Proceedings of the Asia-Pacific Advanced Network ***2011*** *v. 32, p. 218-236.*
http://dx.doi.org/10.7125/APAN.32.24
ISSN 2227-3026

Night Lights and Economic Activity in India: A study using DMSP-OLS night time images

Laveesh Bhandari* and Koel Roychowdhury

Indicus Analytics Pvt. Ltd., 2nd Floor, Nehru House, 4, Bahadur Shah Zafar Marg, New Delhi, India, 110002.

Emails: laveesh@indicus.net, koel.roychowdhury@indicus.net

* Author to whom correspondence should be addressed; Tel.: +91-11-4251-2400; Fax: +91-11-2337-9466

Abstract

This paper investigates the association between night lights and GDP estimates for India at the district level. While many studies are finding a high degree of association between economic activity as measured through the Gross Domestic Product (GDP) and night lights internationally, there is a lack of understanding of whether and how night light data are correlated with economic activity at the sub-national level in emerging economies. This achieves more significance in economic monitoring and policy-making as estimates of GDP are not available at geographically disaggregated level, and even if available there is a large time lag involved before they are released. Stable light data obtained from night time images of 2008 captured by Defense Meteorological Satellite Program – Operational Linescan System (DMSP-OLS) satellite are used in the study. The data records artificial lights from human habitations from the earth surface and is a surrogate of the level of development of an area. The data on GDP at the district level for the year 2008 have been sourced from Indicus Analytics that has used data from government sources and a method of estimation suggested by the Central Statistical Office of the Government of India. Using multinomial non-linear regression techniques the paper finds that indeed GDP at the district level is significantly explained by night lights in the area. It also finds that the non-linearity is much stronger for metropolitan cities where GDP levels are far higher than a linear model can explain. Conversely, in areas where agriculture and forestry activities are higher, the use of night lights in a linear model overestimates the GDP.

Keywords: DMSP-OLS Stable Lights; District GDP; Sectoral GDP

1. Introduction:

A country's Gross Domestic Product (GDP) is the value of all goods and services produced within a country in a given year. It is arguably the most widely used measure of economic activity whose progress is closely followed by governments and researchers alike [1]. Following GDP growth rates across sectors and zones helps policy-makers assess the changes taking place in the economy and to formulate policies accordingly. It also helps individuals such as investors and employment-seekers form expectations and make decisions. Although it is not a perfect measure of the economic performance of any economy, it's easy interpretation and comparability across regions and time lends it great usefulness as an indicator of the measure of economic activity.

Night time lights data have been used to study economic activities for the last two decades. Elvidge et al [3] first studied the correlation between night time lights and economic activities. Global night time datasets was specifically used by Doll et al [4] for producing a first ever map of GDP-PPP (Purchasing Power Parity) of the world at a resolution of 1 degree by 1 degree (geographical co-ordinates). Sutton [5] used DMSP-OLS data to study spatial patterns of both market and non market indicators. GDP at Purchasing Power Parity (PPP) prices was taken as the economic indicator while ecosystem service product (ESP), subtotal ecological-economic product (SEP), SEP per capita and percent ecosystem service product were taken as non-market indicators. The effectiveness of DMSP-OLS night time datasets to estimate GDP was further extended by Ebener et al [6]. Doll et al [7] created maps of coterminous United States and western Europe at 5 Km spatial resolution using night time radiance data and regional economic productivity data. Since different countries had unique relationships with their light usage based on their cultures, each country was considered separately in this study. The information being derived solely from light sources it applied better for developed countries where industry and service sector comprised 90% of the economy. In developing countries where agriculture was the predominant occupation, the map recorded agricultural activities from towns that emitted light and not from the agricultural fields which occupied a significant portion of landuse.

There are many ways through which night-time lights can help better understand economic activity. There is, of course, it has great potential in better understanding the spatial nature and patterns of economic activity. But that pre-supposes a relationship between lights and economic activity. The first question is – how strong may this relationship be as we delve into finer and finer levels of granularity? The second is – does this relationship hold as strongly for developing countries as it does for developed ones? The third – given the higher share of agriculture, the use of old technologies in manufacturing and transport, and the low utilization of either electricity or fossil fuels in developing countries– what character does this relationship take?

But this is not all. There is another very important potential use of night light imagery specifically and remote sensing imagery in general, especially in the context of developing countries. GDP estimation in developing countries is difficult for many reasons. Data on specific activities is difficult to obtain as many activities are outside the domain of organized sector and many if not most transactions are not recorded in a consistent manner. Even where available there are questions about the veracity of the data obtained by the government. Businesses and individuals may not report the actual scale of their activities. Moreover, even where data are accessed there is a large time gap between the time the economic activity is undertaken, and it is recorded and made available to economic accounting agencies. In addition a range of estimations and interpolations need to be undertaken which further increases the time gap between the economic activity occurring and the GDP estimate being made available. Consequently governments such as in India have to resort to artifices such as quick estimates, provisional estimates, first estimates etc. before a final estimate is made – in some cases years elapse after the activities under question were undertaken. Night lights and remote sensing data have the potential of being available to government economic accounting agencies within a relatively short span of time with little need to generate improved estimates over a period of time. In other words, night lights can potentially become an integral part of the GDP estimation process in developing countries.

This paper uses the DMSP-OLS night time images of 2008. The stable light product for 2008 is used to study the correlations between GDP and the sum of lights for India at the district level (there being 593 districts in 2008). This paper differs from others in many ways.

