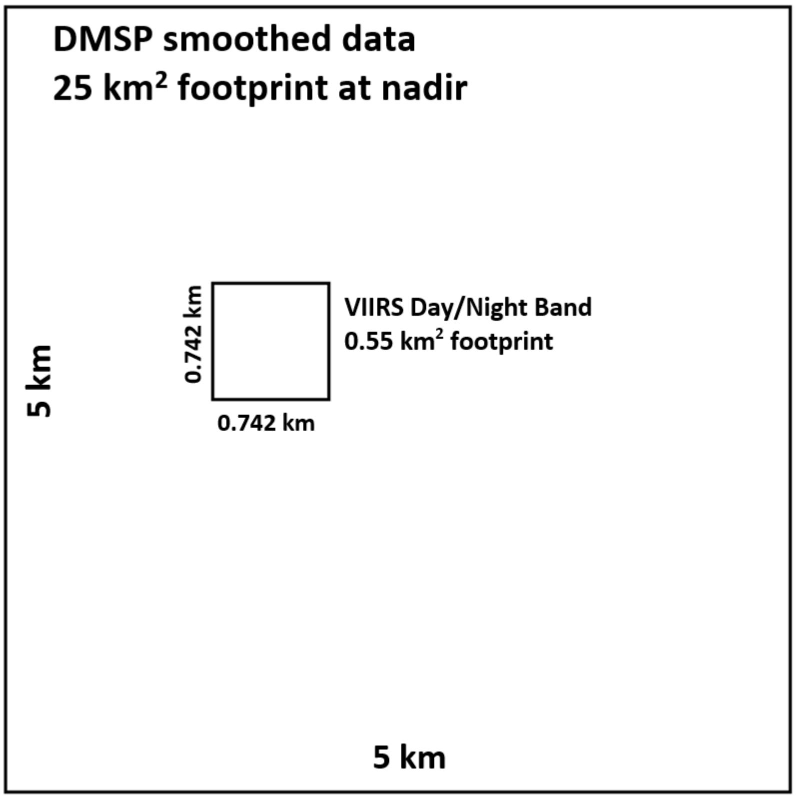


Fig D.1: Comparison of ground footprint at nadir of sensors on DMSP and Suomi/NPP satellites

In addition to greater spatial precision of data coming from the VNL V2.1 and Black Marble annual composites, there are three other reasons why these should be more accurate indicators of true luminosity than are the DMSP data. First, top-coding of DMSP data limits the apparent brightness of urban areas (Bluhm and Krause, 2022) while no similar problems affect VNL V2.1 and Black Marble data. Second, sensors on the Suomi/NPP satellites have dynamic range of seven orders of radiance magnitude while DMSP covers only two orders of magnitude, so DMSP sensors cannot simultaneously detect dimly-lit and brightly-lit areas. Hence, DMSP has relatively coarse 6-bit digital numbers, with just 64 values available (0-63) while the VNL has 14-bit precision (16,384 possible values) and Black Marble has 16-bit precision (65,536 distinct values).

