Supporting Information

Xi Chen and William Nordhaus, Using luminosity data as a proxy for economic statistics

March 26, 2011

Contents

I. Data Sources

Nighttime luminosity

The derivation of nighttime luminosity for cells and countries is described in the text. The statistics on the data by luminosity and year are shown in Table SI-1.

Gross cell product

The Yale G-Econ dataset is organized around geophysical boundaries. The original data set, published in 2006, was available for a single year (1990). We have updated the observations, revised several data and methodological issues, and extended it to three periods (1995, 2000, and 2005). The data and descriptions are available at gecon.yale.edu.

The Yale G-Econ project provides data on gridded output, *gross cell product,* or GCP. In this work, the "cell" is the surface bounded by 1-degree latitude by 1-degree longitude contours. The globe contains 64,800 such grid cells; we have data on 27,442 observations, of which virtually all have reasonably complete data on climate, population, and output. The grid cell is the selected geographic unit because it is the high-resolution geophysical system for which population data are available. It is also the most convenient boundary system for integrating with global environmental data. Additionally, it has the features that the coordinate system is (to a first approximation) statistically independent of economic data (which obviously is not the case for political boundaries) and that the elements are (except at high latitudes) of nearly uniform size.

The conceptual basis of GCP is the same as that of gross domestic product (GDP) and gross regional product as developed in the national income and product accounts of major countries, except that the geographic unit is the latitude-longitude grid cell. Gross cell product is gross value added in a specific geographic region; gross value

added is equal to total production of market goods and services in a region less purchases from businesses. GCP aggregates across all cells in a country to equal gross domestic product. We measure output in purchasing-power-corrected 2005 international U.S. dollars using national aggregates estimated by the World Bank.

II. The Analytical Background

This section derives the equations in the text in detail. We are grateful to Zhipeng Liao for suggestions for improvement in the presentation in this section.

Constructing "true" GDP

For this purpose, we define the different measures as follows:

Y = measured output (GDP in constant 2005 international U.S. $)
Y^* = true output (GDP in constant 2005 international U.S. $)
X = synthetic measure of output (GDP in constant 2005 international U.S. $)
M = measured luminosity (index value)
Z = luminosity-based measure of output (GDP in constant 2005 international U.S. $)
i = grid cell (here 1° latitude by 1° longitude)
j = country
k= country grade (A, B, C, D, E)
t = year
y = log (Y)
x = log (X)
m = log (M)
z = log (Z)
ε_i = measurement error in GDP
ξ_i = measurement error in luminosity
u_i = error in output-luminosity relationship
α, β, μ = structural parameters

For notational purposes, we define $x_i(t)$ as the value of variable x in grid cell i averaged over a year. We omit the time variables as inessential to the exposition. Further assume that there is an unknown true level of output for each country and grid cell, which is measured with error.

$$y_i = (1 - \mu) E\left(y_i^*\right) + \mu y_i^* + \varepsilon_i$$

For the present study, we assume that there is no bias in measured output, so $\mu = 1$. This assumption is not completely innocuous as there may be systematic growth mismeasurement due, say, to incomplete source data or infrequent observations, and the question will be pursued in a future study. Assuming that $\mu = 1$ yields:

(SI-1) $y_i = y_i^* + \varepsilon_i$

Luminosity is subject to measurement error (due to satellite, calibration, and other sources):

(SI-2) $m_i = m_i^* + \xi_i$

There is a structural relationship between luminosity and true output as follows:

(SI-3) $m_i = \alpha + \beta y_i^* + u_i$

The error in Eq. SI-3 arises from several sources. One important reason why the relationship is noisy is that luminosity is sampled at night, whereas economic activity is generally concentrated in the daytime. More important, the light intensity differs greatly across sectors. Often, lights are associated with electricity use. The use of electricity per dollar of output in different sectors provides a rough idea of how light-intensities might vary. In the 2002 U.S. input-output tables, the electricity used per unit output of real estate was 200 times greater than that of software. (See the input-output tables at www.bea.gov for the underlying data.) Similar differences are seen across other sectors. This example suggests that industrial composition across countries and regions is likely to make the output-luminosity relationship in Eq. SI-3 relatively noisy.

We want to construct a luminosity-output proxy from these relationships. We have measurements of all variables over time and space at the national and grid cell levels. However, we need to develop measures of the error of measurement of national and grid cell output, the coefficient on luminosity, as well as the error in the structural relationship in Eq. SI-3.

Our procedure is first to estimate equation Eq. SI-3 using measured output and luminosity. This provides a biased estimate of the coefficient, $\hat{\beta}$, because output is measured with error. We then do an errors-in-variable correction using our prior estimates of the error in GDP to get a corrected estimate of the structural coefficient, $\tilde{\beta}$. The corrected coefficient is calculated as:

(SI-4) $\tilde{\beta} = \left(\dfrac{\sigma_{y^*}^2 + \sigma_{\varepsilon}^2}{\sigma_{y^*}^2} \right) \hat{\beta}$

Here, $\hat{\beta}$ is the estimated coefficient in equation Eq. SI-3; σ_{ε}^2 is the a priori error variance of true output; and $\sigma_{y^*}^2$ is the estimated variance of true output. The consistent estimate of $\tilde{\alpha}$ follows immediately.