One, it studies the link between nighttime lights and GDP at a far more granular level than most other studies. Further, to the knowledge of the authors, it is the first such study on India and among the few that have focused on an emerging market/developing country.

Two, the paper separately studies the relationship for *different kinds of economic activities* – as measured by the primary, secondary and tertiary sectors of the economy. The primary sector of the economy includes the production of raw material and basic foods and includes agriculture and livestock related activities as its most important component. The primary sector also includes mining and quarrying, forestry, farming, grazing, hunting and gathering and fishing. The secondary sector includes all of manufacturing, processing, and construction. This includes activities such as metal working and smelting, automobile production, textile production, chemical and engineering industries, aerospace manufacturing, energy utilities, engineering, breweries and bottlers, construction, and shipbuilding. The tertiary sector comprises of largely the service industry. This sector provides services to the general population and to businesses. Activities associated with this sector include retail and wholesale sales, transportation and distribution, Information technology, entertainment, restaurants, media, tourism, insurance, banking, healthcare, etc. The major urban centers of the country have higher share of their GDP coming from the tertiary sector. However, there are some parts of the country with areas having mixed sectoral contribution with income coming from both the secondary and tertiary sectors. About three fifths of the workforce in India is involved in tertiary sector activities and this ratio is rising. The bulk of the tertiary sector, despite the rapid growth of the IT and financial sectors remains the unorganized sector.

Three, the effect of specific regions such as the metropolitan regions, large towns and capital cities are considered separately in this study. This is because it is evident that the relationship between night lights and economic activity may be different in larger urban agglomerations than in smaller ones.

The rest of the paper proceeds as follows. Section 2 discusses the various data sources. Section 3 develops the model, tests the hypotheses and reports the results. Section 4 investigates the error terms and finds insights into modeling and data issues that impact the predictive nature of the model and Section 5 concludes.

2. Data and Methodology

2.1. Overview

The data on district level GDP at factor cost for the year 2008 was provided by Indicus Analytics (IA). There are 35 states and Union Territories (UTs) in India. GDP estimates for each of these are made available by the government of India through its agency the Central Statistical Organization (CSO). Each of these states and UTs comprise of districts – there were 593 districts in India in 2001 whose boundaries are available from maps made public by the Census of India. IA used a refined version of a methodology to estimate district level GDP that was devised by the CSO, though government estimates for 2008 are not yet available.

The satellite image used in this paper was captured at night by the Operational Linescan System (OLS) sensor onboard the Defense Meteorological Satellite Program (DMSP) group of satellites. The stable light data was obtained from National Geophysical Data Centre (NGDC) website using the latest average DN data series [8]. It contains the lights from cities, towns, and other sites with persistent lighting, including gas flares. Ephemeral events, such as fires have been discarded from this dataset. Data values range from 1-63, with background noise data replaced with a zero. Areas with zero cloud-free observations are represented by the value 255.

To assess the relationship between GDP and night-lights, correlations between various GDP measures (overall, primary, secondary and tertiary sectors) were first observed – high degree of correlations indicated further analysis may be fruitful. A set of multivariate regression models were then tested.

2.2. Satellite image processing

The nighttime lights imagery of India used in this study was carved out from the global stable lights imagery of 2008 collected by the Operational Linescan System (OLS) sensor onboard satellite F16 of the Defense Meteorological Satellite Program (DMSP). The satellite imagery was obtained from the nighttime lights data repository at the National Geophysical Data Center (NGDC), National Oceanic and Atmospheric Administration (NOAA) [8]. The stable lights image contains lights from cities, towns, and other sites with persistent lighting. The ephemeral lights from fires, fishing boats, and other sources are removed. The pixels of the stable lights data, at approximately 1km2 resolution at the equator, show brightness values in Digital Numbers (DNs). Because of 6-bit quantization the DN values range from 0 to 63 [10]. The sum of all lit pixels within each district's boundary for the whole of India were calculated and termed as sum of lights and used in this analysis (Figure 1). All districts were covered in the analysis.

As the OLS sensor does not have on-board calibration, an empirical procedure has been developed to inter-calibrate the stable lights images. Coefficients derived from this empirical analysis were used to calibrate the stable lights image of F162008. Lights from gas flares were masked out using the mask of global gas flares created by Elvidge, so that they are not incorrectly considered as lights [9].