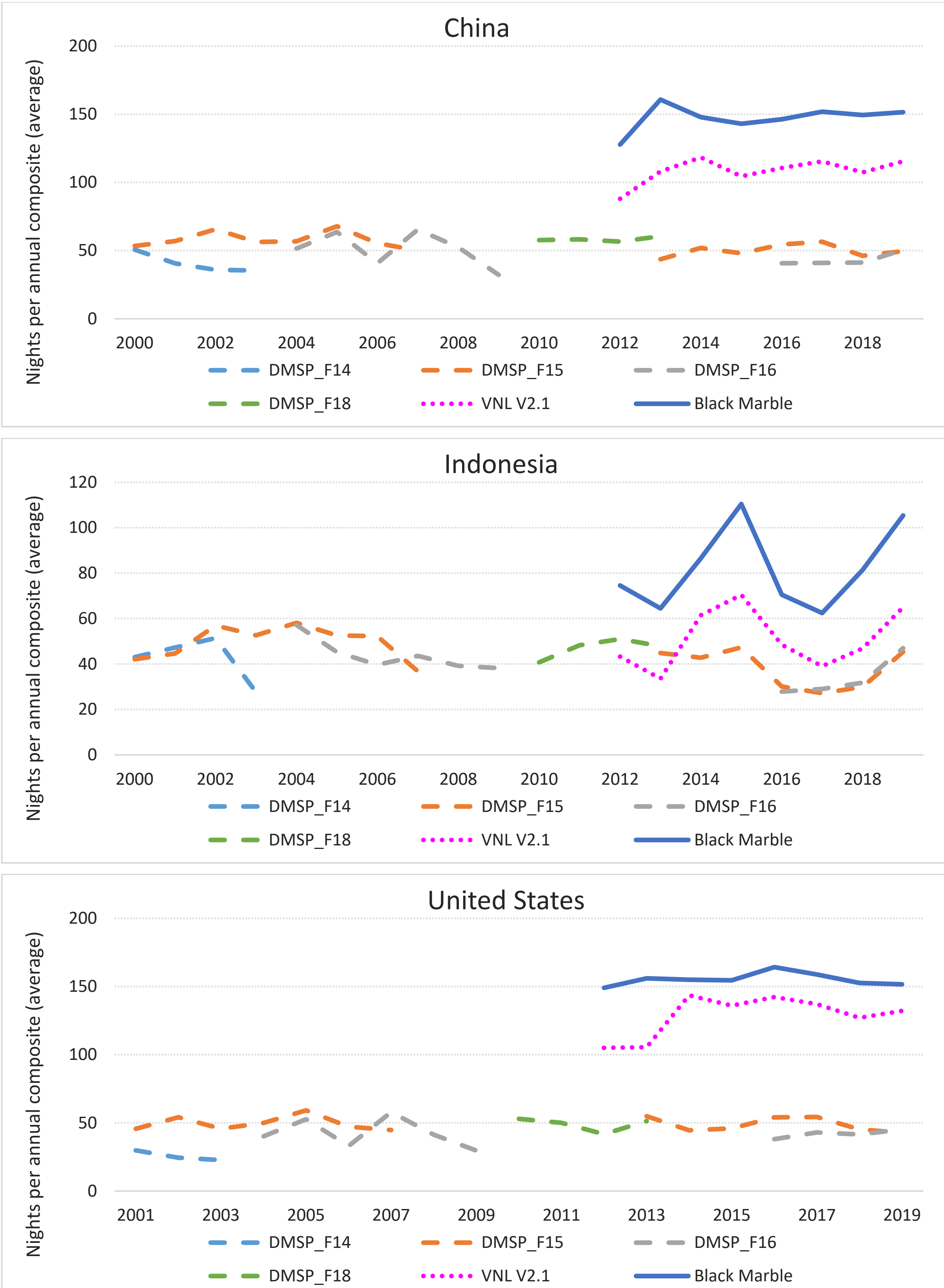


Fig D.2: Average number of nights used when composing the annual composites

The third source of accuracy for Black Marble and VNL V2.1 over DMSP as a proxy for annual economic activity comes from their incorporation of images from far more nights of the year when composing their annual composites. Figure D.2 shows the average number of nights (calculated from the pixel level data) that are used per satellite-year for the three countries we study. While the DMSP composites are based on about 50 nights per year in China, the VNL V2.1 composites are based on around 100 nights per year and the Black Marble ones on 150 nights per year. All else the same, a better representation of annual economic activity should come from a sensor that is using data from almost one-half of the nights every year, compared to one that uses just one-seventh of the nights each year.

The annual composites for Indonesia rely on fewer nights per year than for China, but the ranking of the three sources is the same. On average, the DMSP composites are based on 43 nights per year, the VNL V2.1 composites are based on 51 nights per year and the Black Marble composites use 82 nights per year. These averages hide some temporal variation, with both VNL and Black Marble composites in 2015 based on almost twice as many nights as the composites in earlier years. A possible cause of this variation is the *El Nino* weather pattern, where rain that is usually centered over Indonesia shifts eastward into the central Pacific, resulting in more cloud-free nights for parts of Indonesia in *El Nino* years such as 2015.

The difference between the night lights data sources in the number of nights per year used for their annual composites is even larger for the United States, which on average is less cloudy than either China or Indonesia. The Black Marble composites are based on an average of 155 nights per year, the VNL composites on 130 nights, and the DMSP composites are drawn from just 45 nights, on average. Across the three countries, it is clear that there is considerably better temporal coverage of annual economic activity with the Black Marble data compared to the DMSP data, with almost three-times as many nights used to form the composites (ca. 130 nights for Black Marble versus 46 nights for DMSP). The VNL composites are based on about twice as many nights as the DMSP ones so at least in terms of annual comprehensiveness, we would rank the three NTL sources as Black Marble as most comprehensive, then VNL, and DMSP last.