We then estimate the luminosity-output proxy as follows by inverting Eq. SI-3:

(SI-5) $\quad \hat{z}_i = (1/\tilde{\beta})(m_i - \tilde{\alpha}_i)$

where $\hat{z}_i$ is the log of our luminosity-output proxy and $\tilde{\beta}$ and $\tilde{\alpha}$ are the corrected coefficients from equation (SI-4).

Next, we construct a synthetic or combined measure of output by taking weighted averages of conventional measures of output and our luminosity-output proxy:

(SI-6) $\quad \hat{x}_i = (1-\theta)y_i + \theta\hat{z}_i$

where

$\hat{X}_i$ = new synthetic measure of output
$\hat{x}_i = \ln(\hat{X}_i)$
θ = weighting fraction on luminosity.

The key variable of this study is θ, which is the share (or weight) of the luminosity-output proxy. The central question we address is whether we can significantly improve conventional measures of output using luminosity. If the measurement error of the luminosity-output proxy is low relative to the measurement error of conventional output estimates, then luminosity can be a useful proxy.

Define $V(\theta)$ as the mean squared error (MSE) of $\hat{x}_i$ as a function of the weight, θ. We proceed intuitively by first assuming that all parameters are known. In this case, we can derive $V(\theta)$ as a function of θ as follows (we omit the grade superscript in the derivation until the next section):

$$V(\theta) = E\left[(1-\theta)y + \theta\hat{z} - y^*\right]^2$$

$$= E\left[(1-\theta)(y^* + \varepsilon) + \frac{\theta}{\beta}(\beta y^* + u) - y^*\right]^2$$

$$= E\left[(1-\theta)\varepsilon + \frac{\theta}{\beta}u\right]^2$$

$$= (1-\theta)^2\sigma_\varepsilon^2 + \frac{\theta^2}{\beta^2}\sigma_u^2$$

If all the parameters are known, then minimizing $V(\theta)$ with respect to θ yields a unique value for the optimal weight, θ^*:

$$V'(\theta^*) = 0 = -2(1-\theta^*)\sigma_\varepsilon^2 + \frac{2\theta^*}{\beta^2}\sigma_u^2$$

or

(SI-7) $\qquad \theta^* = \dfrac{\beta^2 \sigma_\varepsilon^2}{\beta^2 \sigma_\varepsilon^2 + \sigma_u^2}$

However, because the parameters in $(\mathrm{SI}-7)$ are unknown, we need to find an appropriate estimator of θ. We assume σ_ε^2 is known from external evidence (see below). Further, assume that that σ_u^2 and β can be consistently estimated as $\tilde{\sigma}_u^2$ and $\tilde{\beta}$, respectively. It can be shown that θ^* is the uniformly consistent estimator of θ.

A sketch of the proof is the following. Define $\tilde{\theta}_n^*$ as the estimator of the optimal weight θ^* for sample size n:

(SI-8) $\qquad \tilde{\theta}_n^* = \dfrac{\tilde{\beta}^2 \sigma_\varepsilon^2}{\tilde{\beta}^2 \sigma_\varepsilon^2 + \tilde{\sigma}_u^2}$.

Because $\tilde{\beta}$ and $\tilde{\sigma}_u^2$ are consistent estimators, $\tilde{\theta}_n^* \to \theta^*$ in the probability limit. So the estimator in (SI-8) is a consistent estimator of $\tilde{\theta}_n^*$ which minimizes the asymptotic MSE of the synthetic output measure. Note that $\tilde{\theta}_n^*$ is not necessarily an unbiased estimator in small samples. Further work using Monte Carlo and bootstrap techniques will be necessary to determine the properties of $\tilde{\theta}_n^*$ for the actual sample.

Differences in reliability by country

There are several modifications of equation Eq. SI-8 that can be pursued, but we use only one in the present study. We know both from direct observation and from studies that the reliabilities of national statistical systems vary greatly. We therefore will be particularly concerned about the applicability of our estimates to different countries. For this purpose, we will divide countries into different "grades," k = A, B, C, D, and E. This leads to the modification of Eq. SI-8 in which the optimal weights depend on the error variance estimates of output by country grade:

(SI-9) $\qquad \tilde{\theta}_n^{k*} = \dfrac{\tilde{\beta}^2 (\sigma_\varepsilon^k)^2}{\tilde{\beta}^2 (\sigma_\varepsilon^k)^2 + \tilde{\sigma}_u^2}$

where $(\sigma_\varepsilon^k)^2$ is the variance of the measurement error of output for country grade k.

III. Estimated Parameters

Equation Eq. SI-9 contains three parameters for countries in each grade.

