Proceedings of the Asia-Pacific Advanced Network ***2011*** *v. 32, p. 192-217.*
http://dx.doi.org/10.7125/APAN.32.23
ISSN 2227-3026

Assessing Income Distribution at the District Level for India Using Nighttime Satellite Imagery

Mayuri Chaturvedi [1,*], Tilottama Ghosh [1,2] and Laveesh Bhandari [1]

[1] Indicus Analytics Pvt. Ltd., 2nd Floor, Nehru House, 4, Bahadur Shah Zafar Marg, New Delhi – 110002.

[2] Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, 216 UCB, Boulder, CO 80309, USA.

E-Mails: mayuri.chaturvedi@indicus.net; tilottama.ghosh@noaa.gov; laveesh@indicus.net

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

Abstract: Several studies have been carried out relating nighttime lights with economic activity. But most studies relating nighttime lights with economic activity have focused on associating higher *totals* in economic activity with higher sum of lights across regions. The question addressed in this paper is how best to model the relationship of nighttime lights with not just the wealthy but also the relatively worse-off *within* a region. The implications of such an exercise are immense with respect to ascertaining income distribution aspects of any area. The methods developed in this paper explore the relation between households in different income brackets at the district level for India, and the radiance-calibrated nighttime image of 2004. Besides the radiance-calibrated data of 2004, estimates of household incomes and number of households in different income brackets, made by Indicus Analytics (specialized economic research firm, based in New Delhi, India) were used. The results were mapped and insights were drawn for all districts based on their socio-economic profile. These results illustrate the advantage of using this easily available data for determining income inequalities, especially in information-deficient countries such as India.

Keywords: Radiance-calibrated nighttime images; income distributions; income inequalities

1. Introduction

Income inequality, that is, the unequal proportion in the distribution of income, is a matter of policy concern in many countries. In India, for instance, while the top 10% of India's population enjoys 31.1% of the country's income, the lowest 10% has only 3.6% of the total income share [1]. The Government of India (GoI) has been concerned about rising inequalities and uneven distribution of the benefits of growth. Accordingly, the thrust of the 11th Five-Year Plan (2007-12) was on inclusive growth, allowing all groups of people to contribute to and benefit from economic growth. The forthcoming 12th Five-Year Plan is expected to deepen and sharpen the focus on inequalities [2].

While socio-economic applications abound, the availability of reliable and frequent data on economic growth and development is rare and is characterized by several shortcomings. There are issues of under-reporting, over-reporting, misreporting, inappropriate sampling (when the sample is not representative of the population), and inappropriate weighting (when inaccurate proportions are placed on certain segments of the sample to get the population averages) associated with data collected through various national sample surveys. Other difficulties include, the lack of standardized national income accounting methods and reliable methods of data collection, inefficiency of the surveyors, and the subjective response of the respondents. At times, the political and economic situation in an area also inhibits data collection [3, 4, 5]. Because of the enormous expenses involved in conducting socio-economic surveys, the frequency of these surveys that are nationally representative is less, usually at an interval of five or ten years. Moreover, there is always a time lag between the time when data are collected and when they are finally published. Furthermore, data collected through these surveys are usually done at the country level or the state level, and not at the sub-state level, or at finer spatial resolutions and suffer from MAUP (Modifiable Areal Unit Problem) [6, 7].

In India the two important sources of consumption data (used for estimating incomes) are the socio-economic data collected by the National Sample Survey Organization (NSSO) quinquennially at the district level [8], and by the National Accounts Statistics (NAS) annually at the pan-India level [9]. The latest data of NSSO on consumer expenditure, which comprises, level and pattern of household consumer expenditure, nutritional intake, commodity-wise consumption, adequacy of food, differences in level of consumption among socio-economic groups, use of durable goods, etc., were collected for 2009-2010. The latest NAS data on private consumption expenditure and domestic savings were also collected for 2009-2010.

Some of these drawbacks of available sample-survey data can be circumvented using remote sensing data that are accessible at higher spatial and temporal resolution. Remote sensing data in the form of nighttime satellite imagery shows great potential in this respect. The nighttime satellite imagery has been used for several socio-economic studies like mapping urban extent and its ecological impact [10, 11, 12], estimating urban populations and intra-urban population density [13, 14, 15], estimating impervious surface area [16], mapping greenhouse gas emissions [17, 18, 19], mapping fire and fire prone areas [20], estimating and mapping GDP at the national and sub-national levels [5, 6, 21, 22, 23, 24, 25], and creating global grids of GDP and poverty [26]. However, the relationship between nighttime imagery and income distributions has not been explored before to the knowledge of the authors.

In this paper, we developed a method to investigate the relation between households in different income brackets at the district level for India, and the nighttime image. The objective of this study was to examine how good a proxy of income distribution is the nighttime lights image at finer spatial resolutions. The paper also tried to identify the regions for which this method does not give close approximations, and the possible underlying socio-economic factors behind such results. The method developed for examining the relationship between nighttime lights and households in different income distributions is discussed in the next section.

2. Methods

The different datasets used for examining the relationship between nighttime lights and income distribution at the district level for India are as follows:

2.1. Nighttime Lights Imagery

The nighttime lights image was used to calculate the sum of light intensity values at the district and state level for India. A blended stable lights and fixed gain image of 2004 collected by satellite F14 was used for this purpose. Since 1994, the National Geophysical Data Center (NGDC) of the National Oceanic and Atmospheric Administration (NOAA) has been archiving and processing data collected by the Operational Linescan System (OLS) sensor flown on the Defense Meteorological Satellite Program (DMSP) satellites. The stable lights nighttime image is a global annual cloud free composite produced by averaging the visible band data of the cloud-free segments of individual orbits. In the stable lights product the ephemeral lights from fires and other sources are removed on the basis of their high brightness and short duration [27]. The operational data are collected in a high gain setting to make possible the detection of moonlit clouds. However, the six bit quantization and the limited dynamic range causes saturation in the urban centers. But it also detects dim lighting in the populated rural and suburban areas. The second product, the fixed gain product, is generated by merging data collected in low, medium, and high fixed gain settings. The fixed gain data have unsaturated values in the bright cores of urban centers but are unable to detect the dimmer lights. Merging the stable lights and fixed gain product, taking only the unsaturated values, facilitates the creation of the superior radiance-calibrated images, which helps to overcome the limited dynamic range of the operational stable lights data [28]. The spatial resolution of the nighttime raster image (grid) is about 1 km^2 at the equator. The radiance-calibrated image of 2004 was used in this study. Bounding latitudes and longitudes (38°N, 68°W, 5°S, and 98°E) were used to 'carve out' India from the global nighttime image of 2004. The image has a dynamic range of Digital Numbers (DNs) from 0 to 4550. The radiance-calibrated image of India was re-projected from geographic coordinates to Albers conic equal area projection.

A mask of global gas flares was created by Elvidge in a previous study [29]. Since gas flares are bright point sources of light with no shielding to the sky, they form circular lighting features having bright centers and wide rims in the nighttime images. They need to be masked out so that they don't get erroneously detected as nighttime lights from human settlements. The same latitudinal and longitudinal extent that was used to 'carve out' India was used to delineate the mask extent for India, and this mask was then applied on the nighttime image of India. This helped to mask out the on-shore gas flares in the north-eastern states of Assam, Arunachal Pradesh, and Nagaland (Figure 1).