E. Descriptive Statistics

Table E.1: Descriptive Statistics for China

	2000-11			2012-19		
	Mean	Std Dev	Trend	Mean	Std Dev	Trend
All county-level units						
GDP	4.27	1.42	14.88**	5.58	1.28	6.45**
DMSP	9.03	1.64	3.59**	9.10	1.52	6.21**
VNL	n.a.	n.a.	n.a.	8.62	1.35	8.18**
Black Marble	n.a.	n.a.	n.a.	11.11	1.35	3.97**
Urban core counties						
GDP	5.78	1.31	15.70**	7.04	1.16	6.87**
DMSP	10.33	0.94	2.89**	10.48	0.91	4.11**
VNL	n.a.	n.a.	n.a.	10.30	1.05	6.11**
Black Marble	n.a.	n.a.	n.a.	12.77	1.04	4.86**
Non-metro counties						
GDP	4.05	1.30	14.76**	5.36	1.15	6.39**
DMSP	8.84	1.64	3.69**	8.90	1.48	6.52**
VNL	n.a.	n.a.	n.a.	8.38	1.21	8.48**
Black Marble	n.a.	n.a.	n.a.	10.86	1.21	3.84**
Provinces						
GDP	9.15	1.19	15.40**	10.45	0.99	6.69**
DMSP	13.80	0.96	2.77**	13.92	0.84	4.44**
VNL	n.a.	n.a.	n.a.	13.66	0.79	6.05**
Black Marble	n.a.	n.a.	n.a.	16.13	0.78	4.12**

Notes: All variables are in logs (inverse hyperbolic sine). The trend is the percentage annual growth rate from a regression of the logarithm on a time trend (and on satellite dummies were appropriate). The ** and * denote trends that are statistically significantly different from zero at $p<0.01$ and $p<0.05$.

Table E.2: Descriptive Statistics for Indonesia

	2004-11			2012-19		
	Mean	Std Dev	Trend	Mean	Std Dev	Trend
All district-level units						
GDP	8.59	1.26	2.50*	8.96	1.26	5.26**
DMSP	7.73	2.42	-0.22	7.85	2.26	7.50**
VNL	n.a.	n.a.	n.a.	7.39	2.12	13.86**
Black Marble	n.a.	n.a.	n.a.	10.17	1.78	5.81**
Urban cores (Kota)						
GDP	8.98	1.46	4.32	9.44	1.43	5.58*
DMSP	8.46	1.16	3.27	8.61	1.09	3.92*
VNL	n.a.	n.a.	n.a.	8.42	1.30	7.96**
Black Marble	n.a.	n.a.	n.a.	10.95	1.24	5.15**
Non-metro (Kabupaten)						
GDP	8.49	1.18	2.10	8.84	1.18	5.18**
DMSP	7.55	2.61	-0.10	7.66	2.42	8.38**
VNL	n.a.	n.a.	n.a.	7.13	2.20	15.31**
Black Marble	n.a.	n.a.	n.a.	9.97	1.84	5.97**
Provinces						
GDP	11.25	1.33	5.47	11.74	1.26	5.36
DMSP	10.19	1.46	-0.83	10.23	1.46	6.14
VNL	n.a.	n.a.	n.a.	9.82	1.49	10.78**
Black Marble	n.a.	n.a.	n.a.	12.42	1.47	5.89

Notes: All variables are in logs (inverse hyperbolic sine). The trend is the percentage annual growth rate from a regression of the logarithm on a time trend (and on satellite dummies where appropriate). The ** and * denote trends that are statistically significantly different from zero at p<0.01 and p<0.05. Sample sizes are reported in Table 2.

Table E.3: Descriptive Statistics for the United States

	2001-11			2012-19		
	Mean	Std Dev	Trend	Mean	Std Dev	Trend
All counties						
GDP	13.80	1.58	2.09**	13.93	1.57	1.35**
DMSP	10.00	1.19	-2.62**	9.76	1.24	-0.45**
VNL	n.a.	n.a.	n.a.	9.20	1.39	-0.22**
Black Marble	n.a.	n.a.	n.a.	11.59	1.45	-2.76**
Counties in urban cores						
GDP	15.43	1.83	2.25**	15.58	1.85	2.32**
DMSP	11.02	0.99	-1.23**	10.80	1.05	-0.38**
VNL	n.a.	n.a.	n.a.	10.58	1.41	-0.16**
Black Marble	n.a.	n.a.	n.a.	13.00	1.39	-1.43**
Non-metro counties						
GDP	13.54	1.36	2.07**	13.67	1.35	1.20**
DMSP	9.84	1.14	-2.84**	9.60	1.18	-0.47**
VNL	n.a.	n.a.	n.a.	8.98	1.26	-0.23**
Black Marble	n.a.	n.a.	n.a.	11.37	1.33	-2.98**
States						
GDP	18.97	1.02	2.09**	19.11	1.03	1.89**
DMSP	13.67	1.03	-1.36**	13.46	1.04	-0.14
VNL	n.a.	n.a.	n.a.	14.62	1.03	-0.23
Black Marble	n.a.	n.a.	n.a.	17.06	1.09	-0.88**