Errors in luminosity equation $\left(\tilde{\sigma}_u^2\right)$ *and coefficient* $\left(\tilde{\beta}\right)$

These two coefficients are derived primarily from estimates of the luminosity equation. We show the equation for the cross-section of grid cells for all countries, high-density countries, low-density countries, and countries for different grades in Table SI-5. The analogous equations are made for time series as well as for countries.

Errors of measurement for conventional output $(\sigma_\varepsilon^k)^2$

The country classification is based on the Heston-Summers and related studies (2-4) with some adjustments by the authors. The list of countries by grade is shown in Table SI-4. The errors of measurement for individual countries and cells are available in a background document (1). There is substantial work on errors in high-income countries, but very little on developing countries. Future work will provide bootstrap and Monte Carlo estimates of the optimal weights as a function of the errors for both the coefficients and the measurement errors. At present, we do know that the errors arising from the coefficient estimates are very low because they are very precisely determined (see Table SI-5).

IV. Intercalibrated Lights

As discussed in the main text, there are multiple versions of nighttime lights. We relied primarily on stable lights, but also analyzed raw lights for all years and calibrated lights for 2006. At present, there are no publicly available data on stable lights that are intercalibrated across satellites and time. However, there is a preliminary set of intercalibrations for the "avg_lights_x_pct" product (5). The "avg_lights_x_pct" product is derived from raw lights, not stable lights.

To test for a possible difference in results, we examined the published study on intercalibration (5). These corrections estimate quadratic functions of the original series for each year and satellite. One problem with this approach is that several regressions contain negative intercepts. This leads to negative values for nighttime lights in cells with zero lights in the original data. In some years, indeed, we found that about 60% of grid cells with intercalibrated lights have a negative value for intercalibrated lights.

As a final check, we did some preliminary analysis for the data with positive constant terms in the intercalibrating equation (and therefore no negative lights). The coefficients are generally less precise than for stable lights, but the basic results for the weights are very close to our preferred estimates.

V. Lower Bound Reliability for the Luminosity-output Proxy

The lower bound for the reliability of the luminosity-output proxy can be estimated as follows. Begin with the definition of the luminosity-output proxy from Eqs. (SI-3) and (SI-5) for a consistent estimator, where we reintroduce necessary notation for grid cell i, satellite k, and year t:

$$\hat{z}_{i,k,t} = (1/\tilde{\beta})(m_{i,k,t} - \tilde{\alpha}_{i,k,t})$$

If we take the difference between the estimate of the luminosity-output proxy for the same grid cell and year and different satellites k and $k+1$, we obtain the difference in observations of the proxy as:

$$\hat{z}_{i,k,t} - \hat{z}_{i,k+1,t} = (1/\tilde{\beta})(u_{i,k,t} - u_{i,k+1,t})$$

Assuming that the measurement errors of nighttime lights are identical and independently distributed with variance σ_u^2 and that $\tilde{\beta}$ is 1, this implies that the variance across satellites for grid cells is

(SI-10) $E[(\hat{z}_{i,k,t} - \hat{z}_{i,k+1,t})^2] = 2\sigma_u^2.$

The left-hand side of (SI-10) is the squared measurement error across satellites, while the right-hand side is two times the squared measurement error of nighttime lights. It is easily verified that the SERs shown in Tables SI-2a and SI-2b are higher than σ_u^2 in Eq. SI-10. This indicates that the lower bound on measurement errors of the output-luminosity proxy based on a single year-satellite observation of lights are approximately the values shown in those tables divided by √2. Similar reasoning shows that if the growth rate of output is small from one year to the next, then the same estimate applies for the same satellite in different years.

Applying Eq. SI-10 to the regression results in Tables SI-2a and SI-2b, we can take the average of the 12 SER estimates where there are multiple satellites for a single year (for example, the estimate of the SER for satellites F12 and F14 for 1997). For the 1° x 1° grid cells the estimated lower bound measurement error is 26 logarithmic percent, while for the 0.5° x 0.5° scale the estimated lower bound measurement error is 28 logarithmic percent.

VI. Tables and Figures

year	f10	f12	f14	f15	f16
1992	12,124	.	.	.	.
	36,952	.	.	.	.
1993	12,776	.	.	.	.
	36,971	.	.	.	.
1994	12,818	13,034	.	.	.
	37,718	38,484	.	.	.
1995	.	14,829	.	.	.
	.	41,408	.	.	.
1996	.	14,556	.	.	.
	.	40,562	.	.	.
1997	.	15,040	11,748	.	.
	.	41,847	34,734	.	.
1998	.	16,225	12,738	.	.
	.	43,781	36,026	.	.
1999	.	15,767	12,984	.	.
	.	44,733	37,162	.	.
2000	.	.	13,924	16,859	.
	.	.	39,593	44,568	.
2001	.	.	13,597	16,424	.
	.	.	38,908	43,518	.
2002	.	.	13,163	16,920	.
	.	.	38,969	44,354	.
2003	.	.	13,874	12,594	.
	.	.	39,296	35,531	.
2004	.	.	.	12,847	14,953
	.	.	.	35,707	40,076
2005	.	.	.	13,569	13,501
	.	.	.	37,307	36,870
2006	.	.	.	13,479	14,421
	.	.	.	36,961	39,523
2007	.	.	.	13,520	16,873
	.	.	.	37,060	44,047
2008	.	.	.	.	16,434
	.	.	.	.	43,083