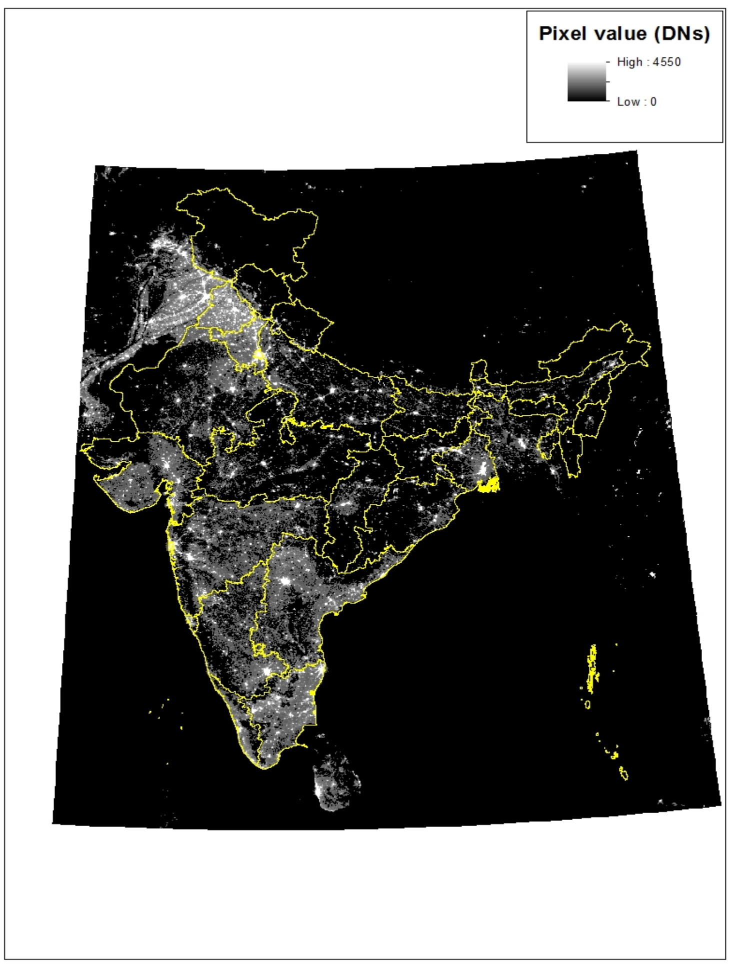


Figure 1. Radiance-calibrated nighttime image of India, 2004

2.2. Data on Number of Households in Different Income Categories

Data on number of households in different income categories for the year 2004 were obtained from Indicus Analytics' data repository. These data were used in this study to examine their relationship with nighttime lights image through regression analyses. Indicus Analytics, a specialized economic research firm, based in New Delhi, uses intensive analytics to create reliable data sets on household level at a highly granular level. Indicus uses both secondary and primary survey data from various data sources, which are accessed, logged and processed to produce refined data. These include the NSSO surveys, National Data Survey of Savings Patterns of Indians (NDSPPI), District Level Household Survey, Census of India, RBI datasets, etc. These surveys provide detailed insights of the expenditure and savings habits of households, the household characteristics and their family structure, and their earnings and occupations.

Expenditure data of households are either under-reported by the better-off segments or over-reported by the poorest households. Indicus uses a combination of techniques to obtain correct estimates. At first sums of all products level expenditures undertaken by households on a range of product groups that are available from the NSSO are obtained. Next, how much each of these goods and services is over or under-reported for every income group and for every product segment is calculated. This facilitates adjustment for the under-reporting for every income group and for every product segment and also ensures that the estimates are in accord with those of the NAS.

Data on savings potential of households is very difficult to obtain, especially those which are obtained from generic consumer expenditure surveys. The Ministry of Finance has undertaken household surveys aimed particularly at determining the savings behavior of Indian households. Through a sample survey of 40 thousand Indian households in various demographic and income groups, the savings potential of different households were examined, which were then calibrated with data on aggregate household savings available from NAS to acquire actual savings at the household level [30].

Annual household income data, which is an aggregate of expenditure and savings data, of number of households in six income brackets for the year 2004 were obtained from Indicus' data repository. The six income brackets are -

1. < Rs. 75,000 (or, < Rs.0.75 lakh)
2. Rs.75,001 - Rs.150,000 (or, Rs.0.75 lakh – Rs.1.5 lakh)

3. Rs.150,001 - Rs.300,000 (or, Rs.1.5 lakh – Rs.3 lakh)
4. Rs.300,001 - Rs. 500,000 (or, Rs.3 lakh – Rs.5 lakh)
5. Rs.500,001 - Rs.1,000,000 (or, Rs.5 lakh – Rs.10 lakh)
6. > Rs.1,000,000 (or, > Rs.10 lakh)

The numbers of households in the first three income brackets (< Rs.0.75 lakh, Rs.0.75 lakh – Rs.1.5 lakh, Rs.1.5 lakh – Rs.3 lakh) were added up and classified as population belonging to the '*lower'* income group. The next two income brackets (Rs.3 lakh – Rs.5 lakh, Rs.5 lakh – Rs.10 lakh) were added up and classified as '*middle'* income group, and > Rs.10 lakh was classified as the '*upper'* income group. These income categories were organized this way for the purpose of tractability in our analysis.

2.3. State and District Shapefiles for India

Vector boundaries in the form of shapefiles for the 593 districts of India (according to the 2001 census) were obtained from Indicus' data archive. The district vector data were indigenously digitized by personnel at Indicus. The districts were 'dissolved' to get the 28 states and 7 Union Territories (UTs) of India. The shapefiles were re-projected from geographic coordinates to Albers Equal Area conic projection. The district and state shapefiles were then overlaid on the radiance calibrated nighttime lights image of India and the district and state level 'sum of lights' data were extracted. The sum of lights is the sum of the DNs within every district or state and UT boundary of India.

2.4. Preliminary Analysis

The states and UT vector shapefile of India was overlaid on the radiance-calibrated image of 2004 and the sum of lights were extracted. Data on the sum of lights for the states of India was plotted against the number of households in each of the defined income categories- upper, middle and lower. A look at the scatter plot for sum of lights and households in the upper income group gives an interesting picture of how well lights can capture the households' income data across Indian states. The figures below (Figure 2a, 2b and 2c) provide the scatter plots for the 20 biggest states of India with sum of lights on the X-axis and number of households in the lower, middle and upper income groups on the Y-axis. The states with the higher total GDP are captured on the top-right in the scatter (for example Maharashtra, Andhra Pradesh, Uttar Pradesh, Tamil Nadu), and the states with lower GDP are seen in the bottom left (for example Jammu and Kashmir, Bihar, Chhattisgarh). Other middle income states lie between these two extremes above or below the line of best fit in the three graphs, depending on the relative distribution of income in these states. These graphs effectively portray that other than a few outliers above and below the line of best fit, lights can be used to capture income levels of

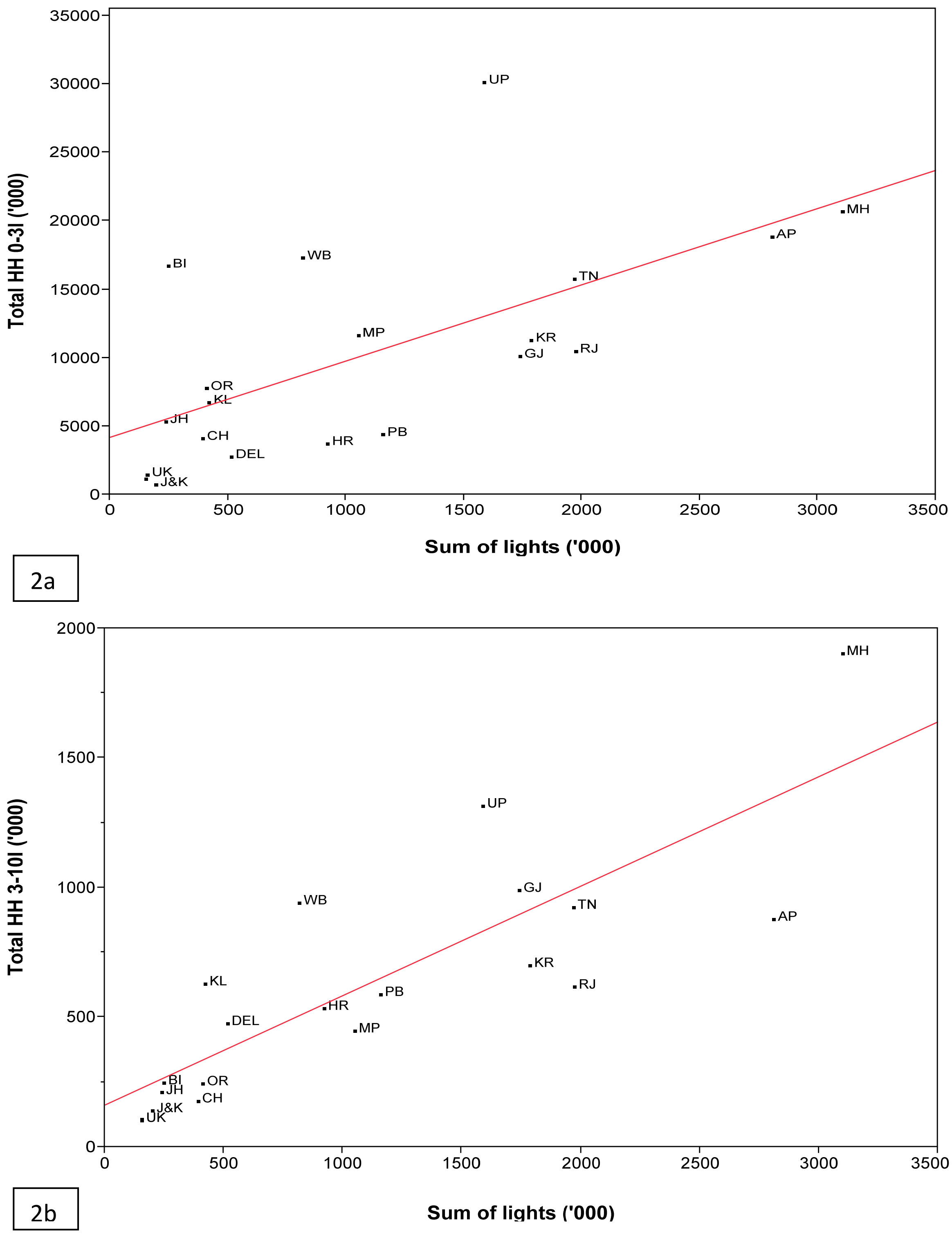


Figure 2. (a) Sum of lights and lower income households with annual household income of less than 3 lakh **(b)** Sum of lights and middle income households with annual household income between 3 lakh and ten lakh.

