Planes, Trains, and Automobiles: Night-time Lights of the USA

Dickinson, Jeffrey

2020

Planes, Trains, and Automobiles: Night-time Lights of the USA

Jeffrey Dickinson*

Friday 21st August, 2020

Abstract

This paper seeks to advance understanding of the lights-income relationship by linking the newest generation of night-time satellite images, the VIIRS images, to nationwide, panel data on 3,101 US counties, including data on both population and income. I leverage the quality and frequency of those data sources and the VIIRS lights images to decompose the links between population growth, official GDP growth, and nighttime lights growth at the county level. I use a between-county estimator to identify the effects of time-invariant infrastructure features on night-time light. Roads, rail, ports, and airports I find to be strong contributors to increases in light. I find GDP growth is weakly linked with night-time lights though light growth is strongly linked with population growth even when controlling for substantial non-linearities which appear to be present.

JEL Codes O51, C82, R10, R11, R12

Keywords: night-time light, GDP, population, infrastructure, regional development

*author contact: jeffrey.dickinson@graduateinstitute.ch, Department of International Economics, Graduate Institute of International and Development Studies, CP 1672 - CH-1211 Genève 1, Suisse. I am deeply grateful to my supervisors Nicolas Berman and Jean-Louis Arcand for their extensive support and comments. All remaining errors are my own.

1 Introduction

The literature using nighttime lights satellite images as a proxy measure for human activity dates back to the late 1970's, but the watershed papers in the economics literature were those by Henderson et al. (2012) (n=3,015), and Chen and Nordhaus (2011) (n=11,559). These two papers proposed that nighttime lights could be used as a proxy indicator for income, and they analyzed the correspondence between national accounts data and night-time lights at the highest level of aggregation, the country. They find a strong relationship between income and lights. The authors in Henderson et al. (2012) faced limitations with their data: the reference national accounts data from many low-income countries could be noisy making identification of the exact parameters linking income, GDP, and population difficult. Henderson et al. (2012) notably lacked data on population growth, meaning they were not able to decompose light changes into income and population components. More recent work, using high-quality cross-sectional data from Sweden, has suggested that light growth is more closely linked with population movements than with fluctuations in income (Mellander et al., 2015). Levin and Zhang (2017) utilizes data from the newer VIIRS satellite, the same lights dataset used in this paper, and analyzes lights-income relationship for all the urban areas on the globe (n=4,153) in the months of January 2014 and July 2014. They find that lights are more closely related with *national* income per capita than with population.

With respect to papers whose analysis utilizes nighttime lights at a more detailed level, e.g. at a higher spatial resolution, the literature has been growing. Hodler and Raschky (2014) examine the presence of stronger growth in regions associated with the leader of a country, and find a significant result. Mellander et al. (2015), perhaps the paper most similar in spirit to this one, is a well-cited paper which examines the relationship between economic activity, population, enterprise density, and nighttime light in Sweden using cross-sectional analysis. The authors find that light growth corresponds most to nighttime population density (population), rather than daytime enterprise density. Mellander et al. (2015) also argue that night-time light is only weakly correlated with income, although in their OLS regressions night-time light appears to increase by 0.424 units with an increase of one unit of Total Wage Incomes. Two new papers have recently been published using night-time lights for localized analysis. One recently published paper measures the effects on light of flooding in cities around the globe, and finds that low-lying areas in cities recover as fast as other areas, and there appear to be no permanent effects of flooding on city development (Kocornik-Mina et al., 2020). Frick et al. (2019) uses night-time lights data to analyze the effect of special economic zones on economic activity. They find that key determinants to the success of special economic zone was linked with pre-existing industrial infrastructure in the surrounding area, and the existence of large markets in which to sell outputs. Bleakley and Lin (2012) uses night-time lights from the years 1996-7 to test for path-dependence around certain natural water features in the United States. The authors find that portage sites, sites where, in the past, transport boats could not pass and thus cities arose, are likely to still be of a substantial size around 100 years after the portage sites were relevant. Smith and Wills (2018) is a recent paper which leverages the global nighttime lights coverage to estimate the fraction of the population below the poverty line, and they find that spillovers from economic activity rarely disseminate to rural populations. An overview of the capabilities and some applications of night-time lights data can be found in Donaldson and Storeygard (2016).

The principal contribution of this paper is to further collective understanding of the lights-income relationship by linking lights to panel data, including data on both population and income, administrative data of the highest quality, which are available at a fine resolution. I also compare the nighttime light measure alongside electrical consumption data at the county level in California over the sample years. Previous authors have suggested that electrical consumption data may be of a similar value to NTL as a proxy indicator (Mellander et al., 2015; Henderson et al., 2012). I find that electrical consumption does correlate with higher levels of GDP and population, though in the within-county model we only see an effect of increases in the population on an increase in non-residential light.

The United States, in contrast with the data from Sweden used in Mellander et al. (2015), is a much larger landmass and total population (10m vs. 350 m), and has substantial heterogeneity with respect to landmass and shape, demographic composition, population density, and geographical characteristics such as mountains, lakes, rivers, and coastlines. This is evident when we consider places like California, which has only 58 counties per 40m citizens, Alaska, which is has enormous counties but is sparsely populated, Arizona, which is mostly desert and borders Mexico, Washington which has dense deciduous and evergreen forest, mountains, and a shared border with Canada, as well as Hawaii, an island halfway between the US and Japan in the Pacific ocean.