Table SI-1. Statistics of stable-lights data by satellite and year

Table provides statistics for grid-cell data on satellites for 1° x 1° grid cells for stable-lights luminosity after merging with G-Econ3.4. The numbers in the column are the mean and standard deviation of the cell totals. The sample size is 19,815 for each satellite year.

Standard errors of regressions	f10 (t+1 on t)	f12 on f10	f12 (t+1 on t)	f14 on f12	f14 (t+1 on t)	f15 on f14	f15 (t+1 on t)	f16 on f15	f16 (t+1 on t)
1992	0.494								
1993	0.429								
1994		0.586	0.594						
1995			0.373						
1996			0.427						
1997			0.412	0.358	0.471				
1998			0.414	0.326	0.399				
1999				0.334	0.406				
2000					0.404	0.345	0.355		
2001					0.420	0.378	0.382		
2002					0.435	0.396	0.389		
2003						0.369	0.386		
2004							0.360	0.315	0.389
2005							0.366	0.284	0.404
2006							0.400	0.345	0.410
2007								0.318	0.345

Table SI-2a. Standard errors of regressions of log 1° x 1° grid cell luminosity over satellites or years for stable lights

Table shows the standard errors of estimates of regressions (SER) of the logarithm of luminosity for grid cells across satellites or across time for the stable lights series used in the main analysis. The unshaded columns show the SER for regressions of luminosity in t and t+1, while the shaded columns show the SER for luminosity of satellite k on satellite k+1 for year t.

Standard errors of	f10 (t+1 on t)	f12 on f10	f12 (t+1 on t)	f14 on f12	f14 (t+1 on t)	f15 on f14	f15 (t+1 on t)	f16 on f15	f16 (t+1 on t)
1992	0.557								
1993	0.465								
1994		0.563	0.583						
1995			0.416						
1996			0.479						
1997			0.449	0.374	0.501				
1998			0.454	0.358	0.442				
1999				0.372	0.437				
2000					0.428	0.373	0.388		
2001					0.459	0.379	0.418		
2002					0.470	0.431	0.426		
2003						0.403	0.422		
2004							0.416	0.376	0.445
2005							0.414	0.346	0.454
2006							0.451	0.399	0.465
2007								0.354	0.398

Table SI-2b. Standard errors of regressions of log 0.5° x 0.5° grid cell luminosity over satellites or years for stable lights

Similar to Table SI-2a, this table shows the standard errors of estimates of regressions (SER) of the logarithm of luminosity for 0.5° x 0.5° grid cells across satellites or across time. Compared to the numbers in Table SI-2a, the SER for 0.5° x 0.5° grid cells are consistently larger except for year 1994.

Correlation coefficent	f10	f12	f14	f15	f16
1992	0.800				
1993	0.800				
1994	0.797	0.802			
1995		0.803			
1996		0.802			
1997		0.805	0.822		
1998		0.808	0.821		
1999		0.823	0.825		
2000			0.824	0.806	
2001			0.831	0.816	
2002			0.842	0.803	
2003			0.819	0.822	
2004				0.801	0.815
2005				0.798	0.805
2006				0.800	0.807
2007				0.791	0.805
2008					0.799

Table SI-3. Correlation coefficients of luminosity and GDP

Table shows the correlation coefficients of log GDP and log luminosity for grid cells by year for stable lights. Note that approximately 22 percent of grid cells with positive output have stable lights = 0 and are therefore excluded.