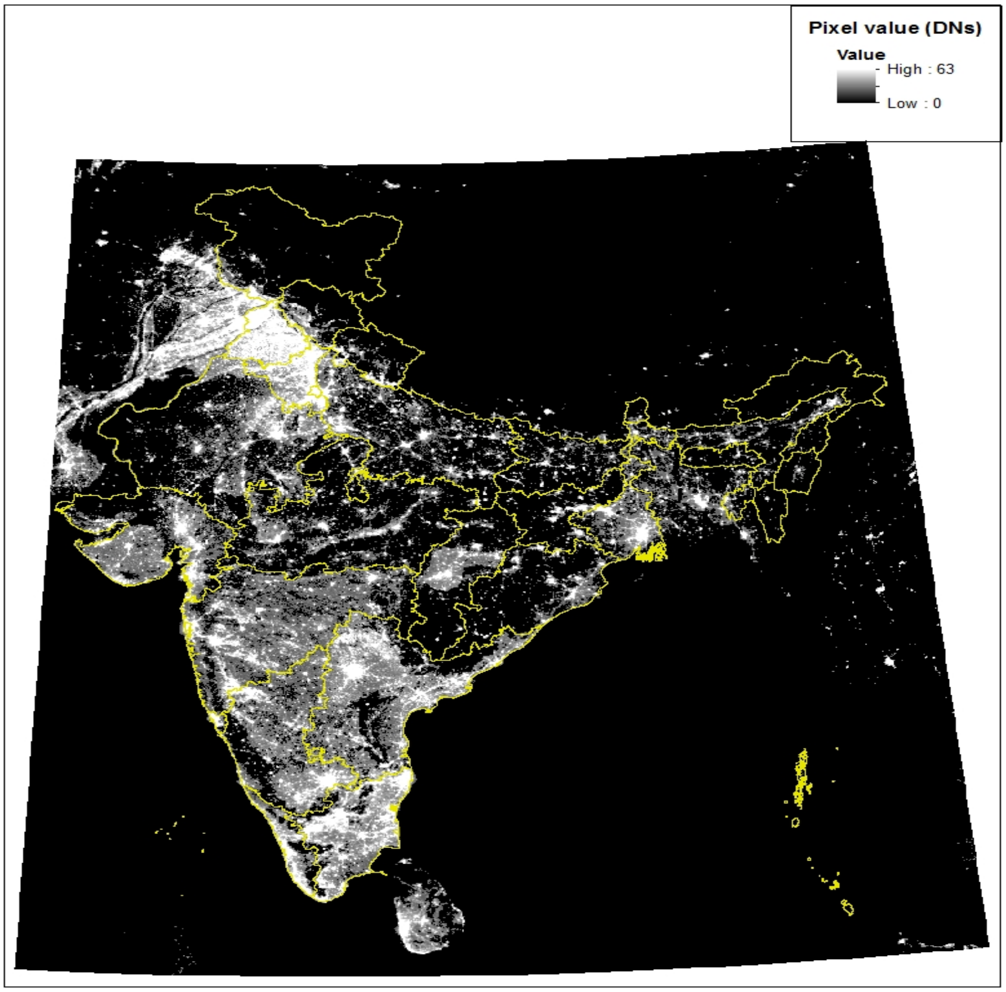


Figure 1. Stable Lights Image of India, 2008.

2.3. GDP data processing

The district level GDP data from IA was first published in 2006 and is currently updated annually. This data provides sectorwise estimates of GDP on twelve distinct sources of the economy for each district of the country. The data is collated at the district and sub district level from information obtained from other government data sources. These include Ministry of Agriculture (data on agriculture); Ministry of Environment (data on Forestry); Ministry of Animal Husbandry (data on Fishing); Indian Bureau of Mines (data on Mining and Quarrying); National Sample Survey Organization and Economic Census (data on Manufacturing); Ministry of Rural Development (data on Construction, electricity, gas and water supply), Census of India (demography) and Reserve Bank of India (financial sector activity). In addition to these, primary survey data on household characteristics, family structure, income, expenditure and saving practices are also analyzed and validated to obtain an insight into the GDP at district level.

The GDP at the district level is also sometimes termed as District Domestic Product (DDP); the DDP for each sector is estimated by distributing the state level GDP of the sector into each district. Depending on the availability of the data for the sector, the distribution is based on a method of estimating the district level value of production and creating two indices using principle inputs and outputs. The first index is based on production function and is of the form *Y (K, L, M)* where Y is the output, K is the capital, L is labour and M is a general variable related to land and natural resources. The second index is an additive index based on normalized sectoral data. The estimates of these variables are cross – checked with heuristics and secondary sources.

3. Results

3.1. Correlation Analysis

GDP at the district level for India was correlated with the sum of lights. It was found that there was very low linear correlation between GDP and the sum of lights as obtained from DMSP-OLS images. On further examination a strong logarithmic relationship was noted between natural log of DDP and natural log of sum of lights. The correlation coefficient (r) is the highest for correlation with the total GDP (r = 0.87) while it varied from 0.73 to 0.87 in correlations with the log of sectoral GDPs. Both outliers and clustering are noted in the correlations.

Districts such as Mumbai, Delhi, Chennai, Thane and Pune are noted as outliers in the correlation between the sum of lights and total GDP (Figure 2). These districts contain either metro cities (Mumbai, Delhi, Chennai, Pune) or cities that are satellites of metros (Thane for example).