In striving to obtain the most accurate point estimates for the effects of population and income on nighttime lights, I estimate two functional forms for the nighttime light production function: log-linearized cobb-douglas, and translog. The translog functional form allows me to control for the potential nonlinear relationships with respect to interactions among the independent variables and also capture the higher-order behavior of the light production function.

A second principal goal of this paper is to uncover key drivers night-time light growth by analyzing which important infrastructure features may be driving light, holding constant income and population changes. This is done using a between-groups estimator, a procedure which is designed to permit identification of the effects on night-time light of geographic features which are invariant in the sample period. Nighttime light is increasing in nearly all of the infrastructure elements examined here including rail, roads, ports, and airports.

The rest of the paper will proceed as follows: section 2 introduces further details on the methodology used in the paper. Section 3 discusses the data sources and availability including a detail description of the VIIRS nighttime lights data. Section 4 presents the results, and section 5 concludes.

2 Methodology

Based on the results of Mellander et al. (2015), night-time light appears to be a proxy for consumption more than production. A principal approach of this paper is to use empirical tools reveal the links between population growth, income growth, and night-time light as measured. Using night-time light as the dependent variable makes the most sense in the context because the satellite images from the VIIRS are somewhat noisy, while they are very precise in the dimension of how they record the texture of activity across space.[1] As such, using the night-time lights measure as the dependent variable makes more sense, I argue, than trying to use night-time light to predict income or population size. The latter is left for future research. The general model, a night-time light production function, states simply that night-time light, as measured from the VIIRS sensors is a function of income, population, and other factors:

[1]This will be discussed further in the data section. The night-time lights images must undergo processing in order to remove image distortions which are orthogonal to changes in human-made light.

$$NTL_{ct} = \beta_1[Income_{ct}] + \beta_2[Population_{ct}] + \beta_3[Area_{ct}] + \alpha_c + \varepsilon_{ct} \tag{1}$$

Where c indexes the county, t indexes the year, and α_c are the county fixed effects. The area variable controls for any potential relationship between the size of the county and the measurement of the lights that may not be captured by the income and population variables. Based on previous papers, most notably Hu and Yao (2019), there is reason to believe that income and population may not enter the night-time light production function linearly. This is an important consideration for our purposes as nonlinearities may mask desired effects of interest. In that case I will also estimate the following specification, which includes squared terms and interaction terms among all three key independent variables. The intuition behind the squared terms is that there could be strongly diminishing effects in the way that income and population enter the production function. The interaction terms are included to capture the possibility that the lights-income or lights-population relationship could be stronger in larger counties or smaller ones. The third main variable besides income and population being the area of the county, which controls just for the total size of the county, as there is quite a large variation. The second potential specification is therefore the following:

$$NTL_{ct} = \beta_1 X + \beta_2(X^2) + \beta_3(x_1 \times x_2 ...) + \varepsilon_{ct} \tag{2}$$

Between-county Estimation

There are certain geographic characteristics of the counties which we would like to analyze, but it is difficult because these counties do not have any variation in those characteristics within the sample period, 2012-2018. In order to obtain identification, all variables are collapsed to their group means. This procedure is similar to the strategy employed in Henderson et al. (2012), who also employ the within-transformed country-level data, and then in their case they used long-differences instead of group means. Identification of the effect of the infrastructure or geographic features then comes from comparing counties which have infrastructure or features exclusively to other counties within the same state that lack those features. Given the size of the sample (n=21,728 county-years) and the survey period I feel this is the most appropriate approach to consider the effects of geographic variables.

3 Data

Table 1 details years of data availability. The LandScan data has the best coverage through time, while the VIIRS nighttime lights series starts only in 2012. The binding constraint on our sample is therefore the population data as we have no estimates for population at the county level past 2018, and I am able to leverage the years 2012-2018.

	ACS Pop. 5-yr Est.	LandScan Pop. Est.	BLS GDP	VIIRS Nighttime Lights
Years of Availability	2009-2018	2000-2018	2001-2018	2012-2020

Table 1: Data Availability

3.1 Bureau of Labor Statistics GDP Data

Over the past few years the Bureau of Labor Statistics (BLS) has been releasing local-area calculations for gross domestic product. In the BLS GDP statistics, GDP is calculated using the income approach. Based on the availability of data, the Bureau of Economic Analysis (BEA) utilizes the income method for calculating GDP. *"GDP is computed as the sum of compensation of employees, taxes on production and imports less subsidies, and gross operating surplus. The initial regional estimates are then scaled to the national estimates so that all BEA estimates are reconciled"* (Aysheshim et al., 2020). Principal sources of the GDP data are the Department of Labor's Quarterly Census of Earnings and Wages, aircarrier traffic statistics, DOT surface transportation data, bank branch deposits, and other proprietary government sources. A full accounting of all sources and information used in the calculation of GDP at the county level are described in Aysheshim et al. (2020). There is substantial variation in the GDP data, some counties produce millions of dollars, while others produce well under 100k per annum.