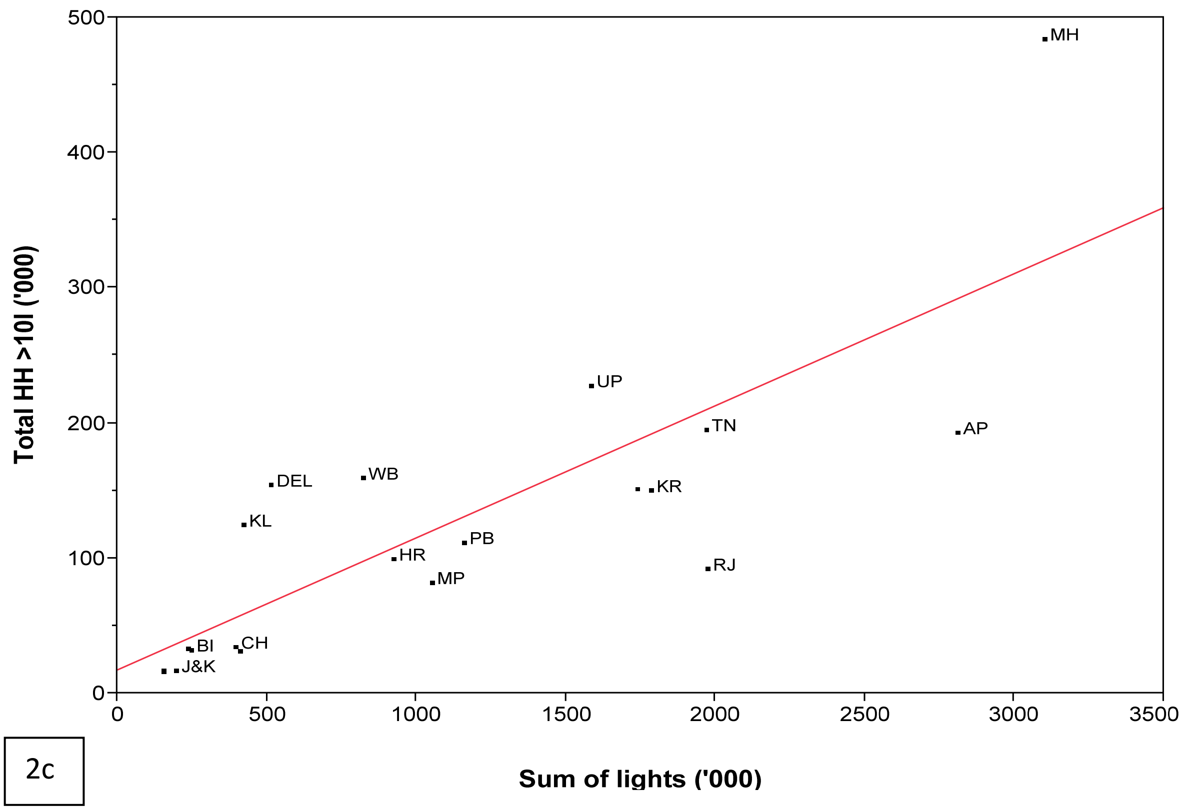


Figure 2. (c) Sum of lights and upper income households with annual household income of more than ten lakh.

households in various states in India. (See Table 1 in Appendix for exact figures of all States and Union Territories).But a closer examination of these graphs hints at some underlying dynamics that are clearly boosting the position of some states while pulling down those of others relative to the presence of lights there. For example, Maharashtra and Andhra Pradesh are not very different in terms of sum of lights (or even total GDP) although the difference in number of upper income households is fairly high for these states in Figure 2c. Both states are well lit in the nightlights image (Figure 1), but the region in and around Mumbai (India's commercial capital) in Maharashtra is very brightly lit. Mumbai also happens to house the richest industrialists and the biggest stars of the Indian film industry – Bollywood, apart from other high income official dignitaries. Hence, presence of a very high income zone in Maharashtra places it as an extreme outlier in terms of the number of affluent households. Similarly, Madhya Pradesh and Rajasthan's relative positions in the same figure (2c) may leave some reservations in mind. Though both are less developed states in the Indian economy, and both have similar number of affluent households as can be seen in the graph, the difference in sum of lights is almost double between them. A look at the nightlights image (Figure 1) gives a probable explanation. While most of Rajasthan is seen to be dark, there is consistent lighting all along the international border with Pakistan boosting sum of lights in those border districts, and a few scattered districts with high level of development (Jaipur, Udaipur, etc.). Uttar Pradesh is again one of the states with high sum of lights and higher number of households in any income category. But when compared to its nightlights image in Figure 1, it tells us that even a variegated lighting pattern, with a few bright pockets and most of the remaining area being in the dark, can boost the state's totals despite being below other states in the hierarchy of overall development.

Another noteworthy feature of these graphs is that most of the states with the highest population densities figure above the fitted line in one or more of the income categories graphs. These include Uttar Pradesh, Delhi, Kerala, West Bengal, and Bihar (in the lower income households' graph, Figure 2a). While greater population does imply relatively higher use of lights by way of sheer numbers, some studies have also shown that several very highly developed and very densely populated regions maybe anomalously dark relative to their population due to the presence of large densely populated areas (like New York City and the Los Angeles region in the United States of America) [31]. Thus population differentials seem to influence the relationship between nightlights and households' incomes to some degree, though the exact role of population in determining this relationship maybe somewhat complex.

This visual analysis suggests that it is not only the total wealth or GDP of the states that influences the brightness level caused by the sum of lights, but it is probably the distribution of income which plays a more important role. The examination at the state level suggests that a detailed probe into the relationship between nighttime lights and households in different income brackets at a finer spatial resolution, that is, at the district level, might provide better insights into the association between the nighttime lights and households in different income brackets. Understanding the relationship between nighttime lights and income distribution at the district level and mapping the errors would help to develop predictive models of income distribution from nighttime lights image. Once satisfactory models are built, they could be used to estimate the number of households in different income brackets for years when the data are not collected, or provide an alternative estimate or supplemental estimate of income distribution. As pointed out by Browning and Crossley [32], 'Several error prone measures are better than one.'

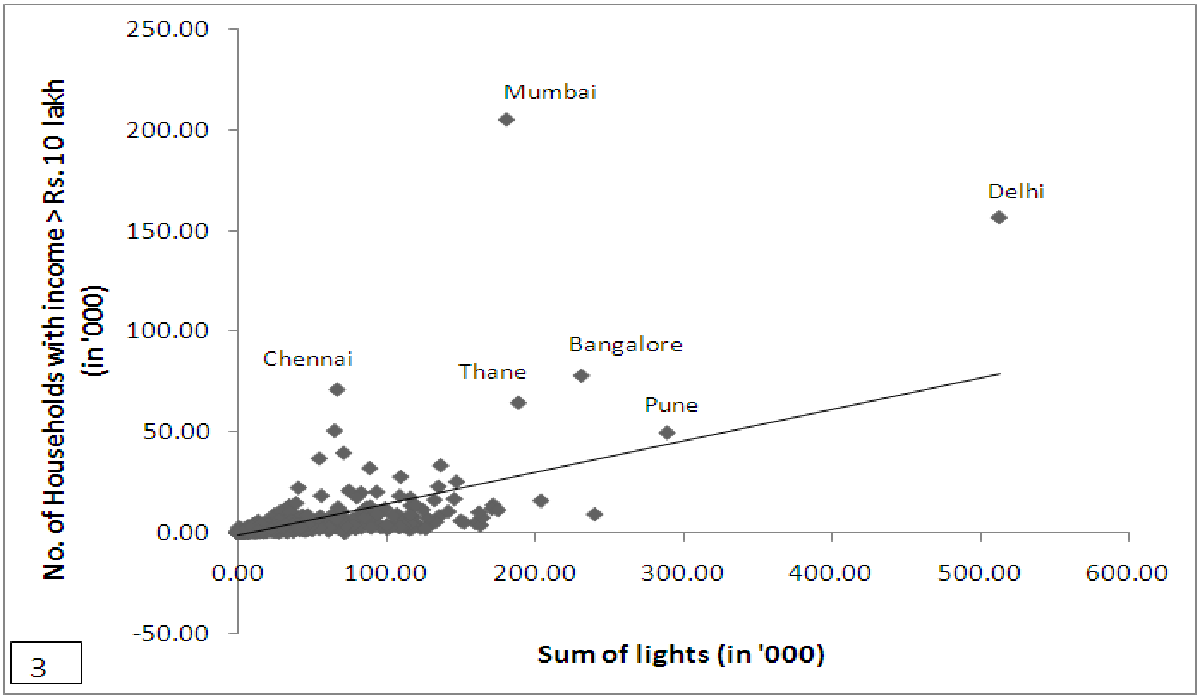
2.5 The relationship between nighttime lights and income distribution