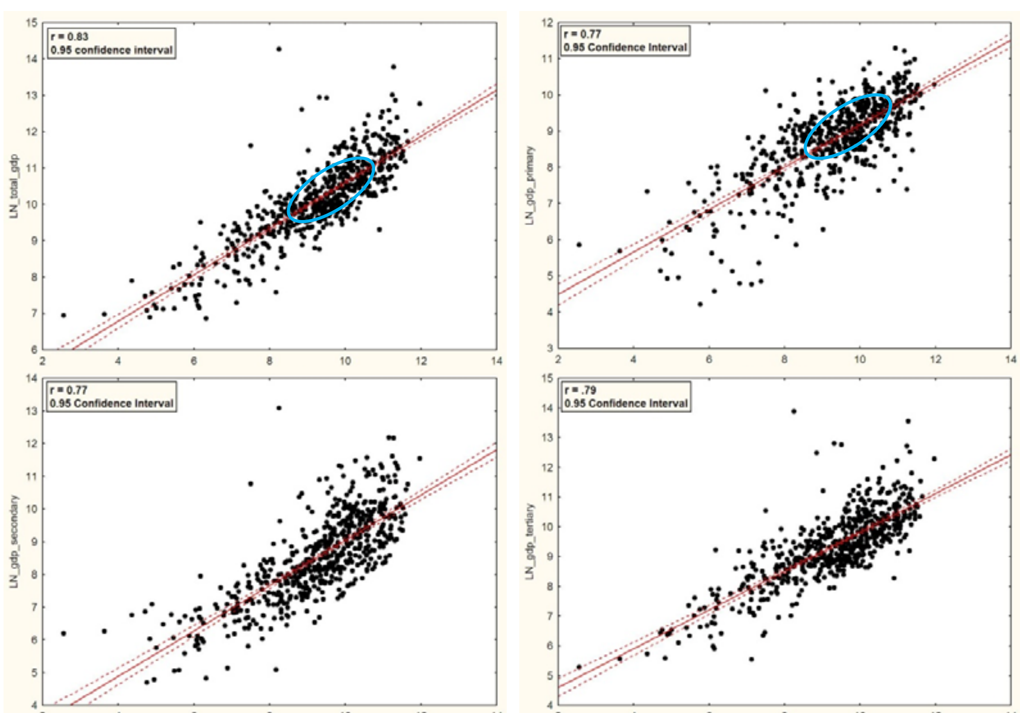
A clustering was noted between 8 to 12 log sum of lights (highlighted in blue circle in figure 2a). These districts include urban areas such as Chennai and Kolkata as well as rural districts of Harda, Barwai (Madhya Pradesh) and Dhamtari (Chhattisgarh). This shows that for more than 50% of the districts of the country moderate sum of lights was recorded by the DMSP-OLS dataset that had a logarithmic value between 8 and 12. These districts also accrue medium to high GDP in total. In order to look into the detailed effects of lights on different sectors of the economy (primary, secondary and tertiary) correlations were calculated between natural log of sum of lights and the natural log of GDP obtained from these sectors (Figure 2).

The correlations between the log sum of lights and the sectoral GDP (primary, secondary and tertiary sectors) are shown in Figure 2. About 54% of the districts are clustered between 7 to 11.5 log sum of lights and 8.5 to 10 log GDP from primary sector. It includes very bright districts such as Rangareddi in Andhra Pradesh, Raigarh and Nagpur in Maharashtra and Delhi to some darker districts such as Garhwa and Chatra in Jharkhand, Munger and Banka in Bihar and Anantnag in Jammu and Kashmir. These districts have an average of 30% to 65% of their GDP contributed by the primary sector.

In the correlation between log GDP secondary sector and log sum of lights, Mumbai and Aurangabad districts of Maharashtra form the outliers. Mumbai and Delhi are noted as outliers in the correlation between the log of sum of lights and log GDP tertiary sector. Tertiary sector contribute to 66.5% and 78% of the GDP respectively for these two districts.

The bulk of the positive outliers were found to be larger metros, districts in their vicinity, or those containing larger cities. This suggests that standard linear or logarithmic models are not adequate to capture the variation in GDP of larger urban agglomerations.

Figure 2: Correlations between log GDP and log Sum of Lights. (a) Correlation between log total GDP and log sum of lights; (b) Correlation between log GDP primary sector and log sum of lights; (c) Correlation between log GDP secondary sector and log sum of lights; (d) Correlation between log GDP tertiary sector and log sum of lights



Moreover, Figure 2 also shows that there is an upper limit to night-lights (stable lights product) which builds in a level of non-linearity that would need to be better explained. Since larger metros tend to have higher populations (such as Mumbai and Kolkata), and also have a far greater level of economic activity due to better economic and physical infrastructure, the modeling exercise aimed at studying each of these factors along with night lights in explaining the GDP of a district needs to be different for these locations.

3.2. Model for Regression Analysis

The models developed for assessing the relationship between lights and GDP at the district level were developed using insights from the correlations observed in the earlier section. The results showed that while lights do predict GDP to a great extent for most of the districts, there are other factors too that may be playing a role in this interaction that night lights are not able to adequately capture. The discussion below motivates the model.

Types of urban areas: As mentioned before correlation analysis showed that larger cities (e.g. metropolitan areas) tended to have higher level of GDP than a linear night lights model can explain. Consequently we introduced dummy variables that take the value 1 when that district contains a metropolitan area and 0 if it does not (DMetro). Next another dummy variable takes the value 1 when the district contains satellite/suburban areas to metropolitan cities and 0 if not (DSubmetro). Capital cities of the various states may have a different economic structure and are identified by a dummy variable as well (DCapital). Last, there are certain large cities that are not classified as metropolitan but have a high degree of economic activity (DLargeCity). Whether they behave differently is tested by another dummy variable that takes the value 1 if it is such a city and 0 if not. The coefficients for these dummy variables are expected to have a positive value –the ordering being – highest for metros, satellite cities, capital cities and lowest for other large cities.