3.2 Census Bureau ACS County-Level Population Data

Population estimates come from American Community Survey (ACS) 5-year estimates of the county-level population. These are calculated using data sampled from the county on a rolling basis over the course of 5 years. ACS data are the main survey data that are collected from communities in the United States in the intercensal period.

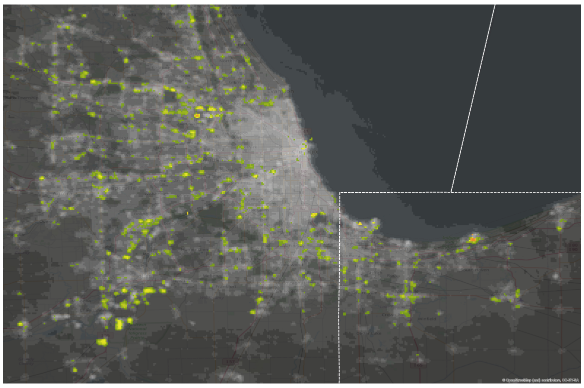
3.3 LandScan Gridded Population Data

LandScan gridded population data is a global dataset in the form of an integer-based raster. The population is inferred using an algorithm and a mix of sources, the principal source being daytime satellite imagery of human settlements. These data are quite popular, and have been used in other economics research when comparable administrative population data are not available.

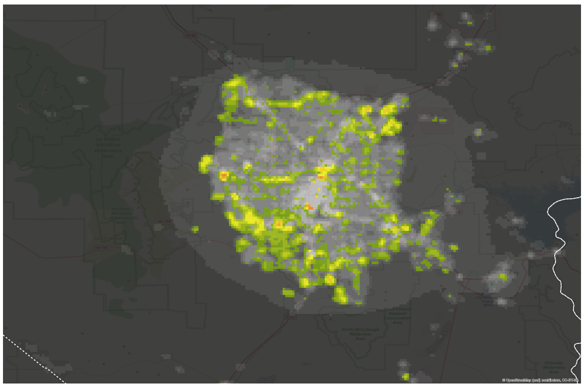
3.4 VIIRS Night-time Lights Data

The Suomi-NPP Satellite project, which started in 2011, is a joint civilian venture of the United States National Aeronatuic and Space Administration (NASA), the Department of Defense, and the National Oceanographic and Atmospheric Administration. The Visible Infrared Imaging Radiometer Suite (VIIRS) is intended to capture human-made light and overcomes many limitations of the previous Defense Meteorological Satellite Program (DMSP) satellite images. The newer Suomi NPP satellite, which contains the VIIRS, has an automatic gain sensor which adjusts to allow great sensitivity, meaning the device can better capture much lower and higher levels of light (Elvidge et al., 2017). The resolution of the new VIIRS images, available from 2012-2019, with data available on a daily frequency or in monthly composite forms, is extremely high, with pixels being around 742m across compared to the DMSP pixels which were 3km across (Carlowicz, 2012; Elvidge et al., 2017). This sensitivity is of extreme interest to researchers in attempting to pinpoint precise locations which are centers of economic activity, and will reduce limitations around night-time lights data coming from heavily saturated urban areas. The Suomi-NPP satellite flies over the earth to capture imaging using the spectroradiometer, a device used to capture light similar to the capture device in a digital camera, and passes by around 1:30am and 1:30pm local time each day (Carlowicz, 2012). Raw data from the sensor are then processed to remove non-human generated disturbances such as aurora borealis, stray light, natural fires and other light which could potentially introduce noise. A detailed accounting of the processing of the data can be found in (Elvidge et al., 2017).

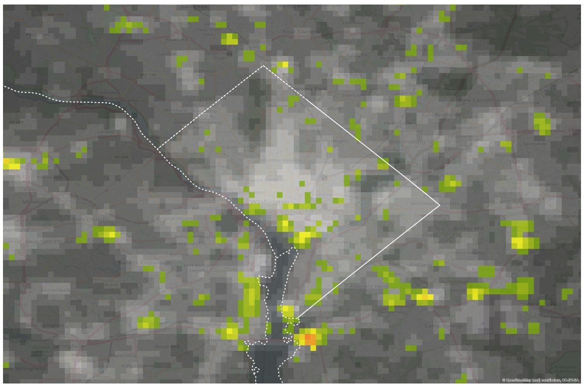
Some examples of night-time lights images of major US cities are shown in Figure 1. Long-run changes in night-time light are shown in green-red colors to demonstrate intensity. Chicago, IL is shown in the upper left panel, and is seen to be quite spread out over space. Las Vegas, NV is an interesting example because of its intensity relative to the darkness of the nearby unpopulated desert. Washington, DC provides a good illustration of how, despite high density of lights, light intensity can still be distinguished at a high resolution. The dark red spot just south of Washington, DC is National Harbor, an area of major development for the DC metropolitan area over the last few years. The major development within DC over that period was the Southwest Waterfront, which can also be seen as the glowing yellow dot at the southern tip of DC where the Potomac River meets the Anacostia. Lastly, one of the wealthiest, most expensive, and most productive regions in the country is depicted in Northern California from Berkeley to San Jose, revealing pockets of development along the way.