Notes: All variables are in logs (inverse hyperbolic sine). The trend is the percentage annual growth rate from a regression of the logarithm on a time trend (and on satellite dummies where appropriate). The ** and * denote trends that are statistically significantly different from zero at $p<0.01$ and $p<0.05$. Sample sizes are reported in Table 3.

F. Measurement Error Adjustments

To mitigate effects of measurement error in luminosity estimates, in case these contribute to the aggregation-sensitivity of GDP-luminosity elasticities that we find, we use Black Marble (BM) estimates as instrumental variables when estimating equation (3) in the main text:

$$z_{jt} = \psi x_{jt} + u_{jt} \tag{3}$$

where z_{jt} is observed GDP and x_{jt} is observed luminosity. The BM estimates are highly correlated with the potentially mis-measured luminosity estimates, with first-stage *F* statistics over 600 in some cases. Any error in BM estimates (noting they should have the least error, due to sensor properties and data processing, as discussed in Appendix D, and from results of Zhang and Gibson (2022) where models with BM data are closer to truth than models using other luminosity estimates, according to Vuong (1989) tests) should not correlate with errors in DMSP or VNL data. Recently, two other IV strategies have been used for potentially mismeasured luminosity data: Chor and Li (2024) and Kim et al. (2023) use a 1-year lag of annual VNL composites to instrument for VNL data, in estimates at prefectural and county level in China, while Beyer et al. (2022) use the count of nights providing the images used for the composites, in a cross-country study. Instrumenting with lags assumes there is no serial correlation in measurement errors for luminosity, but there are at least two reasons to doubt this: North-South patterns in missing data due to summer glare (Gibson, 2021) and errors from not adjusting for snow cover (Zhang and Gibson, 2022). Both of these features will tend to affect the same areas each year, inducing a serial correlation in the measurement errors. The number of nights used for the composites will also proxy for geographic factors that alter the amount of annual cloud cover (Gibson et al, 2020). These geographic factors are likely to directly affect economic activity, and rainfall fluctuations are also linked to economic growth (Barrios et al., 2010). Indeed, rainfall is a widely used instrumental variable for economic growth. These additional pathways would tend to violate the exclusion restrictions needed for the number of nights providing the images to be a valid instrumental variable where the only path of effect on economic growth is via the quality of the luminosity signal.

Table F.1 has county-level and province-level results for China from our IV approach, for the 2012-19 annual sum of lights. The estimates can be compared with OLS estimates in Table 1 of the main text. Column (1) shows the mean-reversion in the DMSP estimates; the elasticity of DMSP with respect to BM is 0.51 at the county level and 0.29 at the province level. The assumption of random errors requires an elasticity of 1.0 when the mis-measured variable is regressed on the true(r) variable: $lnx_{DMSP,jt} = \theta + \lambda lnx_{BM,jt} + v_{jt}$. This shows that a second adjustment is needed, with the IV estimate of the elasticity of GDP with respect to DMSP luminosity scaled to account for attenuating effects of the mean-reverting measurement error in DMSP data (given that IV is consistent only for random errors). Column (2) has the estimates of the instrumental variables elasticities, adjusted for mean-reversion (MRA-IV), which are 0.12 at the county level and 0.44 at the province level, giving a province-to-county ratio of 3.7 (as compared to the ratio of 8.4 estimated using OLS in Table 1 in the main text).[2]

[2] Results using DMSP data (columns (1) and (2)) control for fixed effects for year, for county or province, and for satellite (reasons for using satellite fixed effects instead of averaging data from two DMSP satellites in years with two satellites providing readings are discussed in Gibson et al., 2020. Satellite fixed effects were also used by Chen and Nordhaus (2011) in their first study using luminosity to proxy for economic activity statistics). So