Several studies have been carried out relating nighttime lights with economic activity [5, 6, 21, 22, 23, 24, 25]. But most studies relating nighttime lights with economic activity have focused on associating higher *totals* in economic activity with higher sum of lights across regions. The question addressed in this paper is how best to model the relationship of nighttime lights with not just the wealthy but also the relatively worse-off *within* a region. Accordingly, we propose two simple hypotheses with respect to the relationship between nightlights satellite imagery and households' income data.

Among all the residents of an area, it is not difficult to note that the relatively better-off will use more lighting simply by virtue of them being able to afford it. Thus, assuming electricity consumption to be a normal good, **Hypothesis One** states that nightlights should be more closely associated with the richer in any given region than with the poor.

But, in a densely populated and less-developed economy like India, the number of households in lower income brackets is far higher than the number of households that are relatively better-off. Thus, an immediate challenge of trying to associate lights with households with different levels of income within a region is that we can only observe total light in a region at night, and cannot by any means segregate the lights consumed by rich from lights consumed by the poor. Hence, **Hypothesis Two** suggests that sum of lights as seen from a region when linked with number of households, will most likely tend to under-estimate the number of poor households and over-estimate the rich households.

A scatter plot of the total number of households in any income category and sum of lights in that area shows a non-linear trend. Figure 3 gives such a scatter plot for the highest income cut. Taking the natural logs of both the variables and plotting the data for all the 585 districts gives a clear monotonic relationship between lights and number of households in any income category. Figure 4a plots this relationship for the highest income cut. Another feature noted in these diagrams is an apparent difference in the level of relationship of certain groups of cities with lights. Most of the districts with big economies and good infrastructure are visible at a plane higher than the rest of the districts. For example, Delhi, Mumbai are obvious outliers even in the logarithmic data scatter plot. Being metropolitan areas, these districts not only have a very high concentration of population, in every income class, but are also major industrialized centers. These infrastructurally strong economies with extensive road and rail networks, ample street lighting, large number of highways and other publicly lighted spaces, attract and support a vibrant base of economic activity in neighboring areas too. These high-income neighborhoods act as suburbs for these mega-cities. Such suburbs like Noida and Thane are also seen to be among the specs generally above the rest in the scatter plot.



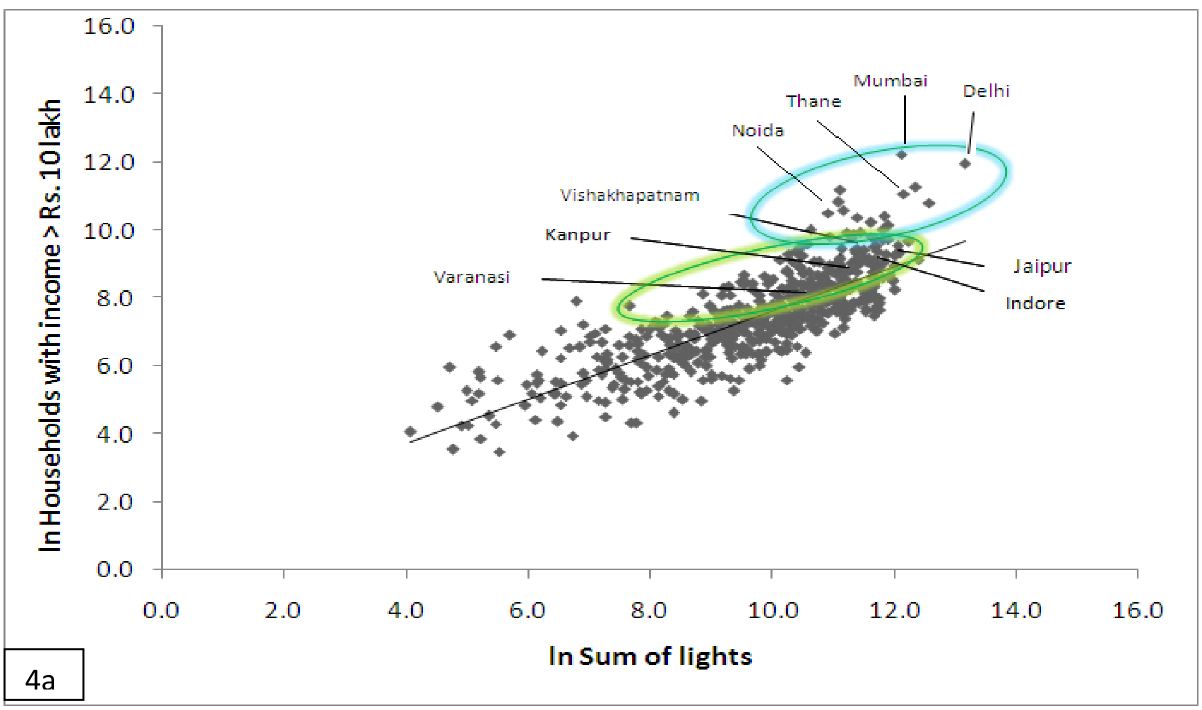


Figure 3. Scatter plot for sum of lights and households in the upper income bracket **(4a)** Scatter plot for natural log of the sum of lights and natural log of number of households in the upper income bracket

Apart from these selected pockets of massive economic activity, there are other less conspicuous but large cities that are centers of heightened economic activity in the heartland of the states. The presence of large industrial clusters (Indore, Bokaro), or commercially important ports/harbors (Vishakhapatnam, Surat), or locations along the river basins and in fertile belts (Nasik, Varanasi, Kanpur) make them places of relatively flourishing market activity in the contours of their respective states. These large districts offer better livelihood opportunities and hence attract people from the nearby less developed villages or countryside.

But is it just commerce that drives higher income levels in a district? Public consumption, as opposed to private consumption, could also boost the economy of a place that may not be adequately captured by nightlights. In that respect, districts with higher government spending and large public infrastructure should also be on a higher plane than the remaining districts. The places other than commercially vibrant cities that have such strong infrastructure are the administrative capitals of states, with a good amount of state machinery contributing to the district's economy (for example Jaipur, Thiruvananthapuram, etc).

Thus, examination of the scatter plots of lights and households in each income bracket not only suggested the use of a logarithmic model, but also that certain districts needed to be treated differently than the rest given the uneven spread of economic development in India, with a few urban cores attracting a more than proportionate share of the pie.

In this paper we classified districts into five broad categories based on their commercial and administrative importance: metropolitan districts, suburbs of metros, large industrialized towns, State administrative capitals, and all the remaining as 'others'. Accordingly, four dummy variables were created –

Metros – Districts in which the top 8 metropolitan cities, on the basis of their population numbers, are located. These are Mumbai, Delhi, Chennai, Bangalore, Kolkata, Ahmadabad, Hyderabad and Pune.

Sub-metro – The 13 highly populated and industrialized districts (for example, Noida and Thane) are the districts located around the major metros.

Large Towns – These include districts other than the Metros and Sub-metros which are the industrial hubs in their respective states (for example, Kanpur, Nasik and Coimbatore). They are classified on the basis of their market size.

Capitals – The districts in which the 34 capitals of the states and Union Territories (UTs) are located. It is to be noted that Chandigarh is the capital of both the states of Punjab and Haryana, and also the UT of Chandigarh. Jammu and Kashmir has two capitals, Srinagar in summer, and Jammu in winter, and they are located in the districts of Srinagar and Jammu, respectively.

Not only are these districts centers of heightened economic activity, they also help us in capturing the high population zones that are likely to contribute most to economic activity to a fair degree. So, while keeping the underlying relationship between sum of lights and incomes the same for all districts, these dummies let the levels of this relationship vary for districts with different economic forces of attraction operating within. Although six out of the eight metros are also administrative capitals, the differential between say Delhi and Shimla or between Mumbai and Bhopal in terms of their economic contribution to the country's growth accorded the use of separate dummies for State capitals and major metropolitan districts in our analysis.