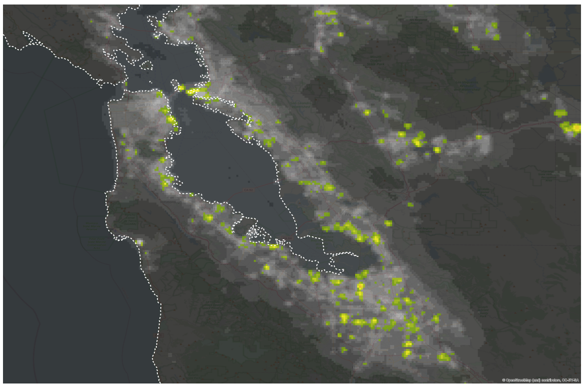
(a) Chicago, IL



(b) Las Vegas, NV



(c) Washington, DC



(d) San Francisco, CA

Figure 1: Night-time Lights of Four Major US Cities;
Layers: Basemap: Open Street Map, CC License; Night-time Lights Annual Image (2019); Changes in NTL 2012-2019 - Green = small change, Red = large change

3.5 California Electrical Consumption Data

California's state energy agency, California Energy Commission, makes available electrical consumption data at the county level for all counties in California.[2] These data are available at

[2] https://ecdms.energy.ca.gov/elecbycounty.aspx

VARIABLES	(1) Total NTL	(2) Total NTL	(3) Total NTL	(4) Total NTL
Commerical Elec. Cons.	0.712***			
	(0.0178)			
Residential Elect. Cons.		0.772***		
		(0.0243)		
Combined Elect. Cons.			0.763***	0.593
			(0.0183)	(0.557)
Observations	406	406	406	406
R-squared	0.869	0.806	0.868	
Number of Counties				58

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

Table 2: California Nighttime Lights (log) Regressed on the Log of Electrical Consumption

the county level from 1990-2018. They are administrative in nature and are therefore, to the best of my knowledge, do not represent a sample of electrical consumption data. A regression of NTL on electrical consumption can be seen in table 2. As we can see, nighttime light is strongly correlated with electrical consumption, slightly more so with non-residential electrical consumption.

3.6 Infrastructure Data

Infrastructure data, including the location of ports, rail, navigable waterways, and Fortune-500 business headquarters have been collected from the U.S. federal government's Homeland Infrastructure Foundation Level Database (HIFLD) website, which is funded under the Department of Homeland Security. Airport locations were taken from open data sources.[3] Data on primary roads, which includes interstates and principal highways, was collected from the US Census Department.

Tables 9 and 10 show the counties with the most and least light, and are included in the appendix. The variance in light is substantial, from Robertson County, KY, the county with the least total light, to Yukon-Koyukuk County, AK with the most light.

4 Results

[3]https://ourairports.com/

VARIABLES	N	mean	median	sd	min	max
Total Nighttime Light	406	54822	17507	112144	755.6	822111
BLS GDP	406	41730000	7615000	97600000	47224	710900000
LS Population	406	668138	181767	1453000	1140	10140000
ACS Population	406	669915	181536	1452000	1057	10120000
miles^2	406	2727	1554	3097	48.56	20118
km^2	406	7063	4024	8020	125.8	52104
Non-residential Elec. Cons.	406	3315	781.4	7021	4.008	49193
Residential Elec. Cons.	406	1585	553.2	3090	9.291	21162
Total Elec. Con.	406	4901	1474	10032	13.89	69946

Table 3: Summary Statistics of Variables Used in Electrical Consumption Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Total NTL	Total NTL	Total Elec	Total Elec.	Resid. Elec.	Resid. Elec.	Comm. Elec.	Comm. Elec.
Area	0.486***		0.147***		0.209***		0.0472***	
	(0.0206)		(0.0143)		(0.0205)		(0.0133)	
BLS GDP	0.551***	0.261***	0.235***	0.0419	0.392***	0.0993	-0.00390	-0.00551
	(0.0572)	(0.0790)	(0.0272)	(0.0337)	(0.0503)	(0.131)	(0.0484)	(0.0382)
ACS Population	0.0974	-1.239	0.672***	0.525*	0.555***	0.374	0.878***	0.712***
	(0.0637)	(0.926)	(0.0292)	(0.300)	(0.0562)	(0.393)	(0.0545)	(0.178)
Constant	-3.670***		-5.638***		-7.688***		-4.616***	
	(0.296)		(0.182)		(0.274)		(0.213)	
Observations	406	406	406	406	406	406	406	406
R-squared	0.922		0.981		0.956		0.964	
Number of Counties		58		58		58		58
County FE		yes		yes		yes		yes

Columns 1,3,5,7: clustered standard errors (county) in parentheses
Columns 2,4,6,8: cluster-robust standard errors (county) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: California Electrical Consumption Regressions

4.1 California Electrical Consumption Regressions

Table 3 contains the summary statistics of variables used in the electrical consumption regressions, and table 4 shows the results of regressions those regressions. The availability and granularity of the California data permit the direct comparison of the value-added of night-time lights over electrical consumption data. Columns 1-2 are the regression of only the California night-time lights using the same set of parsimonious controls as earlier. We see in column 1 and 2 that nighttime lights tracks with BLS GDP in California as well as the area, and this relationship is significant both in the global and the within regressions. With respect to the electrical consumption data, they track more closely with increases in the population as we see in column 3, and in column 4, which is the within-county transformed regression, none of the independent variables are significant. Looking at columns 5-8 which are residential (5-6) and non-residential (7-8) electrical consumption separated out, we see that population moves with electrical consumption, but that income moves with electrical consumption less, and income is only statistically significant in column 5, global-OLS with year fixed effects.