Table F.1: IV estimates of GDP-luminosity elasticities for China, 2012-2019

Relationship estimated:	DMSP-BM Mean-reversion test (1)	GDP-DMSP MRA-IV (2)	GDP-VNL IV (3)
		County-level	
ln(sum of lights)	0.507	0.117	0.173
	(0.035)	(0.022)	(0.017)
Number of observations	30446	30446	18736
First-stage *F* statistic	n.a.	207.51	643.23
R-squared (Within)	0.076	0.028	0.028
		Province-level	
ln(sum of lights)	0.291	0.437	0.549
	(0.213)	(0.969)	(0.356)
Number of observations	403	403	248
First-stage *F* statistic	n.a.	1.87	10.47
R-squared (Within)	0.038	0.079	0.079
Province-to-county elasticity ratio		3.7	3.2

Notes: Clustered standard errors in ( ). MRA-IV is Mean-Reverting-Adjusted IV.

The results in column (3) use VNL data as the potentially mis-measured luminosity variable, with the BM estimates as the instrumental variable. There is no known mean-reversion issue with VNL estimates so an adjustment for mean-reverting error is not carried out and it is just a standard instrumental variables estimate in column (3). The GDP-luminosity elasticities are 0.17 at county level and 0.55 at province level, which then gives a province-to-county ratio of 3.2 (compared to the ratio of 3.6 estimated using OLS in Table 1 in the main text). Thus, the pattern of aggregation-sensitivity of the GDP-luminosity elasticities persists, even if using the less error-prone VNL estimates and an instrumental variables framework that should further mitigate any measurement error concerns. Notably, the fact that switching the estimator from OLS to IV only slightly changes the province-to-county elasticity ratio (from 3.6 to 3.2) when using VNL estimates but had a larger effect when using DMSP estimates (from 8.4 to 3.7) is indirect evidence that the VNL estimates are less plagued by measurement error than are the DMSP estimates.

Table F.2 has the county-level and state-level results for the United States, which can be compared with the OLS results in Table 3 of the main text. The degree of mean-reversion in the DMSP estimates is seen from the elasticity of DMSP with respect to BM being 0.52 at the county level and 0.66 at the state level. Using DMSP data with the MRA-IV estimator yields GDP-luminosity elasticities of 0.20 at the county level and 0.05 at the state level, so just as with the OLS results in Table 3, spatial aggregation lowers the GDP-luminosity elasticity estimates for the United States. When the VNL data are used, the results in column (3) show that the GDP-luminosity elasticity is 0.44 at the county level and 0.07 at the state level and so the state-to-county ratio is 0.2 when using measurement error mitigation approaches applied to either the DMSP estimates or the VNL estimates. So our main results, of a downward bias when there is spatial aggregation with US data and an upward bias for China, still hold.

the unit of observation is the satellite-area-year triplet. To not overweight years with two satellites providing data, observations are given a weight of 0.5 (while they are weighted 1.0 for the years with only one satellite in operation). Satellite fixed effects are not needed for VIIRS, so there are fewer observations in column (3), where area-year doublets are used rather than the satellite-area-year triplets in columns (1) and (2).

Table F.2: IV estimates of GDP-luminosity elasticities for the United States, 2012-2019

Relationship estimated:	DMSP-BM Mean-reversion test (1)	GDP-DMSP MRA-IV (2)	GDP-VNL IV (3)
		County-level	
ln(sum of lights)	0.517	0.202	0.442
	(0.036)	(0.050)	(0.057)
Number of observations	40417	40417	24872
First-stage *F* statistic	n.a.	202.34	151.98
R-squared (Within)	0.089	0.070	0.070
		State-level	
ln(sum of lights)	0.656	0.045	0.068
	(0.144)	(0.116)	(0.115)
Number of observations	663	663	408
First-stage *F* statistic	n.a.	20.71	14.43
R-squared (Within)	0.121	0.006	0.006
Province-to-county elasticity ratio		0.2	0.2

Notes: Clustered standard errors in ( ). MRA-IV is Mean-Reverting-Adjusted IV.

Finally, we show that even with the IV strategy of Kim et al. (2023) and Chor and Li (2024), where a 1-year lag of VNL luminosity is used as the instrument, our main findings regarding the aggregation-sensitivity of GDP-luminosity estimates still hold. The results in Table F.3 show that the province-to-county ratio of elasticities is 4.0 for China, while the corresponding ratio of state-to-county elasticities is 0.4 for the US. Even though we argue that using lagged luminosity as the instrumental variable is less valid than our strategy of using the accurate BM luminosity estimates as the instrument, for mitigating effects of any measurement errors in luminosity estimates, spatial aggregation still has the same effect (which is also the same effect as in the OLS results)—inflating the GDP-luminosity elasticity for China and reducing it for the United States. Hence, aggregation-sensitivity of the GDP-luminosity elasticities is unlikely to be due to the effects of any measurement error in luminosity estimates.