Country	Grade	Country	Grade	Country	Grade
Afghanistan	E	Central African	D	Gibraltar	B
Albania	C	Chad	E	Greece	B
Algeria	D	Chile	B	Greenland	B
American Samoa	C	China	C	Grenada	C
Andorra	C	Christmas Island	D	Guadeloupe	C
Angola	E	Cocos Island	D	Guam	C
Anguilla	C	Colombia	C	Guatemala	C
Antarctica	D	Comoros	D	Guinea	D
Antigua and Barbuda	C	Congo	E	Guinea Bissau	C
Argentina	B	Cook Island	D	Guyana	D
Armenia	C	Costa Rica	C	Haiti	D
Aruba	C	Cote d'Ivoire	C	Heard Island and	D
Australia	A	Croatia	C	Honduras	C
Austria	A	Cuba	D	Hong Kong	C
Azerbaijan	C	Cyprus	D	Hungary	C
Bahamas	C	Democratic Republic of Congo	D	Iceland	C
Bahrain	C	Czech Republic	C	India	C
baker and Howland	D	Denmark	A	Indonesia	C
Bangladesh	C	Djibouti	E	Iran	C
Barbados	C	Dominica	C	Iraq	E
Belarus	D	Dominican Republic	C	Ireland	A
Belgium	A	Ecuador	C	Isle of Man	C
Belize	C	Egypt	C	Israel	B
Benin	C	El Salvador	C	Italy	A
Bermuda	C	Equatorial Guine	C	Jamaica	C
Bhutan	D	Eritrea	E	Jan Mayen	D
Bolivia	C	Estonia	C	Japan	A
Bosnia and Herzegovinia	C	Ethiopia	C	Jarvis Island	D
Botswana	C	Falkland Island	C	Johnston Atoll	D
Bouvet Island	D	Faroe Island	C	Jordan	C
Brazil	C	Federated States of Micronesia	D	Kazakhstan	C
British Indian Ocean Territory	D	Fiji	C	Kenya	C
British Virgin Islands	C	Finland	B	Kiribati	D
Brunei	C	France	A	Kuwait	C
Bulgaria	C	French Guiana	C	Kyrgyztan	C
Burkina Faso	C	French Polynesia	C	Laos	D
Burundi	C	French Southern Territories	D	Latvia	C
Cambodia	D	Gabon	C	Lebanon	C
Cameroon	C	Gambia	C	Lesotho	D
Canada	A	Georgia	C	Liberia	D
Cape Verde	C	Germany	B	Libya	D
Cayman Island	C	Ghana	C		

Country	Grade	Country	Grade	Country	Grade
Liechtenstein	B	Northern Mariana	D	St. Pierre and Miquelon	C
Lithuania	C	Norway	A	Saint Vincent and the Grenadines	C
Luxembourg	A	Oman	C	Sudan	E
Macau	C	Pakistan	C	Suriname	D
Macedonia	C	Palau	D	Svalbard	C
Madagascar	C	Panama	C	Swaziland	C
Malawi	C	Papua New Guinea	D	Sweden	A
Malaysia	C	Paraguay	C	Switzerland	A
Maldives	C	Peru	C	Syria	C
Mali	C	Philippines	C	Taiwan	C
Malta	D	Pitcairn Islands	D	Tajikistan	D
Marshall Is.	D	Poland	B	Tanzania	C
Marshall Islands	D	Portugal	B	Thailand	C
Martinique	C	Puerto Rico	B	Timor Leste	E
Mauritania	C	Qatar	C	Togo	D
Mauritius	C	Reunion	C	Tokelau	D
Mayotte	D	Romania	C	Tonga	D
Mexico	C	Russia	C	Trinidad and Tobago	C
Micronesia	C	Rwanda	C	Tunisia	C
Midway Is.	C	Saint Helena	C	Turkey	C
Moldova	C	Samoa	C	Turkmenistan	D
Monaco	B	San Marino	C	Turks and Caicos Islands	C
Mongolia	D	Sao Tome and Principe	C	Tuvalu	D
Montserrat	D	Saudi Arabia	D	Uganda	E
Morocco	C	Senegal	C	Ukraine	C
Mozambique	D	Serbia and Montenegro	C	United Arab Emirates	C
Myanmar	E	Seychelles	C	United Kingdom	A
Namibia	D	Sierra Leone	C	United States	A
Nauru	D	Singapore	B	Uruguay	B
Nepal	C	Slovakia	C	Uzbekistan	D
Netherland Antil	C	Slovenia	C	Vanuatu	D
Netherlands	A	Solomon Islands	C	Vatican City	D
Netherlands Anti	C	Somalia	E	Venezuela	C
New Caledonia	C	South Africa	C	Vietnam	C
New Zealand	B	South Georgia and South Sandwich Islands	E	Virgin Islands	C
Nicaragua	C	South Korea	B	Wake Island	C
Niger	E	Spain	B	Wallis and Futuna	D
Nigeria	C	Sri Lanka	C	West Bank and Gaza	E
Niue	D	St. Helena	C	Yemen	E
Norfolk Island	D	St. Kitts and Nevis	C	Zambia	C
North Korea	E	St. Lucia	C	Zimbabwe	D