4.2 Cobb-Douglas Estimation

Summary statistics for the principal regression variables can be found in table 5 The county population variable, LandScan version, the smallest county has 85 residents, Loving, TX while the largest has 10,140,000, Los Angeles, CA. The ACS 5-year estimates are very similar. The results presented in Table 6 are the estimates of the Cobb-Douglas nighttime light production function. Columns 1-3 are the most parsimonious specifications, omitting all interactions, where column 2 uses the ACS 5-year estimates and column 3 uses LandScan estimates. The estimated equation is the the same as equation 2, though the squared terms are omitted.

All variables are in their log form and all columns include state·year fixed effects. Column 1 shows that area contributes significantly to the total amount of light, but so does income, almost as much as area, with population growth tracking the least strongly with night-time light. Column 2 is the county fixed-effect within estimates of the same, income and population are significant, though population has a significant and negative effect on night-time light. With columns 2-3, there is not much difference between the results for the point estimates for income with ACS population estimates and LandScan population data, though the point estimates for the effect of population on night-time light are now slightly smaller and they remain negative and statistically significant. In columns 4-6 I have now incorporated an interaction of the key variables of interest with the area variable. The intuition for this is that the relationship between lights and population or lights and income might be changing with the size of the county in question. Adding the interaction terms in columns 3-6 increases the R^2 from .90 to .92 in column 1,3 , from .68 to .69 for the ACS County-FE estimates in column 2 and 4, and from .68 to .69 in columns 3 and 6, where the LandScan data are used.

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max	(13) p10	(14) p25	(15) p50	(16) p75	(17) p90
Total NTL (sum of pixel values)	21,728	17,485	66,982	447	2922000	2,292	3,590	6,476	13,506	31,997
BLS GDP Data	21,728	5533000	24240000	2753	710900000	163270	335217	876331	2616000	9275000
LandScan Population Estimates	21,728	103,045	333,748	81	10140000	4,821	10,569	24,921	67,781	205,340
ACS 5-year Estimates	21,728	104,246	332,430	86	10120000	5,144	11,021	26,017	68,958	208,518
Miles	21,728	1,160	3,710	15.67	147066	311.5	443.6	636.2	950.3	1,884
km2	21,728	3,004	9,610	40.57	380898	806.9	1,149	1,648	2,461	4,880
Log of BLS GDP	21,728	13.86	1.604	7.921	20.38	12	12.72	13.68	14.78	16.04
Log of LS Population Estimates	21,728	10.24	1.505	4.407	16.13	8.481	9.266	10.12	11.12	12.23
Log of Total NTL	21,728	8.934	1.072	6.105	14.89	7.738	8.186	8.776	9.511	10.37
Log miles 2	21,728	6.557	0.8	2.813	11.90	5.745	6.097	6.457	6.858	7.542
log of ACS 5-year Pop. Est.	21,728	10.29	1.486	4.466	16.13	8.546	9.308	10.17	11.14	12.25
Has Port	21,728	0.03	0.16	0	1.00	0	0	0	0	0
Has Rail Infrastructure	21,728	0.88	0.32	0	1.00	0	1	1	1	1
Has Primary Road	21,728	0.45	0.50	0	1.00	0	0	0	1	1
Has Airport	21,728	0.32	0.47	0	1.00	0	0	0	1	1
Has Multiple Airports	21,728	0.07	0.26	0	1.00	0	0	0	0	0
Has All Four Elements	21,728	0.08	0.27	0	1.00	0	0	0	0	0

Table 5: Variables Used in County-Level Night-time Lights Regressions

VARIABLES	(1) Total NTL	(2) Total NTL	(3) Total NTL	(4) Total NTL	(5) Total NTL	(6) Total NTL
Area	0.444***			1.234***		
	(0.00757)			(0.0474)		
BLS GDP	0.475***	0.0717***	0.0688***	-0.534***	-0.0193	-0.0664
	(0.00889)	(0.0275)	(0.0266)	(0.0580)	(0.243)	(0.241)
ACS Pop	0.153***	-0.480***		0.821***	-1.370**	
	(0.0105)	(0.104)		(0.0674)	(0.660)	
LS Pop			-0.360***			0.387
			(0.0603)			(0.335)
Area × ACS Pop				-0.178***	0.208**	
				(0.00911)	(0.1000)	
Area × BLS GDP				0.0775***	0.0522	0.0548*
				(0.00847)	(0.0333)	(0.0329)
ACS Pop × BLS GDP				0.0407***	-0.0334***	
				(0.00127)	(0.00877)	
Area × LS Pop						-0.0486
						(0.0486)
LS Pop × BLS GDP						-0.0300***
						(0.00888)
Observations	21,707	21,707	21,707	21,707	21,707	21,707
R-squared	0.901	0.685	0.688	0.927	0.690	0.692
Number of Counties		3,101	3,101		3,101	3,101
State-Year FE	yes	yes	yes	yes	yes	yes
County FE		yes	yes		yes	yes