Table 1: Different types of districts

DMetro	***DSubmetro***	***DCapital***	
1. Ahmadabad (Gujarat)	1. Chandigarh (Chandigarh)	1. Aizawl (Mizoram)	18. Kamrup (Assam)
2. Bangalore (Karnataka)	2. Faridabad (Haryana)	2. Andamans (Andaman & Nicobar Islands)	19. Khordha (Orissa)
3. Chennai (Tamil Nadu)	3. Gautam Buddha Nagar (Uttar Pradesh)	3. Bangalore (Karnataka)	20. Kohima (Nagaland)
4. Delhi (Delhi)	4. Ghaziabad (Uttar	4. Bhopal (Madhya Pradesh)	21. Kolkata (West

	Pradesh)		Bengal)
5. Hyderabad (Andhra Pradesh)	5. Gurgaon (Haryana)	5. Chandigarh (Chandigarh)	22. Lakshadweep (Lakshadweep)
6. Kolkata (West Bengal)	6. Hugli (West Bengal)	6. Chennai (Tamil Nadu)	23. Lucknow (Uttar Pradesh)
7. Mumbai (Maharashtra)	7. Ludhiana (Punjab)	7. Dadra & Nagar Haveli (Dadra & Nagar Haveli)	24. Mumbai (Maharashtra)
8. Pune (Maharashtra)	8. Nagpur (Maharashtra)	8. Daman (Daman & Diu)	25. North Goa (Goa)
	9. North 24 Parganas (West Bengal)	9. Dehradun (Uttaranchal)	26. Papum Pare (Arunachal Pradesh)
	10. South 24 Parganas (West Bengal)	10. Delhi (Delhi)	27. Patna (Bihar)
	11. Surat (Gujarat)	11. East Khasi Hills (Meghalaya)	28. Pondicherry (Pondicherry)
	12. Thane (Maharashtra)	12. East Sikkim (Sikkim)	29. Raipur (Chhattisgarh)
	13. Visakhapatnam (Andhra Pradesh)	13. Gandhinagar (Gujarat)	30. Ranchi (Jharkhand)
		14. Hyderabad (Andhra Pradesh)	31. Shimla (Himachal Pradesh)
		15. Imphal West (Manipur)	32. Srinagar (Jammu & Kashmir)
		16. Jaipur (Rajasthan)	33. Thiruvananthapuram (Kerala)
		17. Jammu (Jammu & Kashmir)	34. West Tripura (Tripura)

DLargeCity			
1. Agra (Uttar Pradesh)	17. Dhanbad (Jharkhand)	33. Kanniyakumari (Tamil Nadu)	49. Panipat (Haryana)
2. Ajmer (Rajasthan)	18. Durg (Chhattisgarh)	34. Kannur (Kerala)	50. Patiala (Punjab)
3. Allahabad (Uttar Pradesh)	19. Ernakulam (Kerala)	35. Kanpur Nagar (Uttar Pradesh)	51. Purbi Singhbhum (Jharkhand)
4. Ambala (Haryana)	20. Gulbarga (Karnataka)	36. Kolhapur (Maharashtra)	52. Rajkot (Gujarat)
5. Amritsar (Punjab)	21. Guntur (Andhra Pradesh)	37. Kota (Rajasthan)	53. Rangareddi (Andhra Pradesh)
6. Anantapur (Andhra Pradesh)	22. Gwalior (Madhya Pradesh)	38. Kozhikode (Kerala)	54. Rupnagar (Punjab)
7. Barddhaman (West Bengal)	23. Haora (West Bengal)	39. Krishna (Andhra Pradesh)	55. Salem (Tamil Nadu)
8. Bareilly (Uttar Pradesh)	24. Hisar (Haryana)	40. Madurai (Tamil Nadu)	56. Solapur (Maharashtra)
9. Bathinda (Punjab)	25. Indore (Madhya Pradesh)	41. Medinipur (West Bengal)	57. Sundargarh (Orissa)
10. Belgaum (Karnataka)	26. Jabalpur (Madhya Pradesh)	42. Meerut (Uttar Pradesh)	58. Thanjavur (Tamil Nadu)
11. Bhavnagar (Gujarat)	27. Jalandhar (Punjab)	43. Moradabad (Uttar Pradesh)	59. Thiruvallur (Tamil Nadu)
12. Bikaner (Rajasthan)	28. Jalpaiguri (West Bengal)	44. Murshidabad (West Bengal)	60. Thrissur (Kerala)

13. Bokaro (Jharkhand)	29. Jamnagar (Gujarat)	45. Mysore (Karnataka)	61. Tiruchirappalli (Tamil Nadu)
14. Coimbatore (Tamil Nadu)	30. Jodhpur (Rajasthan)	46. Nanded (Maharashtra)	62. Udaipur (Rajasthan)
15. Cuttack (Orissa)	31. Kancheepuram (Tamil Nadu)	47. Nashik (Maharashtra)	63. Vadodara (Gujarat)
16. Dakshina Kannada (Karnataka)	32. Vellore (Tamil Nadu)	48. Nellore (Andhra Pradesh)	64. Varanasi (Uttar Pradesh)

Snow covered regions: The northernmost parts of India contain some pockets that tend to be snow covered for much of the year; snow covered regions tend to reflect greater light and consequently the night lights may show a higher value. The dummy variable for snow covered regions is therefore intended to capture this and is expected to have a negative coefficient. (DSnow)