Subsequently, three models were developed for upper, middle and lower income households by taking the natural log of the number of households in the three income brackets and the natural log of the sum of lights, and including the four dummies – Metros, Sub-metros, Large towns, and Capitals (Equation 1). It is to be noted that data for nine districts of Delhi were combined together and treated as Delhi. Also four districts in Jammu and Kashmir (Kargil, Punch, Leh (Ladakh), Rajouri) and internationally disputed area in the state for which household data were unavailable were not included in the analysis.

$$\mathbf{Ln\ Y = \alpha + \beta_1 (Ln\ X_1) + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5} \qquad (1)$$

3. Results

Regressing the log of the number of households in the lowest income bracket against the log of the sum of lights and the four dummies gave an adjusted R^2 of 0.61. All the predictor variables except the dummy 'Large towns' were significant at the 95% confidence interval. Regressing the log of the number of households in the middle income bracket against the log of the sum of lights and the four dummies, provided an R^2 of 0.75 and all the predictor variables were significant at the 99% confidence interval. For the upper income bracket the regression gave an adjusted R^2 of 0.76. All the predictor variables were also significant for households in this income bracket at the 99% confidence interval (Table 1).

Table 1. *Results of the logarithmic multivariate regression model*

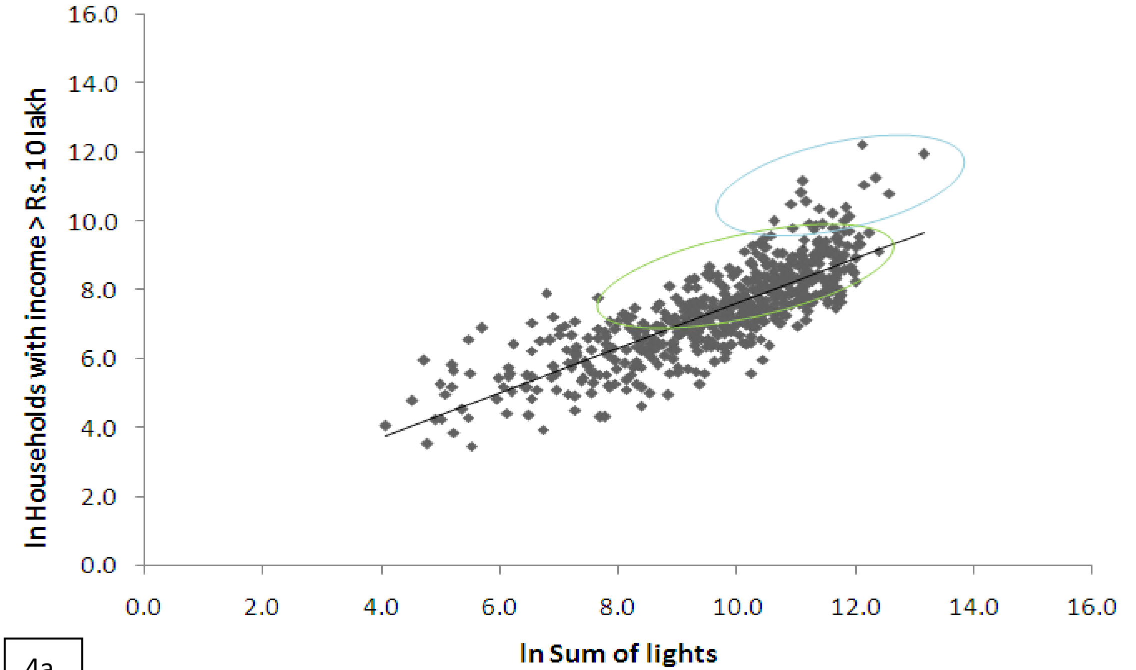
Predictor Variables	Natural log of number of households in the 'lower' income group	Natural log of number of households in the 'middle' income group	Natural log of number of households in the 'upper' income group
Natural log of Sum of Lights, β_1	0.45*	0.51*	0.55*
Dummy for Metro cities, β_2	1.01*	1.60*	2.11*
Dummy for Suburbs of Metros, β_3	$0.37^{\$}$	1.35*	1.67*
Dummy for Large Towns, β_4	$0.18^{\#}$	0.55*	0.72*
Dummy for Capital cities, β_5	$-0.30^{\$}$	0.46*	0.76*
Intercept, α	8.08*	4.26*	1.86*
Adjusted R^2	0.61	0.75	0.76

* Significant at the 99% Confidence Interval, $^{\$}$ Significant at the 95% Confidence Interval, $^{\#}$ Significant at the 90% Confidence Interval

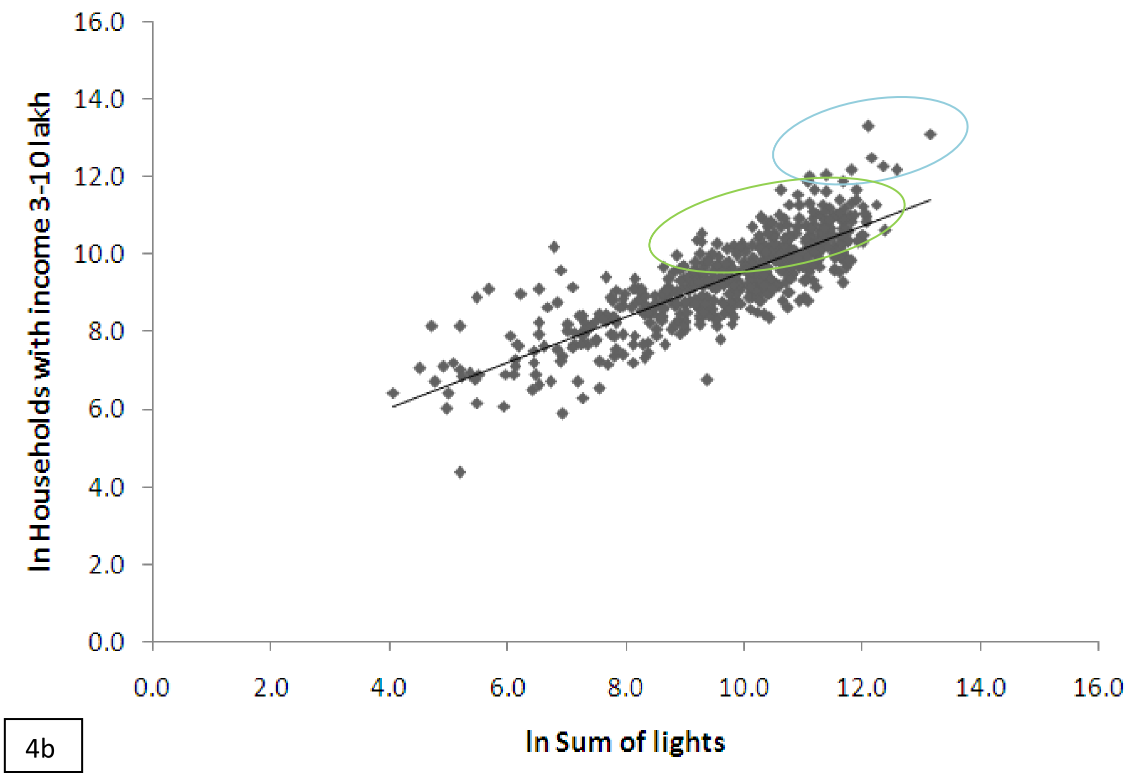
Table 1 suggests that as we move from the lower to upper income brackets, the relationship between sum of lights and number of households tightens i.e. lights are better able to estimate households in more affluent categories as depicted by the higher adjusted R^2 values for the middle and upper income class. The magnitude of the coefficient for the nightlights variable (β_1) also increased for the more affluent segment, indicating a steepening of the slope for the association of lights with the rich than with the not so rich. Hence, it is safe to conclude that the proposed Hypothesis One holds true with respect to the relation of nightlights with households' income.

It is also noteworthy that the coefficients for all the dummies increase monotonically as we go up the income categories. In fact, the co-efficient for the dummy for Metropolitan districts (β_2) more than doubles for the upper income households as compared to the lower income households. There is a similar but perhaps even starker surge in β values for the other three dummies for Suburbs, Large towns and Capital districts as we look across the income groups. Therefore, these dummies provide a much bigger boost to the middle and upper income segments than the lower income segments for the same relationship of lights with number of households in the respective income categories. It is also noticeable that the β's are consistently highest for the metropolitan dummies, followed by the dummy for Suburbs of Metros.