Cols. 1,4 clustered standard errors (state) in parentheses
Cols. 2,3,5,6, cluster-robust standard errors (county) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: County-Level Night-time Lights Regressions

Column 4 contains the most interesting results, possibly because the effects are not well-identified within-counties. Column 4 shows that area is increasing light significantly, the level effect of GDP is now negative and significant, population is positive and significant, though area×Population is negative and significant, area×GDP is positive and significant. It is important to note that this effect is the most important, the GDP*area effect, and this appears to be consistently estimated regardless of specification. Furthermore, the interaction on income×GDP is also significant and positive, though the effect is relatively smaller than the others.

4.3 Translog Functional Form

In Table 7 we have the full specification including all interactions and squared terms. Columns 1-2 contain the ACS population data, while 3-4 contain the estimates with the LandScan population data. All of the squared terms are statistically significant in all columns. GDP appears to have increasing returns to scale with the pooled OLS model, but decreasing returns on the within-county estimation. With respect to the interaction terms it is more of a mixed bag. Nearly all of the interaction terms are significant. The area × GDP is positive, while area × population is negative. The effect of income × population, if present, does not appear to be identified within-county. The specifications in table 2 represent the preferred specifications, where it is not necessary to conserve degrees of freedom, given the importance of the second order terms.

4.4 Between-County Economic Geography Regressions

Table 8 is the results of the between-group estimation. The procedure is used in the case that we have variables of interest which are unchanging in the period of data availability, in this case those variables are infrastructure variables. In this estimation the same fully specified model is used as before, but the ACS population variable is now dropped, since the LandScan estimates appear to perform better. These regressions are meant to demonstrate the importance of geographic variables in the production of nighttime light, which would be impossible without the use of this econometric procedure. My preferred specification would be column 2 or column 4. All of the level terms are significant except for GDP, which is negative and not significant.

VARIABLES	(1) Total NTL	(2) Total NTL	(3) Total NTL	(4) Total NTL
Area	0.410***		0.357***	
	(0.0547)		(0.0583)	
BLS GDP	0.0340	0.219	0.0161	0.281
	(0.0822)	(0.268)	(0.0903)	(0.274)
ACS Pop	0.277***	2.233***		
	(0.0884)	(0.679)		
LS Pop			0.345***	0.952**
			(0.0937)	(0.415)
BLS GDP^2	0.0147**	-0.0309***	0.0195***	-0.0265**
	(0.00714)	(0.00886)	(0.00577)	(0.0103)
ACS Pop^2	0.0721***	-0.190***		
	(0.00719)	(0.0316)		
LS Pop^2			0.0942***	-0.0413**
			(0.00507)	(0.0180)
Area^2	0.0613***		0.0612***	
	(0.00359)		(0.00372)	
Area × BLS GDP	0.0461***	0.0741**	0.0609***	0.0662**
	(0.0108)	(0.0331)	(0.0112)	(0.0337)
Area × ACS Pop	-0.136***	0.160**		
	(0.0109)	(0.0754)		
Area × LS Pop			-0.152***	-0.0792
			(0.0110)	(0.0485)
ACS Pop × BLS GDP	-0.0408***	0.0159		
	(0.0127)	(0.0178)		
LS Pop × BLS GDP			-0.0667***	0.00236
			(0.00892)	(0.0218)
Observations	21,707	21,707	21,707	21,707
R-squared	0.934	0.696	0.935	0.695
Number of Counties		3,101		3,101
State-Year FE	yes	yes	yes	yes
County FE		yes		yes

cols. 1,3 clustered standard errors (state) in parentheses
cols 2,4 cluster-robust standard errors (state) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: County-Level Night-time Lights Regressions, Full Specification

VARIABLES	(1) Total NTL	(2) Log NTL	(3) Log NTL	(4) Log NTL
Area	0.372*	0.357**	0.481***	0.341**
	(0.192)	(0.172)	(0.165)	(0.173)
BLS GDP	-0.828***	-0.305	-0.318	-0.294
	(0.295)	(0.269)	(0.265)	(0.270)
LS Pop	1.541***	0.652**	0.720***	0.657**
	(0.280)	(0.263)	(0.258)	(0.264)
Area^2	0.0496***	0.0616***	0.0537***	0.0620***
	(0.0122)	(0.00950)	(0.00908)	(0.00946)
BLS GDP^2	0.0556***	0.0450***	0.0461***	0.0446***
	(0.0165)	(0.0155)	(0.0153)	(0.0154)
LS Pop^2	0.111***	0.103***	0.103***	0.103***
	(0.0109)	(0.00865)	(0.00874)	(0.00864)
Area $\times$ BLS GDP	0.138***	0.0612*	0.0625*	0.0617*
	(0.0350)	(0.0328)	(0.0320)	(0.0329)
Area $\times$ LS Pop	-0.246***	-0.152***	-0.156***	-0.152***
	(0.0356)	(0.0316)	(0.0312)	(0.0317)
LS Pop $\times$ BLS GDP	-0.135***	-0.104***	-0.107***	-0.104***
	(0.0211)	(0.0179)	(0.0178)	(0.0179)
Has Port	0.206***	0.123***	0.122***	0.110***
	(0.0514)	(0.0305)	(0.0298)	(0.0304)
Has Primary Road	0.105***	0.105***	0.106***	0.101***
	(0.0139)	(0.0102)	(0.0102)	(0.0102)
Has Rail	0.0887***	0.0624***	0.0629***	0.0622***
	(0.0234)	(0.0174)	(0.0173)	(0.0174)
F500 HQ	0.0468	-0.0473	-0.0507	-0.0449
	(0.0477)	(0.0368)	(0.0358)	(0.0366)
Has Airport	-0.00214	-0.00172		-0.0106
	(0.0182)	(0.0130)		(0.0133)
Has Multiple Airports	0.118***	0.0181		0.0139
	(0.0407)	(0.0251)		(0.0250)
Airport Count			0.0286***	
			(0.00957)	
Has All Four				0.0556***
				(0.0209)
Observations	3,101	3,101	3,101	3,101
R-squared	0.899			
State FE	no	yes	yes	yes