Table 2: Snow covered districts

DSnow	
1. Papum Pare (Arunachal Pradesh)	21. West Sikkim (Sikkim)
2. Tawang (Arunachal Pradesh)	22. South Sikkim (Sikkim)
3. Sirmaur (Himachal Pradesh)	23. East Sikkim (Sikkim)
4. Mandi (Himachal Pradesh)	24. North Sikkim (Sikkim)
5. Shimla (Himachal Pradesh)	25. West Tripura (Tripura)
6. Kinnaur (Himachal Pradesh)	26. Dhalai (Tripura)
7. Kullu (Himachal Pradesh)	27. South Tripura (Tripura)
8. Chamba (Himachal Pradesh)	28. North Tripura (Tripura)
9. Kangra (Himachal Pradesh)	29. Almora (Uttaranchal)
10. Solan (Himachal Pradesh)	30. Dehradun (Uttaranchal)
11. Lahul & Spiti (Himachal Pradesh)	31. Bageshwar (Uttaranchal)
12. Leh (Ladakh) (Jammu & Kashmir)	32. Uttarkashi (Uttaranchal)
13. Kargil (Jammu & Kashmir)	33. Tehri Garhwal (Uttaranchal)
14. Pulwama (Jammu & Kashmir)	34. Champawat (Uttaranchal)
15. Baramula (Jammu & Kashmir)	35. Garhwal (Uttaranchal)
16. Udhampur (Jammu & Kashmir)	36. Chamoli (Uttaranchal)
17. Kupwara (Jammu & Kashmir)	37. Pithoragarh (Uttaranchal)
18. Srinagar (Jammu & Kashmir)	38. Nainital (Uttaranchal)
19. Doda (Jammu & Kashmir)	39. Rudraprayag (Uttaranchal)
20. Anantnag (Jammu & Kashmir)	40. Darjiling (West Bengal)

Sum of Lights: As mentioned higher night lights should be significantly related to higher GDP and consequently have a large positive coefficient. (SOL)

Population: Analysis of the outliers revealed that typically the very low population districts or those with high populations had high (positive or negative) outlier values. It is difficult to ascertain whether superior economic activity is attracting more people due to the enhanced opportunities, or whether higher interaction on account of more people is leading to the high concentration of economic activity; one may be reinforcing the other. But what is essential to note is the fact that wherever population densities are very high or very low sum of lights are unable to capture economic activity to the full extent. This could be for different reasons – high population concentration areas have high buildings and consequently night-lights capture a lower proportion of the total lights, or where populations are low, night-lights are not adequately captured by current in remote sensing technologies. (POP)

We therefore test the following:

GDP = f(SOL, POP, DMetro, DSubmetro, DCapital, DLargeCity, DSnow)

Following the literature we use a logarithmic specification i.e.

Ln GDP = Ln SOL + Ln POP +DMetro + DSubmetro + DCapital + DLargeCity + DSnow (1)

Table 3: Regression Results for District Level GDP: Coefficients

(Multivariate OLS; Dependent variable District Level GDP for 2008)

Predictor Variables	Model 1	Model 2	Model 3
Natural log of Sum of Lights	0.36*	0.34*	0.34*
Natural log of Population	0.52*	0.44*	0.45*
Dummy for Metropolitan Districts		1.64*	1.64*
Dummy for Suburbs of Metro cities		0.96*	0.96*
Dummy for Capital Districts		0.72*	0.71*
Dummy for Large Towns		0.53*	0.54*
Dummy for Snow-Covered Districts			0.09
Constant	-0.59	0.57	0.46
Adjusted R2	0.79	0.87	0.87

*All variables significant at 99% (except Dummy for snow-covered districts)

Table 3 shows the results of the regression. Overall we find that a very high degree of variation is explained by the model. However, lights and population together explain about four fifths of the variation in GDP. Addition of the various dummy variables does not impact the coefficients of sum of lights and population but they do improve the explanatory power of the model. Overall a percent increase in night-lights is associated with a 0.34 percent increase in GDP of a district, keeping population and district type constant.

The introduction of type of city dummy variables provides an interesting insight. First, larger cities are not explained as well by the sum of lights, and more importantly the larger the city the greater the GDP over and above that predicted by a purely lights and population model. This implies that there is some non-linearity in the relationship. Two broad possibilities exist: lights are not adequately captured by the night-light data, and/or lights themselves are relatively lower in these cities. Larger cities are generally marked by many buildings with greater number of floors and overall lower night-lights relative to the level of economic activity. While this can be tested, it also has important implications on our understanding of energy efficiency of large population concentrations. We believe that this would be an interesting question to tackle using night-lights data, however we do not pursue it in this study in the interest of tractability. The second possibility is that night-lights are not being adequately captured – this is indeed true since there is a cap on the lights that can be captured by the current remote sensing technologies.

The introduction of dummy variables for snow covered areas however does not have the expected impact in a statistically significant manner. The snow covered regions in India are in very sparsely populated terrain with low levels of economic activity – consequently it could be argued that there is little light to be reflected.

Overall however, we find that we are able to explain a high level of variation in GDP data. The next sets of regressions analyze the relationship between night-lights and components of the GDP – Primary, Secondary and Tertiary. See table 4.