Table SI-4. Country grades

	All countries	A grade countries	B grade countries	C grade countries	D grade countries	E grade countries	Countries with log(GDP	Countries with log(GDP
log (gdp density)	0.872	0.765	0.914	0.893	0.913	0.739	0.977	0.289
	(763.12)	(507.62)	(298.84)	(511.07)	(176.60)	(66.59)	(641.35)	(46.03)
Year 1993	0.132	0.0952	0.235	0.165	0.0251	-0.0816	0.146	0.0621
	(6.34)	(2.76)	(4.11)	(6.08)	(0.31)	(-0.48)	(6.97)	(1.04)
Year 1994	0.102	0.0077	0.22	0.179	-0.0208	-0.109	0.122	-0.000585
	(4.86)	(0.22)	(3.85)	(6.56)	(-0.25)	(-0.65)	(5.81)	(-0.01)
Year 1995	0.257	0.207	0.508	0.327	0.0907	-0.273	0.296	0.0589
	(8.70)	(4.23)	(6.33)	(8.51)	(0.78)	(-1.16)	(10.00)	(0.71)
Year 1996	0.215	0.172	0.42	0.292	0.0194	-0.313	0.256	-0.026
	(7.30)	(3.53)	(5.24)	(7.61)	(0.17)	(-1.34)	(8.67)	(-0.31)
Year 1997	0.188	0.104	0.379	0.275	0.0486	-0.342	0.225	-0.0128
	(6.65)	(2.22)	(4.92)	(7.46)	(0.44)	(-1.52)	(7.96)	(-0.16)
Year 1998	0.267	0.144	0.513	0.384	0.0612	-0.31	0.313	0.0412
	(9.44)	(3.08)	(6.67)	(10.44)	(0.55)	(-1.38)	(11.06)	(0.51)
Year 1999	0.202	0.0914	0.473	0.29	0.16	-0.204	0.253	-0.0327
	(7.14)	(1.94)	(6.14)	(7.86)	(1.44)	(-0.91)	(8.93)	(-0.41)
Year 2000	0.324	0.212	0.557	0.426	0.302	0.116	0.365	0.127
	(10.63)	(4.17)	(6.71)	(10.71)	(2.55)	(0.49)	(11.93)	(1.47)
Year 2001	0.241	0.108	0.526	0.346	0.241	-0.179	0.295	-0.0175
	(7.87)	(2.13)	(6.33)	(8.67)	(2.04)	(-0.75)	(9.62)	(-0.20)
Year 2002	0.243	0.102	0.583	0.326	0.292	-0.0845	0.275	0.0477
	(7.94)	(1.99)	(7.00)	(8.15)	(2.46)	(-0.35)	(8.95)	(0.55)
Year 2003	0.022	-0.155	0.405	0.14	0.0288	-0.413	0.0733	-0.198
	(0.72)	(-3.06)	(4.88)	(3.52)	(0.24)	(-1.73)	(2.39)	(-2.28)
Year 2004	-0.0956	-0.225	0.238	0.0251	-0.196	-0.563	-0.0678	-0.26
	(-2.97)	(-4.21)	(2.72)	(0.60)	(-1.58)	(-2.27)	(-2.10)	(-2.87)
Year 2005	-0.168	-0.272	0.179	-0.0487	-0.325	-0.639	-0.152	-0.282
	(-5.22)	(-5.08)	(2.04)	(-1.16)	(-2.62)	(-2.57)	(-4.72)	(-3.12)
Year 2006	-0.183	-0.296	0.198	-0.0683	-0.346	-0.622	-0.179	-0.265
	(-5.71)	(-5.54)	(2.26)	(-1.63)	(-2.79)	(-2.51)	(-5.56)	(-2.93)
Year 2007	-0.118	-0.15	0.191	-0.0217	-0.322	-0.53	-0.124	-0.165
	(-3.68)	(-2.80)	(2.19)	(-0.52)	(-2.59)	(-2.14)	(-3.83)	(-1.83)
Year 2008	-0.125	-0.129	0.214	-0.0262	-0.341	-0.668	-0.127	-0.146
	(-3.60)	(-2.23)	(2.26)	(-0.57)	(-2.56)	(-2.54)	(-3.63)	(-1.51)
Statellite F12	0.0334	0.0311	-0.0134	-0.00346	0.0925	0.258	0.0039	0.146
	(1.60)	(0.90)	(-0.24)	(-0.13)	(1.12)	(1.56)	(0.19)	(2.46)
Statellite F14	-0.27	-0.251	-0.282	-0.302	-0.206	-0.142	-0.267	-0.293
	(-11.27)	(-6.27)	(-4.32)	(-9.66)	(-2.20)	(-0.76)	(-11.10)	(-4.32)
Statellite F15	-0.0219	-0.0244	-0.046	-0.0589	-0.0126	0.0798	-0.0524	0.0827
	(-0.84)	(-0.56)	(-0.65)	(-1.73)	(-0.13)	(0.40)	(-2.00)	(1.12)
Statellite F16	0.146	0.085	0.069	0.127	0.202	0.367	0.106	0.288
	(5.24)	(1.82)	(0.91)	(3.48)	(1.88)	(1.73)	(3.76)	(3.68)
Constants	-10.37	-8.408	-10.94	-10.82	-11.12	-10.24	-11.66	-5.594
	(-529.25)	(-280.82)	(-198.94)	(-393.87)	(-142.04)	(-63.41)	(-499.57)	(-86.05)
N	353,843	82,339	25,910	203,122	32,421	10,051	283,886	69,957
R-sq	0.623	0.758	0.776	0.564	0.492	0.313	0.593	0.037
Standard error	1.60	1.29	1.18	1.59	1.80	1.93	1.45	1.94

Table SI-5. Regression coefficients of structural equation for the analysis of cells