Figures 4b and 4c show the scatter graphs for the middle and lower income groups without the dummies. Comparing these figures (4b and 4c) with the figure for the upper income category (4a) it was seen that the logarithmic scatter plot for the highest income group was the most concentrated along the fitted line, as already noted by the higher adjusted R^2 value for upper income segment.


16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.0
2.0
4.0
6.0
8.0
10.0
12.0
14.0
16.0
ln Households with income > Rs. 10 lakh
ln Sum of lights

4a


16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.0
2.0
4.0
6.0
8.0
10.0
12.0
14.0
16.0
ln Households with income 3-10 lakh
ln Sum of lights

4b

Figure 4. (a) Scatter plot for natural log of the sum of lights and natural log of the number of households in the upper income bracket **(b)** Scatter plot for natural log of the sum of lights and natural log of number of households in the middle income bracket.

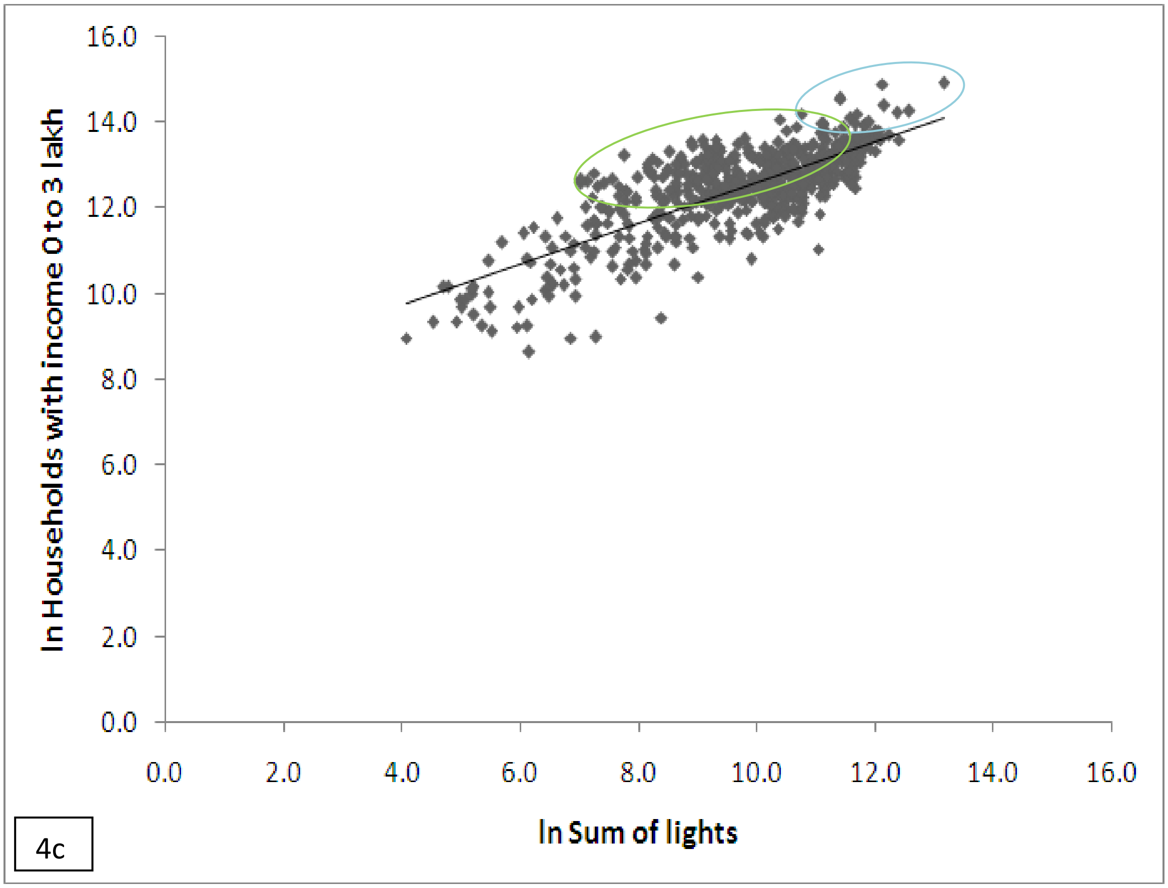


Figure 4. (c) Scatter plot for natural log of the sum of lights and natural log of the number of households in the lower income bracket.

Also seen in blue and green circles are the districts that the four dummy variables intend to capture. These commercially important big districts are not only towards the top right in the scatter graphs, they are also located above the rest of the districts, on a higher level. This level difference is seen to expand for the higher income category household graph, implying higher β values for these dummies in the model for the higher income group.

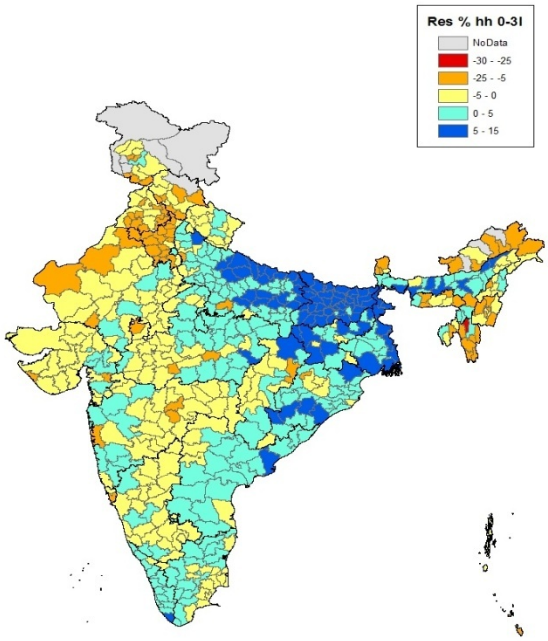
3.1. *Error Maps*

Error maps were drawn using the predicted errors as a percentage of the actual dependent variable (log of the number of households in each of the three income categories). Mapping the residual errors for the number of households in the three income brackets enabled us to understand the districts for which the sum of lights over- or under-estimated the predicted income distribution values.

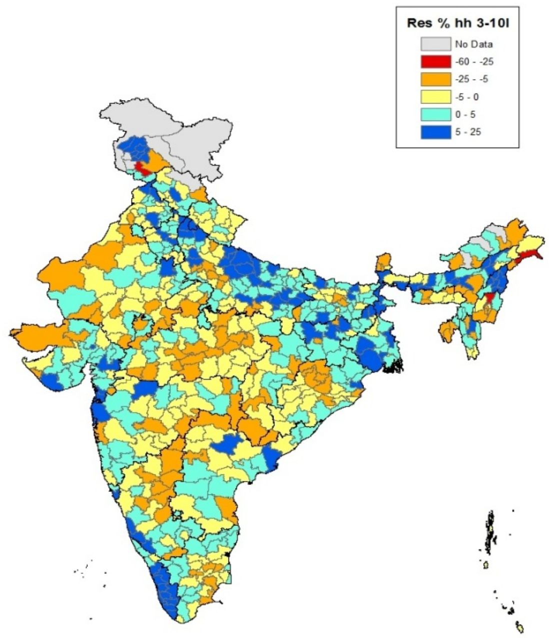
The error map of the lowest income group showed that the regression model using the sum of lights and the four dummies underestimated the predicted number of households within the income bracket of 0-3 lakh by up to 15% for most of the districts of the northern state of U.P., districts of the eastern states of Bihar, Jharkhand, West Bengal (W.B.), Orissa, and the north-eastern state of Assam. Even for most of the districts of the southern states of Andhra Pradesh and Kerala, the number of households in the lowest income bracket predicted by the model was underestimated. The number of households predicted by the model was overestimated by up to 25% for most of the districts of the northern states of Punjab, Haryana, and Himachal Pradesh; the western states of Gujarat and Rajasthan; and the north-eastern states of Sikkim, Arunachal Pradesh, Mizoram, Manipur, and Assam. The greatest overestimation (over 25%) were for the districts of Kolasib and Serchhip in Mizoram, Mahe in Pondicherry, North Sikkim in Sikkim, and the Nicobars of the Andaman and Nicobar Islands (Figure 5a).

In case of the error map for the middle income group (3–10 lakh), a dominant pattern of underestimation by the model, by up to 25%, was seen for the northern state of U.P.; the eastern states of Bihar, Jharkhand, and W.B.; the southern state of Kerala, and the north-eastern states of Assam, Nagaland, and Mizoram. For the rest of the states the districts showed a mixed pattern of over- and under-estimation, The highest percentage of overestimation (greater than 25%) were for the districts of Changlang in Arunachal Pradesh, Udhampur in Jammu and Kashmir, Senapati in Manipur, and Yanam in Pondicherry (Figure 5b).