Table F.3: Using lagged VNL annual composites as the instrumental variable for estimating GDP-luminosity elasticities for China and the United States, 2012-2019

	China	United States
	County-level	County-level
ln(sum of lights)	0.283	0.298
	(0.017)	(0.055)
Number of observations	16394	21763
First-stage *F* statistic	4046.30	457.50
R-squared (Within)	0.061	0.040
	Province-level	State-level
ln(sum of lights)	1.138	0.128
	(0.202)	(0.066)
Number of observations	207	357
First-stage *F* statistic	20.71	40.74
R-squared (Within)	0.297	0.023
Province (or State)-to-county elasticity ratio	4.0	0.4

Notes: Clustered standard errors in ( ).

References for studies cited in Appendices B to F

Abrahams, A., Oram, C., & Lozano-Gracia, N. (2018). Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination. *Remote Sensing of Environment*, *210*, 242-258.

Barrios, S., Bertinelli, L., & Strobl, E. (2010). Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy. *The Review of Economics and Statistics*, *92*, 350-366.

Baugh, K., Elvidge, C., Ghosh, T., & Ziskin, D. (2010). Development of a 2009 stable lights product using DMSP-OLS data. *Proceedings Asia-Pacific Advanced Networks 30*, 114.

Beyer, R., Hu, Y., & Yao, J. (2022). Measuring quarterly economic growth from outer space. *Policy Research Working Paper* No. 9893. The World Bank.

Bluhm, R., & Krause, M. (2022). Top Lights: Bright cities and their contribution to economic development. *Journal of Development Economics*, *157*, 102880.

Chen, X., & Nordhaus, W. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, *108*, 8589-8594.

Chor, D., & Li, B. (2024). Illuminating the effects of the US-China tariff war on China's economy. *Journal of International Economics, 150*, 103926.

Chung, J., & Lam, T. (2004). China's "city system" in flux: explaining post-Mao administrative changes. *The China Quarterly*, *180*, 945-964.

Elvidge, C., Baugh, K., Zhizhin, M., & Hsu, F-C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, *35*, 62.

Elvidge, C., Baugh, K., Zhizhin, M., Hsu, F-C., & Ghosh, T. (2017). VIIRS night-time lights. *International Journal of Remote Sensing*, *38*(21), 5860-5879.

Elvidge, C., Zhizhin, M., Ghosh, T., Hsu, F.-C., & Taneja, J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing 13*, 922.

Ghosh, T., Baugh, K., Elvidge, C., Zhizhin, M., Poyda, A., & Hsu, F-C. (2021) Extending the DMSP nighttime lights time series beyond 2013. *Remote Sensing, 13*, 5004.

Gibson, J., Olivia, S., Boe-Gibson, G., & Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, *149*, 102602.

Henderson, V., Storeygard, A., & Weil, D. (2012). Measuring economic growth from outer space. *American Economic Review*, 1*02*, 994-1028.

Kim, J., Kim, K., Park, S., & Sun, C. (2023). The economic costs of trade sanctions: Evidence from North Korea. *Journal of International Economics*, *145*, 103813.

National Bureau of Statistics [NBS]. China Statistical Yearbook (county-level)[M]. Beijing: China Statistics Press, 2001-2020.

National Bureau of Statistics [NBS]. China City Statistical Yearbook[M]. Beijing: China Statistics Press, 2001-2020.

National Bureau of Statistics [NBS]. Beijing Statistical Yearbook[M]. Beijing: China Statistics Press, 2001-2020.

Román, M., Wang, Z., Sun, Q., Kalb, V., Miller, S., Molthan, A., Schultz, L., Bell, J., Stokes, E., Pandey, B., & Seto, K. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of the Environment*, *210*, 113-43.

Tuttle, B., Anderson, S., Sutton, P., Elvidge, C., & Baugh, K. (2013). It used to be dark here. *Photogrammetric Engineering & Remote Sensing*, *79*(3), 287-297.

Zhang, X., & Gibson, J. (2022). Using multi-source nighttime lights data to proxy for county-level economic activity in China from 2012 to 2019. *Remote Sensing*, *14*(5), 1282.