Table 4: Regression Results for Components of District Level GDP: Co-effcients

(Multivariate OLS; Dependent variable Primary, Secondary and Tertiary Sector District Level GDP for 2008)

Predictor Variables	**Model 4: GDP from Primary Sector**	Model 5: GDP from Secondary Sector	Model 6: GDP from Tertiary Sector
Natural log of Sum of Lights	**0.30***	**0.50***	**0.30***
Natural log of Total Population	**0.55***	**0.22***	**0.55***
Dummy for Metro	**- 0.68***	**1.65***	**1.92***
Dummy for Suburbs of Metro cities	**- 0.42#**	**1.55***	**1.10***
Dummy for Capital	**- 0.57***	**0.97***	**0.98***
Dummy for Large Towns	**- 0.09**	**0.81***	**0.66***
Dummy for Snow-Covered Districts	**- 0.27***	**0.38***	**0.20#**
Constant	**- 1.81***	**0.54**	**- 1.34***
Adjusted R^2	**0.73**	**0.73**	**0.87**

*Significant at 99%, #Significant at 95%, $Significant at 90%,

First consider the adjusted r-squares that indicate that night-lights are better able to explain the variation in the aggregate than at the component level. A better understanding of the underlying economic relationships is therefore required to have better prediction for components of the GDP than GDP in the aggregate.

Next consider Model 4. The coefficients for the larger city dummies are negative – this is natural, since large cities are expected to have significantly lesser primary sector activity than smaller ones. Mining is typically not allowed in the vicinity of larger cities. Agriculture, though allowed is expected to be far lower due to higher land values.

Model 6, on tertiary sector, reveals coefficient values for night-lights that are not very different from those of primary sector. This, it could be argued, may be because a large proportion of the tertiary sector activity in India is in the unorganized sector – that uses traditional technologies, is less dependent upon power, and is highly labour intensive. At least where night-lights are concerned, the tertiary sector GDP appears to follow similar relationship as with the primary sector GDP.

Model 5, on the secondary sector however suggests a different relationship with a significantly higher coefficient value for night-lights. The manufacturing sector is the major sub-component of the secondary sector; and it is more capital intensive. GDP of the secondary sector is therefore found to be more sensitive to night-lights. Moreover, the coefficients of dummy variables for the various classes of cities are found to have the same ordering as expected for GDP in the aggregate. Note that there are certain metros where many kinds of manufacturing activity are not allowed (Delhi for example for polluting industries). Similarly, basic industry (e.g. iron and steel industry, cement manufacturing, etc.) tends to be in far flung areas and away from large cities. Hence the explanatory power of the city dummy variables though statistically significant is limited in the case of manufacturing activity.

Overall, therefore we find that the use of a simple model that allows for non-linearities in the form of different kinds of locations, and population is able to complement the explanatory power of night lights. Further understanding on improving the predictive power of such models would be possible by an analysis of the error terms which we undertake next.

4. Prediction errors

The models were used to predict GDP at the district level for the whole of India. The outlier districts for each model are shown in table 5.

Table 5a. Outliers in the regression for Total GDP

(Highest/lowest error terms)

Top 10 Positive outliers	**Top 10 Negative outliers**
Aurangabad (Maharashtra)	Kolasib (Mizoram)
Mumbai (Maharashtra)	Bilaspur (Himachal Pradesh)
Valsad (Gujarat)	Kaushambi (Uttar Pradesh)
Aizawl (Mizoram)	Harda (Madhya Pradesh)
Jamnagar (Gujarat)	Tikamgarh (Madhya Pradesh)
Dimapur (Nagaland)	Umaria (Madhya Pradesh)
Kottayam (Kerala)	Barwani (Madhya Pradesh)
Thiruvallur (Tamil Nadu)	Datia (Madhya Pradesh)
Panchkula (Haryana)	Bangalore (Karnataka)
Bharuch (Gujarat)	Papum Pare (Arunachal Pradesh)

The table above shows that the positive outliers – greater value of actual GDP than predicted is found in industrial cities such as Jamnagar and Thiruvallur. Aurangabad is a major tourist destination both for international and domestic tourists. The average correction for metropolitan cities is not adequate for Mumbai which continues to show up as a positive outlier. In other words, greater level of industrialization is not being adequately captured by night-lights for these areas. On the other end, negative outliers – greater predicted value than actual is found in areas where there is little industrial activity but a higher level of agriculture activity. In other words, night lights and population sometimes over-estimate the level of economic activity for districts (predominantly in the less developed states of Madhya Pradesh and Uttar Pradesh). A better understanding of these issues is possible if we analyze the outliers at the sectoral level.

Table 5b. Outliers in the regression for Primary Sector GDP
(Highest/lowest error terms)

Top 10 Positive outliers	Top 10 Negative outliers
Mumbai (Maharashtra)	Leh (Jammu & Kashmir)
Aurangabad (Maharashtra)	Bilaspur (Himachal Pradesh)
Kodagu (Karnataka)	Kolar (Karnataka)
Korba (Chhattisgarh)	Lucknow (Uttar Pradesh)
Dantewada (Chhattisgarh)	Bangalore (Karnataka)
Rohtas (Bihar)	Kargil (Jammu & Kashmir)
Dibang Valley (Arunachal Pradesh)	Kancheepuram (Tamil Nadu)
Dimapur (Nagaland)	Karaikal (Pondicherry)
Bathinda (Punjab)	Gaya (Bihar)
Kohima (Nagaland)	Yanam (Pondicherry)

Table 5c. Outliers in the regression for Secondary Sector GDP
(Highest/lowest error terms)