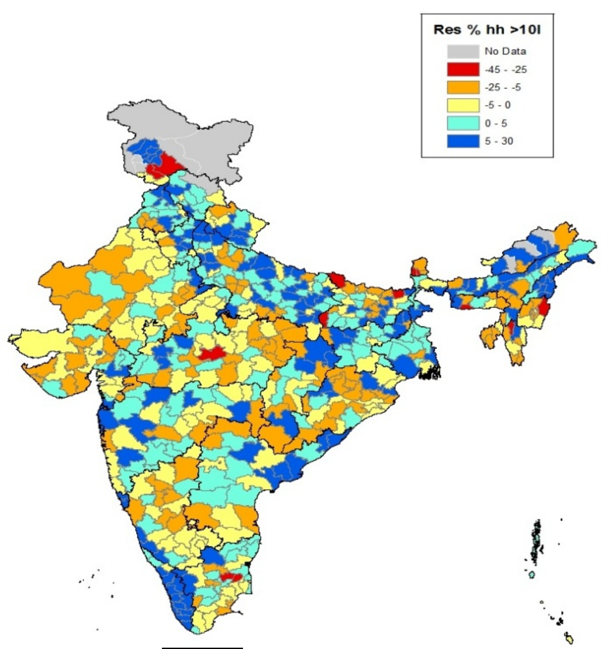
The error map for the households in the highest income category also shows a pattern of underestimation by the model for the districts in the northern states of U.P., Punjab, Himachal Pradesh (H.P.); districts in the eastern states of Bihar, Jharkhand, W.B.; the north-eastern states of Assam and Nagaland; and the southern state of Kerala. The number of households is overestimated by up to 25% for most of the districts of the Western states of Gujarat and Rajasthan, and the southern state of Karnataka. The greatest percentage of overestimation (greater than 25%) were for the districts of Araria and Kaimur in Bihar, South Sikkim and West Sikkim in Sikkim, Raisen in Madhya Pradesh (M.P.), Ukhrul in Manipur, Yanam in Pondicherry, Diu and Daman in Daman and Diu, Udhampur in Jammu and Kashmir, and Ariyalur and Perambalur in Tamil Nadu (Figure 5c).



5a



5b



5c

Figure 5. (a) Percentage residual error map for households in income bracket of 0-3 lakh **(b)** Percentage residual error map for households in income bracket of 3-10 lakh **(c)** Percentage residual error map for households in income bracket of greater than 10 lakh

4. Discussion

The above analysis shows that a good relationship exists between nighttime lights and income distribution at the district level, with the relationship being stronger for households in the higher income brackets. This has important implications with respect to the estimation of income inequalities in India.

The error map for the lowest income group showed an underestimation of number of households in this income category by the nighttime lights for the districts in the states of Bihar, Jharkhand, W.B., Orissa, Assam, Andhra Pradesh and Kerala. This is probably because of the greater number of households in the lowest income group in these highly populated states, and also because more than 70% of the population in these states is rural. Home and outdoor lighting in these states are not proportional to the high population numbers. Also, this maybe because of the lack of government provision of public utilities like street lighting, stable power supply, in these states in comparison to the southern states. On the contrary, the nighttime lights overestimated the number of households in the lowest income group in the districts of the states of Rajasthan, Punjab, and Haryana. Punjab and Haryana are agriculturally rich states, and there are rich farmers who contribute to significant lighting in these two states. Poor people are also overestimated in the border districts of Rajasthan probably because of bright lighting along the border areas with Pakistan.

For the middle income group, the nighttime lights underestimated the number of households in the same states as in the case of the lowest income group, except for Andhra Pradesh and Orissa. In this income category, the number of households is slightly over- or under-estimated by the nighttime lights in the districts of Andhra Pradesh and Orissa. For the districts in Punjab and Haryana, the middle income group households are slightly over or under-estimated, and in fact are underestimated by up to 25% for some of the districts. This showed that the middle income group is better captured by the nighttime lights image. In the case of Rajasthan too, it was seen that the middle income group was better captured by the nighttime lights.

For the highest income group, a mixed pattern of over- and under-estimation was observed for the districts of the states. Since the number of households even in the highest income category was underestimated for most of the districts in U.P, and Bihar, it can be said that the lights are not able to capture the population in the highest income group in these populous, rural, and relatively backward states. Underestimation of the number of households in the highest income group in Kerala requires further investigation. One probable justification may be that Kerala being ranked the highest in Human Development Index [33] among all the states in India has more equitable distribution of wealth, and so the number of households in the highest income group are the most underestimated. Another reason could be the high density of population (Kerala is the third most densely populated Indian state after Bihar and West Bengal), leading to under-estimation of the rich in Kerala.

From the scatter plots and the error maps it is clear that a good relationship exists between nighttime lights and income distribution at the district level, with the relationship being stronger for households in the higher income brackets. The pattern of under-estimation of lower income households is unmistakably clear for most of the eastern states in India, with the highest under-estimation being evident for most of the districts in the densely populated BIMAROU (Bihar, Madhya Pradesh, Rajasthan, Orissa, and Uttar Pradesh) states. There is an over-estimation in very few districts by nightlights for this income group, and occurs only in the relatively sparsely populated districts. The under-estimation is much less for middle-income households, and in fact an emerging pattern of variegated over-estimation of number of households in this income category appears. This pattern of over-estimation spreads further to more districts for upper income households. Hence, we can say that Hypothesis Two also holds true with respect to sum of lights and number of households in a region.

5. Conclusion

Examination of the relationship between nighttime lights and income distribution at the district level for India shows promising results. It is evident from this analysis that population density may also have an important influence in driving these results. While we have implicitly tried to model population in our analysis using dummy variables for different kinds of districts, this complex component needs to be investigated in detail in further studies. The stronger relationship between lights and affluent households is also something that needs further exploration. The insights obtained through this study would help to make improved models for estimating income distribution at finer spatial resolutions and in turn modify the models further. The objective is also to estimate income distribution for years when data are not available or to provide a supplemental means for estimating income distribution even for years for which data are available. We do intend to conduct the study for later years than 2004, when all the required data in the analysis becomes available.

The radiance-calibrated nighttime images are superior products incorporating the best of both the stable lights and the fixed gain products, and they have unsaturated urban cores. Even with coarse spectral and spatial resolution the radiance-calibrated nighttime lights image shows great potential for making economic estimates like no other remote sensing data. It can be hoped that with the launch of the Visible Infrared Imaging Radiometer Suite (VIIRS) in October 2011, which will have on-board calibration and higher spectral and spatial resolution (0.8 km), better low-light imaging data would be made available, which would help to create better income distribution maps.

Developing alternative methods of estimating income distribution at finer spatial resolutions in information deficient countries like India and mapping them would greatly assist the government in developing appropriate socio-economic policies and directing them where they are needed the most.

Acknowledgements

Very special thanks to Dr. Christopher D. Elvidge and his team at NGDC, NOAA, for providing us with the nighttime image of India which was used in this study. Our sincere gratitude to the team at Indicus Analytics for providing us with the income distribution data of India of 2004, and also for the district level shapefiles.

References

1. CIA world factbook - https://www.cia.gov/library/publications/the-world-factbook/geos/in.html. Last accessed on 4 August, 2011.
2. Suryanarayana, M.H.; Agarwal, A.; Seeta Prabhu, K. Inequality-Adjusted Human Development Index for Indian States 2011. *UNDP India 2011*.
3. Min, B. Democracy and Light: Electoral Accountability and the Provision of Public Goods. *Paper presented at the annual meeting of the MPSA Annual National Conference*. Chicago, Illinois, April 6, 2008.
4. Henderson, J.V.; Storeygard, A.; Weil, D.N. Measuring Economic Growth from Outer Space. *NBER Working Paper 15199. National Bureau of Economic Research*. Cambridge, Massachusetts, 2009.
5. Ghosh, T.; Powell, R.; Elvidge, C.D.; Baugh, K. E.; Sutton, P.C.; Anderson, S. Shedding Light on the Global Distribution of Economic Activity. *The Open Geography Journal* **2010**, *3*, 147-160.
6. Doll, C.N.H.; Muller, J.P.; Morley, J.G. Mapping Regional Economic Activity from Nighttime Light Satellite Imagery. *Ecological Economics* **2006**, *5*, 75-92.