Col. 1, POLS, clustered s.e. (state) in parentheses
Cols. 2-4, cluster-robust s.e. (state) in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8: Between-County Procedure: Economic Geography Variables

The direct effect of population on long-term NTL appears to be very strong, 0.65-0.72. Looking at the squared terms, all of them are statistically significant and positive, indicating increasing marginal returns to all of those variables. Night-time light is decreasing in the product of area and population. This means that as an area gets larger and as population increases, the relationship between both area and NTL and population and NTL is diminished. Controlling for population and its derivatives we see that the affect of GDP*Area is positive and quite similar to point estimates from the previous sections, though this variable is not statistically significant at standard levels in this result. Last we have the income*population interaction term, which indicates higher population or higher income results in lower levels of night-time light. This effect appears large in the between-county estimates, though in the previous estimations (Table 6) it was negative and statistically significant, but the point estimate is closer to .05.

Turning to the geographic variables, these are the main variables of interest for the between-county regressions. The geographic control variables include an indicator variable for 1 - Ports, 2 - Principal Roads, 3 - Rail Infrastructure Present, 4 - Airports. There are three different specifications using geographic variables which are presented, column 2 contains a separate indicator for a single airport or multiple airports while column 3 uses the total number of airports as a control. It could be that most of the result was coming from counties with multiple airports, and it was for that reason that this is included. More airports means greater night-time light. Each airport increases light by .0286 according to my estimates, statistically significant at the .01 level. I include an interaction dummy variable that takes 1 if county has all of the following: a port, a major road, rail infrastructure, and at least one airport. This variable then indicates that the simultaneous presence of all of those infrastructure elements are a determinant of night-time light. The point estimates of this effects are .0556, statistically significant at the .01 level. This result can be thought of as the additional marginal benefit of having the combined presence of ports, roads, rail, and airports, over having an individual airport, port, rail, or road. The inclusion of the final interaction term in column 4 does not appear to substantially alter the point estimates the other infrastructure variables except for the effect of a port variable, which declines from 0.122 to 0.110.

5 Conclusion

Using administrative and survey data of the highest quality, pairing these data with the newest VIIRS night-time satellite imagery, I analyzed the relationship between population, income, geographic variables and human-generated night-time light measured at the county level. I find that the strength of the relationship between nighttime lights and population changes is strong, while there is also evidence for a correspondence between light and changes in income, particularly in larger counties. These results hold even after incorporating higher-order terms to account for the potential for nonlinearities in the lights-income-population nexus. I also evaluate the value-added of nighttime lights over electrical consumption data, and find that electrical consumption is more sensitive to changes in population growth than changes in income. Although night-time light only moves slightly with income, light still moves more consistently with income than does electrical consumption data, and night-time light is therefore still likely to be a useful proxy indicator for changes in population or income in small geographic areas for which accurate and timely data are not readily available. I also utilize a between-county estimator to measure the effects of important infrastructure elements on light, and they are found to substantially influence light production. These findings could be useful to future researchers looking to use VIIRS imagery for economic analysis. I argue that based on these results, night-time light is found to be a strong proxy indicator for population changes, while it is only a weak indicator for changes in income.

References

Aysheshim, K., Hinson, J. R., and Panek, S. D. (2020). A primer on local area gross domestic product methodology. *Survey of Current Business*, 100(3):1–13.

Bleakley, H. and Lin, J. (2012). Portage and path dependence. *The quarterly journal of economics*, 127(2):587–644.

Carlowicz, M. (2012). Out of the blue and into the black: New views of the earth at night. [Online; posted 5-December-2012].

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4):171–98.

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

Frick, S. A., Rodríguez-Pose, A., and Wong, M. D. (2019). Toward economically dynamic special economic zones in emerging countries. *Economic Geography*, 95(1):30–64.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, 102(2):994–1028.

Hodler, R. and Raschky, P. A. (2014). Regional favoritism. *The Quarterly Journal of Economics*, 129(2):995–1033.

Hu, Y. and Yao, J. (2019). Illuminating economic growth. Technical report, International Monetary Fund.