Table shows the regression coefficients of log GCP density and dummy variables by year and satellite on log luminosity density for cells, with t statistics in parentheses.

| | Country sample for time series | |
Country grade	All countries	Low density countries
All	0.9571	0.7561
A	0.9980	(b)
B	0.9985	(b)
C	0.9781	0.7440
D	0.7277	0.6843
E	(a)	(a)

| | Country sample for cross section | |
Country grade	All countries	Low density countries
All	0.9216	0.7898
A	0.9784	(b)
B	0.9573	(b)
C	0.9262	0.9195
D	0.8837	0.8810
E	(a)	(a)

Table SI-6. Final results of estimated optimal weight on measured output for countries

Table shows the final estimated optimal weight on measured output (1-θ) for countries of different grades using 17-year growth rates and for cross-sectional density.

(a) = insufficient observations
(b) = no observations

Country grade	Cell sample for time series		
	All cells	High density cells	Low density cells
All	0.9854	0.9753	0.9953
A	0.9991	0.9990	0.9992
B	0.9962	0.9965	0.9987
C	0.9832	0.9700	0.9942
D	0.9636	0.8984	0.9990
E	0.0255	(a)	(a)

Country grade	Cell sample for cross section		
	All cells	High density cells	Low density cells
All	0.9579	0.9363	0.9956
A	0.9860	0.9752	0.9991
B	0.9475	0.9454	0.9440
C	0.9480	0.9208	0.9908
D	0.8981	0.8338	0.7346
E	0.7543	0.2189	0.9543

Table SI-7. Final results of estimated optimal weight on measured output for cells

Table shows the final estimated optimal weight on measured output $(1-\theta)$ for cells of different grades for 17-year growth rates and for cross-sectional density.

(a) = insufficient observations

IV method	Cell sample for time series		
Country grade	All cells	High density cells	Low density cells
All	0.9702	1.0000	0.9884
A	0.9994	0.9972	0.9985
B	0.9865	0.9931	0.9981
C	0.9710	0.9905	0.9883
D	0.9956	0.9904	0.9999
E	0.9567	(a)	(a)

IV method	Cell sample for cross section		
Country grade	All cells	High density cells	Low density cells
All	0.9601	0.9402	0.9977
A	0.9858	0.9742	0.9981
B	0.9478	0.9463	0.9194
C	0.9522	0.9296	0.9945
D	0.9084	0.8615	0.9413
E	0.8740	0.6600	0.9995

Table SI-8. IV regressions of luminosity equation for grid cells

Table shows the estimated optimal weight $(1-\theta)$ on measured output with the instrumental-variable method for cells of different grades using 17-year growth rates and for cross-sectional density.

(a) = insufficient observations

States	Prior estimated error for GDP	Sample size	1-θ
17-year growth	0.8% (annual)	49	0.999
Output level	15.00%	1392	0.930

Table SI-9. Estimated optimal weights on measured output for contiguous United States

Table shows the optimal weights on measured GDP output (1-θ) for U.S. states. The luminosity data are aggregated at the state level, and state real annual GDP is obtained from the U.S. Bureau of Economic Analysis.

Calibrated lights (cell data 2006)	Prior estimated error for GDP	Sample size	1-θ
17-year growth rate	4% (annual)	11,059	0.975
Output level	40.00%	12,390	0.956

Table SI-10. Estimated optimal weights for cell data with radiance calibrated nighttime lights (2006)

Table shows the optimal weights for output growth rate and output density measures using the radiance calibrated nighttime lights (2006). Results show little difference using radiance calibrated products than using stable light time series. The detail description on the calibrated products can be found the Defense Meteorological Satellite Program (DMSP) data download page (http://www.ngdc.noaa.gov/dmsp/download.html).
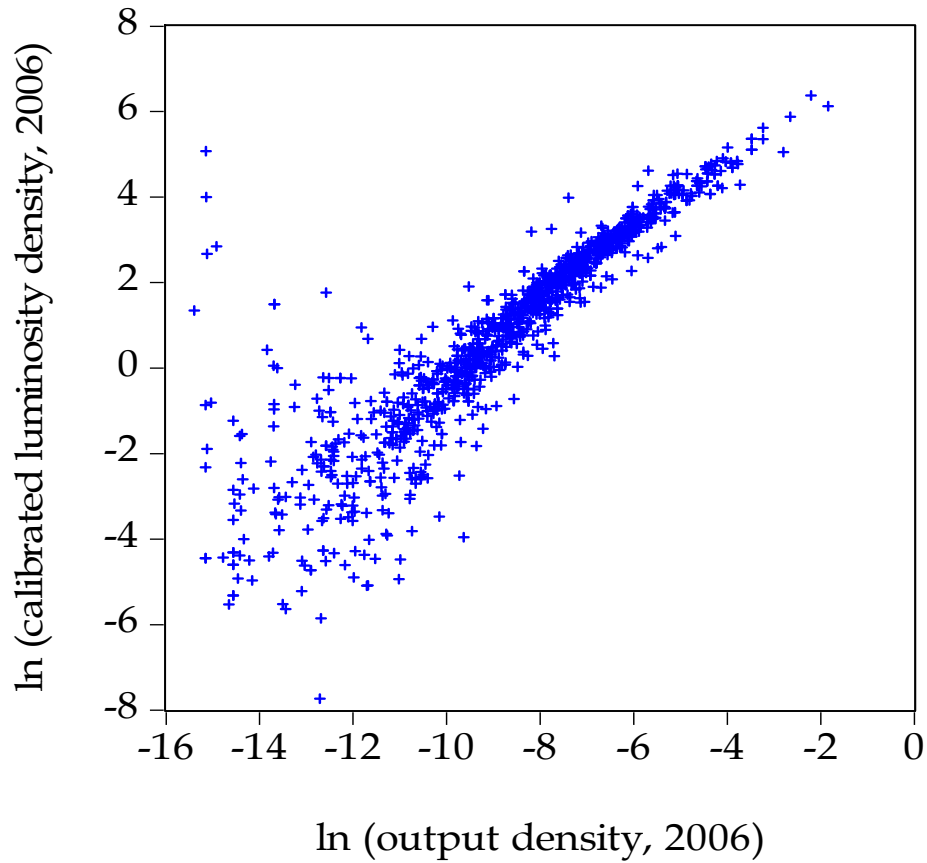
Figure SI-1. Gross cell product (GCP) and luminosity data, United States