7. Roychowdhury, K; Jones, S.; Arrowsmith, C.; Reinke, K.; Bedford, A. The Role of Satellite Data in Census: Case study of an Indian State. *Proceedings of the 30th Asia-Pacific Advanced Network Meeting.* Hanoi, Vietnam, 2010.
8. Ministry of Statistics and Programme Implementation: NSSO Socio-economic Survey, Government of India, <http://mospi.nic.in/nsso_4aug2008/web/nsso/se_nsso.htm> Last accessed on August 4, 2011
9. Ministry of Statistics and Programme Implementation: CSO Report and Publications, Government of India, <http://mospi.nic.in/cso_rept_pubn.htm> Last accessed on August 4, 2011
10. Imhoff, M.L.; Lawrence, W.T.; Elvidge, C.D.; Paul, T.; Levine, E.; Prevalsky, M. Brown, V. Using Nighttime DMSP/OLS Images of City Lights to Estimate the Impact of Urban Land Use on Soil Resources in the United States. *Remote Sensing Of Environment* **1997**, *59 (1)*, 105-117.
11. Small, C. Multi-resolution Analysis of Urban Reflectance. *Workshop on Remote Sensing and Data Fusion over Urban Areas*. Rome, Italy, 2001.
12. Milesi, C.; Elvidge, C.D.; Nemani, R.R.; Running, S.W. Assessing the Impact of Urban Land Development on Net Primary Productivity in the Southeastern United States, *Remote Sensing of Environment* **2003**, *86*, 273-432.
13. Sutton, P.C. Modeling Population Density with Nighttime Satellite Imagery and GIS. *Computers, Environment, and Urban Systems* **1997**, *21 (3/4)*, 227-244.
14. Sutton, P.C.; Roberts, D.; Elvidge, C.D. Baugh, K. Census from heaven: An Estimate of the Global Population using Nighttime Satellite Imagery. *International Journal of Remote Sensing* **2001**, *22 (16)*, 3061-3076.
15. Sutton, P.C.; Elvidge, C.D. Obremski, T. Building and Evaluating Models To Estimate Ambient Population Density. *Photogrammetric Engineering and Remote Sensing* **2003**, *69 (5)*, 545-553.
16. Elvidge, C.D.; Milesi, C.; Dietz, J.B.; Tuttle, B.T.; Sutton, P.C.; Nemani, R.; Vogelmann, J. U.S. Constructed Area Approaches the Size of Ohio. *EOS Transactions AGU* **2004**, *85 (24)*, 233-234.
17. Doll, C.N.H.; Muller, J-P.; Elvidge, C.D. Nighttime Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions. *Ambio* **2000**, *29*, 157–162.
18. Oda, T.; Maksyutov, S. A Very High-resolution Global Fossil fuel CO2 emission inventory Derived Using a Point Source Database and Satellite-observed Nighttime Lights, 1980–2007. *Atmospheric Chemistry and Physics Discussions* **2010,** *10(7)*, 16307-16344.
19. Ghosh, T.; Elvidge, C.D.; Sutton, P.C.; Baugh, K.E.; Tuttle, B.T.; Ziskin, D. Creating a Global Grid of Distributed Fossil Fuel CO2 Emissions From Nighttime Satellite Imagery. *Energies* **2010**, *3*, 1895-1913.

20. Cova, T.J.; Sutton, P.C.; Theobald, D.M. Exurban Change Detection in Fire-Prone Areas with Nighttime Satellite Imagery. *Photogrammetric Engineering and Remote Sensing* **2004**, *70 (11)*, 1249-1257.
21. Sutton, P.C.; Costanza, R. Global Estimates of Market and Non-Market Values Derived From Nighttime Satellite Imagery, Land Cover, and Ecosystem Service Evaluation. *Ecological Economics* **2002**, *41*, 509-527.
22. Sutton, P.C.; Elvidge, C.D.; Ghosh, T. Estimation Of Gross Domestic Product At Sub-National Scales Using Nighttime Satellite Imagery. *International Journal of Ecological Economics and Statistics* **2007**, *8*, 5-21.
23. Ebener, S.; Murray, C.; Tandon, A.; Elvidge, C.D. From Wealth To Health: Modeling The Distribution of Income Per Capita at the Sub-national Level Using Nighttime Light Imagery. *International Journal of Health Geographics* **2005**, *4*, 1-17.
24. Ghosh, T.; Anderson, S.; Powell, R.L.; Sutton, P.C.; Elvidge, C.D. Estimation of Mexico's Informal Economy and Remittances Using Nighttime Imagery. *Remote Sensing* **2009,** 1(3), 418-444.
25. Ghosh, T.; Powell, R.; Anderson, S.; Sutton, P.C.; Elvidge, C.D. Informal Economy and Remittance Estimates of India Using Nighttime Imagery. *International Journal of Ecological Economics and Statistics* **2010**, *17 (P10)*, 16-50.
26. Elvidge, C.D.; Sutton, P.C.; Ghosh, T.; Tuttle, B.T.; Baugh, K.E.; Bhaduri, B.; Bright, E. A Global Poverty Map Derived From Satellite Data. *Computers and Geosciences* **2009**, *35*, 1652-1660.
27. 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.
28. Ziskin, D.; Baugh, K.E.; Hsu, F.C.; Ghosh, T.; Elvidge, C.D. Methods used for the 2006 Radiance Lights. *Proceedings of the Asia Pacific Advanced Network.* Hanoi, Vietnam, 2010.
29. Elvidge, C.D.; Ziskin, D.; Baugh, K.E.; Tuttle, B.T.; Ghosh, T.; Pack, D.W.; Erwin, E.H.; Zhizhin, M. A Fifteen Year Record of Global Natural Gas Flaring Derived From Satellite Data. *Energies* **2009**, *2*, 595-622.
30. Urban Consumer Spectrum of India 2004-05, Indicus Products (document not published)
31. Elvidge, C.D.; Baugh, K.E.; Dietz, J.B.; Bland, T.; Sutton, P.C.; Kroehl, H.W. Radiance Calibration of DMSP-OLS Low-light Imaging Data of Human Settlements. *Remote Sensing of Environment* **1999**, *68*, 77-88.
32. Browning, M.; Crossley, T. Are Two Cheap, Noisy Measures Better Than One Expensive, Accurate One? *American Economic Review* **2009**, 99(2), 99-103.

33. National Human Development Report 2001, Planning Commission, Government of India. http://planningcommission.nic.in/reports/genrep/nhdrep/nhdch2.pdf

Appendix

Table A: Sum of Lights and Total Number of Households in 3 income categories for Indian States for 2004

States	Sum of Radiance-calibrated lights (2004)	Number of households earning less than 3 lakh annually	Number of households earning 3 lakh to 10 lakh annually	Number of households earning more than 10 lakh annually
Andaman & Nicobar Islands	6,556	80,787	9,746	2,011
Andhra Pradesh (AP)	2,805,720	18,959,462	885,993	195,535
Arunachal Pradesh	4,983	227,690	10,463	2,297
Assam	221,189	5,447,201	226,697	30,217
Bihar (BI)	247,592	16,883,440	258,752	34,447
Chandigarh	34,998	199,639	50,207	13,488
Chhattisgarh (CH)	392,232	4,267,675	186,996	36,386
Dadra & Nagar Haveli	7,461	64,832	2,811	648
Daman & Diu	8,428	41,361	4,110	307
Delhi (DEL)	513,143	2,967,884	486,945	156,518
Goa	57,246	257,479	58,424	12,502
Gujarat (GJ)	1,738,450	10,294,230	998,046	153,718
Haryana (HR)	921,001	3,920,312	543,966	101,889
Himachal Pradesh (HP)	152,168	1,293,758	115,816	19,200
Jammu & Kashmir (J&K)	196,184	1,577,505	148,723	19,467
Jharkhand (JH)	237,464	5,485,974	220,203	35,150
Karnataka (KR)	1,784,800	11,427,033	707,847	153,255
Kerala (KL)	417,735	6,929,874	636,988	127,326
Lakshadweep	92	11,274	1,138	122
Madhya Pradesh (MP)	1,050,140	11,815,891	454,809	84,778
Maharashtra (MH)	3,100,640	20,825,559	1,911,836	486,890
Manipur	18,672	436,026	16,349	2,859
Meghalaya	24,038	475,840	27,773	3,562
Mizoram	7,265	166,470	19,406	2,035
Nagaland	8,962	351,355	46,548	6,420
Orissa (OR)	408,842	7,890,793	253,844	33,863
Pondicherry	21,573	243,370	26,290	5,743
Punjab (PB)	1,157,860	4,514,319	598,579	113,959
Rajasthan (RJ)	1,973,570	10,602,928	624,835	94,909
Sikkim	6,133	126,322	8,112	804
Tamil Nadu (TN)	1,968,490	16,070,118	931,568	197,068
Tripura	46,584	745,367	23,733	3,067
Uttar Pradesh (UP)	1,585,980	30,238,608	1,322,748	230,589
Uttarakhand (UK)	155,065	1,760,095	111,436	18,149
West Bengal (WB)	817,732	17,448,100	951,040	162,164