Kocornik-Mina, A., McDermott, T. K., Michaels, G., and Rauch, F. (2020). Flooded cities. *American Economic Journal: Applied Economics*, 12(2):35–66.

Levin, N. and Zhang, Q. (2017). A global analysis of factors controlling viirs nighttime light levels from densely populated areas. *Remote sensing of environment*, 190:366–382.

Mellander, C., Lobo, J., Stolarick, K., and Matheson, Z. (2015). Night-time light data: A good proxy measure for economic activity? *PloS one*, 10(10).

Smith, B. and Wills, S. (2018). Left in the dark? oil and rural poverty. *Journal of the Association of Environmental and Resource Economists*, 5(4):865–904.

State	County	year	Total NTL	BLS GDP	LS Pop	ACS Pop	square_miles	square_km
Alaska	Yukon-Koyukuk	2017	2921585	258303	5366	5396	147066	380898
Alaska	Yukon-Koyukuk	2016	2741543	260813	4795	5423	147066	380898
Alaska	Yukon-Koyukuk	2015	2596611	247510	6657	5466	147066	380898
Alaska	Yukon-Koyukuk	2014	2470665	226243	6693	5464	147066	380898
Alaska	Yukon-Koyukuk	2013	2123825	277385	6840	5564	147066	380898
Alaska	North Slope	2017	1989463	11231169	8976	9831	90793	235153
Alaska	North Slope	2015	1941614	11130682	9379	9795	90793	235153
Alaska	Yukon-Koyukuk	2012	1937930	316396	6834	5624	147066	380898
Alaska	North Slope	2016	1867156	10567213	8218	9718	90793	235153
Alaska	North Slope	2018	1769743	10469543	14320	9872	90793	235153
Alaska	North Slope	2013	1620345	7251453	9388	9786	90793	235153
Alaska	North Slope	2012	1131531	8920976	9343	9692	90793	235153
Alaska	Northwest Arctic	2016	980246	591812	6639	7689	36771	95236
Alaska	Northwest Arctic	2017	925620	680814	7527	7767	36771	95236
Alaska	Northwest Arctic	2013	867246	667707	7685	7725	36771	95236
Texas	Harris	2017	824801	351838304	4844329	4664159	1760	4557
California	Los Angeles	2017	822111	688661568	10132862	10118759	4088	10587
Alaska	Northwest Arctic	2015	811720	577594	7719	7771	36771	95236
Texas	Harris	2013	800395	390463008	4472666	4355158	1760	4557
Texas	Harris	2015	783815	358868384	4676992	4561939	1760	4557
Texas	Harris	2014	779031	392944160	4581052	4458709	1760	4557
California	Los Angeles	2018	757890	710893248	10100543	10105518	4088	10587
California	Los Angeles	2014	747704	630438080	10081448	10048408	4088	10587
Illinois	Cook	2014	743964	350384992	5403468	5257481	962	2492
California	Los Angeles	2015	739414	653885056	10143410	10097037	4088	10587
Alaska	Southeast Fairbanks	2017	735827	640754	6888	6885	26183	67813

Table 9: Top 25 US Counties in Total Light 2012-2018

name_1	name_2	year	Total NTL	BLS GDP	LS Pop	ACS Pop	square_miles	square_km
Kentucky	Robertson	2016	447	26076	1984	2125	101	261
Kentucky	Robertson	2012	459	19574	1867	2216	101	261
Washington	Wahkiakum	2016	515	96746	3414	4167	262	678
Kentucky	Robertson	2013	515	19937	1868	2216	101	261
Kentucky	Robertson	2015	524	24690	1791	2135	101	261
Washington	Wahkiakum	2013	528	64330	3583	4033	262	678
Massachusetts	Nantucket	2016	528	1695910	11101	11124	48	126
Virginia	Highland	2016	533	101481	1918	2209	420	1087
Massachusetts	Nantucket	2013	563	1031003	10910	10567	48	126
Washington	Wahkiakum	2015	564	97635	3586	4027	262	678
Massachusetts	Nantucket	2018	576	1791518	11358	11327	48	126
Massachusetts	Nantucket	2014	594	1116569	11352	10839	48	126
Virginia	Rappahannock	2016	598	267250	6420	7352	265	688
Washington	San Juan	2012	599	492193	14860	15849	181	470
Virginia	Mathews	2016	607	174844	6791	8789	89	231
Georgia	Taliaferro	2016	608	40701	1364	1613	195	506
Washington	San Juan	2015	611	601531	15243	16198	181	470
Massachusetts	Dukes	2016	611	1678037	16831	17316	110	286
Massachusetts	Nantucket	2015	613	1673678	11467	10945	48	126
West Virginia	Wirt	2016	622	58728	5165	5767	232	600
Virginia	Highland	2012	633	46315	1767	2234	420	1087
Massachusetts	Nantucket	2017	633	1722140	11411	11270	48	126
Kentucky	Robertson	2018	638	25531	1804	2135	101	261
Georgia	Glascock	2016	644	45753	2680	2979	144	374
Kentucky	Owsley	2016	648	51987	4396	4473	198	513
Washington	San Juan	2016	658	621278	14145	16304	181	470

Table 10: Bottom 25 US Counties in Total Light 2012-2018