Figure shows the scatter plot of log calibrated luminosity for 2006 and log GCP for all cells at the 1° x 1° scale for the United States (N = 1249). Output density is gross cell product (PPP in billions in 2005 international $) per km². Luminosity density per km² is the radiance calibrated luminosity for 2006. The noisy points to the upper left are mainly grid cells in Alaska.
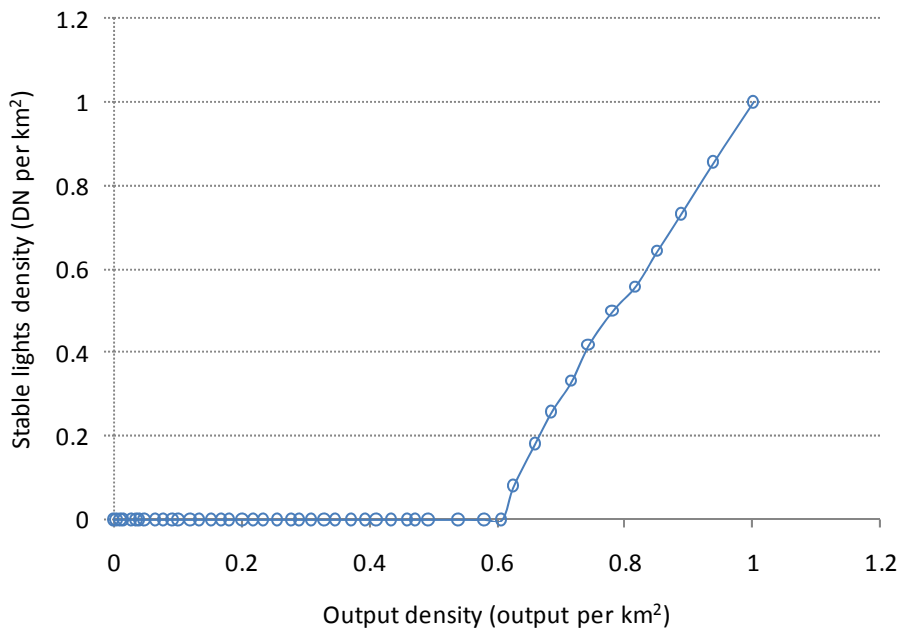
Figure SI-2. Distribution of grid-cell luminosity for Africa by percentile

This graph sorts the 2414 African cells with positive output density into 100 bins to reduce the noise. This shows the relationship between output density and stable lights density for the bottom half (or 50 bins), each represented by a point. Of the 2414 grid cells with RIG > 0.1, almost 40 percent (897) have zero luminosity using the calibrated and stable lights measures. The net result is that luminosity in its current form cannot be reliable used for measuring output in low-density regions.

VII. References

1. Chen X., Nordhaus WD (2010) The Value of Luminosity Data as a Proxy for Economic Statistics, Cowles Foundation for Research in Economics, CFDP 1766.
2. Summers R, Heston A (1991) The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988. *Q J Econ* 106:327–68.
3. Commission of the European Communities—Eurostat, et al. (1993) *System of National Accounts 1993* (United Nations, New York).
4. Johnson SH, Larson W, Papageorgiou C, Subramanian A (2009) Is newer better? Penn World Table revisions and their impact on growth estimates. NBER Working Paper w15455 (National Bureau of Economic Research, Cambridge, MA).
5. Elvidge CD, et al. (2009) A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data. *Energies* 2(3): 595-622.