Top 10 Positive outliers	Top 10 Negative outliers
Aurangabad (Maharashtra)	Kolasib (Mizoram)
Valsad (Gujarat)	The Dangs (Gujarat)
Upper Siang (Arunachal Pradesh)	Madhepura (Bihar)
Mumbai (Maharashtra)	Kawardha (Chhattisgarh)
Jamnagar (Gujarat)	Harda (Madhya Pradesh)
Begusarai (Bihar)	Gulbarga (Karnataka)
Bharuch (Gujarat)	Kokrajhar (Assam)
Bongaigaon (Assam)	Sheopur (Madhya Pradesh)
Bokaro (Jharkhand)	Nanded(Maharashtra)
Thiruvallur (Tamil Nadu)	Chamarajanagar (Maharashtra)

Table 5d. Outliers in the regression for Tertiary Sector GDP

(Highest/lowest error terms)

Top 10 Positive outliers	Top 10 Negative outliers
Aizawl (Mizoram)	Bilaspur (Himachal Pradesh)
Aurangabad (Maharashtra)	Senapati (Manipur)
Mumbai (Maharashtra)	Yanam (Pondicherry)
Panchkula (Haryana)	West Sikkim (Sikkim)
Pathanamthitta (Kerala)	Kolasib (Mizoram)
Kottayam (Kerala)	Raigarh (Chhattisgarh)
Malappuram (Kerala)	Jammu (Jammu & Kashmir)
Kollam (Kerala)	Kaushambi (Uttar Pradesh)
Dimapur (Nagaland)	Shravasti (Uttar Pradesh)
South Goa (Goa)	Umaria (Madhya Pradesh)

Tables 5b to 5d reveal that outlier districts for one sector also tend to be outliers for another. In other words there are certain aspects of GDP that are not explained adequately by the model. There are many possibilities – presence of cantonments (for internal policing or defense) are one possibility; or there may be other kinds of activities that do not generate as much in value added (GDP) as lights would indicate. At the same time, there are many kinds of economic activities that are just not captured by any official data – smuggling is one such activity that leads to a high level of economic value added but is not captured in the GDP estimates. The consequent impact on incomes and expenditures however may show up in night-lights. Tourism is another activity that requires significant generation of night-lights with less than proportional value added.

We also find that the error terms tend to be far higher at the two ends of the economic spectrum – districts with very high GDP as well as those with low GDP and this was also observed in the correlation diagrams. This suggests that the logarithmic specification may not be adequate. A sigmoid specification may yield better predictive power.

3. Conclusion

This paper shows that the information obtained from the night time DMSP-OLS images can (with some improvements) predict GDP at the district level for developing countries such as India. It also finds certain patterns in the error terms that suggest that even higher predictive power is possible with further refinements in the specification as well as inclusion of variables. The results also show that GDP in the aggregate is better predicted than its components; however, this may change if a more sector specific modeling approach is used.

Improvements will need to take three directions. First, night-time imagery that better captures night lights is likely to yield significantly better estimates. With a better capturing of various kinds of night-light radiation across the spectrum, high outliers such as metro cities are expected to become less so. Second, the interaction between population, night lights and economic activity needs to be better understood. Third, most interest in GDP estimates is to adequately capture growth rates. The availability of radiance calibrated night-time imagery for more recent years should help; but if not available, the analysis of historical stable lights that are already in the public domain should reveal greater insights.

Overall we find that despite having a very simple model with limited informational requirements, we are able to explain a large part of the variation in district level GDP. Night-lights data therefore should be considered seriously by policy-makers as another input into better understanding GDP at highly disaggregated levels in India as well as other developing countries.

Acknowledgements

We would like to thank NOAA for making public the night lights data, to C.D. Elvidge for help in using the data, to Tilottama Ghosh and Mayuri Chaturvedi for their most important insights, criticisms and overall help. All errors are ours.

[1] J. Black*, et al.* (2009, 13 August 2011). *Gross Domestic Product*. Available: http://www.oxfordreference.com.ezproxy.lib.rmit.edu.au/views/ENTRY.html?subview=Main&entry=t19.e1403

[2] Indicus Analytics Private Limited, "District Gross Domestic Product series 2001 - 2007," ed, 2009.

[3] C. D. Elvidge*, et al.*, "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption," *International Journal of Remote Sensing,* vol. 18, pp. 1373 - 1379, 1997.

[4] C. N. H. Doll*, et al.*, "Night-time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions," *AMBIO: A Journal of the Human Environment,* vol. 29, pp. 157-162, May 01, 2000 2000.

[5] P. C. Sutton and R. Costanza, "Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation," *Ecological Economics,* vol. 41, pp. 509-527, 2002.

[6] S. Ebener*, et al.*, "From wealth to health: modeling the distribution of income per capital at the sub-national level using nighttime light imager," *International Journal of Health Geographics,* vol. 4, p. 5, 2005.

[7] C. N. H. Doll*, et al.*, "Mapping regional economic activity from night-time light satellite imagery," *Ecological Economics,* vol. 57, pp. 75-92, 2006.

[8] National Geophysical Data Centre. (2007, 20th April). *Defense Meteorological Satellite Program (DMSP) Data Archive, Research, and Products*. Available: http://www.ngdc.noaa.gov/dmsp/dmsp.html

[9] C. Elvidge*, et al.*, "A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data," *Energies,* vol. 2, pp. 595-622, 2009.

[10] Baugh, K.E.; Elvidge, C.D.; Ghosh, T.; Ziskin, D. Development of a 2009 Stable Lights Product Using DMSP-OLS Data. *Proceedings of the Asia Pacific Advanced Network*. Hanoi, Vietnam, 2010.