Internship Report No.08/2019-20

Government of Karnataka
FISCAL POLICY INSTITUTE

Kengeri Post, Bangalore-Mysore Road, Bengaluru-560060
Phone: +91 80 26971000, Fax: +91 80 26971010,
e-Mail: director[at]fpibangalore[dot]gov[dot]in

Internship Report

On

GDP Estimation Using Nightlights data: Relevance and Applicability for India

Himanshu Dixit

March, 2020

Certificate

This internship report titled *"GDP Estimation Using Nightlights data: Relevance and Applicability for India"* is a report on the study taken up at the Fiscal Policy Institute (FPI) in 2019-20.

The internship report is prepared by Himanshu Dixit, Masters in Public Policy, National Law School of India University, Bengaluru under the mentorship of Ms. Shashikala K B, Special Officer (IT), Fiscal Policy Institute.

All opinion and conclusions expressed in the internship report are of the Intern and usual disclaimer applies.

Sujit Kumar Chowdhury
Director, FPI

Institute's Seal

Fiscal Policy Institute
F.P.I.
Bengaluru-560 060

2

Acknowledgement

I would like to thank the Fiscal Policy Institute for supporting me to pursue this research project. My mentor, Shashikala K B, special officer at FPI, has played an instrumental role in helping and motivating me to complete this report. A special thanks and token of respect goes to Prof. M.R. Narayana whose interventions have enriched my understanding of the subject. Lastly, I thank my friends at National Law School of India University whose support and remarks were indispensable to the work.

Table of Contents

Background

Gross Domestic Product (GDP) is a standard measure of the economic activity that denotes the value created through production of goods and services in a given period of time. According to IMF, GDP is an important indicator of the size and performance of the economy and GDP growth rate can shed light on the general well being of the people. Alternatively, it is also referred to as the income of the country.

National Accounts are kept and updated at regular intervals for the purpose of calculating GDP. National accounts contain information and figures such as pertaining to output of firms, changes in inventories, employee compensation, net exports etc. There is no direct way to avail of this information, and must necessarily be obtained from multiple sources. The difficulty of achieving coherence in this complex mix of data which is always incomplete is the primary reason why estimation of GDP is a highly complicated process. National Account statistics is a term often used to signify the process of approximation and statistical adjustments informing the calculation of GDP.

The presence of large formal economy[1] and robust statistical systems in developed nations facilitate GDP calculation with relatively better accuracy. Even so, the patchwork of statistical formulations can't be avoided. Low- and middle-income per capita countries, on the other hand, are marked by high levels of 'informality'[2], which makes it extremely hard to arrive at a good estimation. Capturing the informal economy is not possible in national accounts and is usually done through infrequent surveys. The size and growth of the total economy is thus spuriously projected. Needless to add, the national accounts systems are poorly resourced and maintained, and thus, extremely unreliable.

The situation in India is not alarming but still a cause for concern. According to a World Bank report on statistical capacity of developing countries, India scores 77 on a scale of 100 (Feenstra 2015). India does well in terms of periodicity of data but scores low in methodology and source data categories. Ministry of Statistics and Programme Implementation (MOSPI), Government of India admits the fact that the decentralised character of Indian Statistical System makes it reliant on data collection and

[1] A universally agreed upon definition of formal economy includes 'all firms and entities that are legally registered and remitting taxes on their income.

[2] The most comprehensive definition of informality is given by ILO which covers all the characteristics of informal economy. Informality as such refers to all the labour and economic activities operating outside the formal reach of law and regulation.

maintenance done by a diverse range of autonomous agencies. The quality of data, hence, varies depending on the way these 'source agencies' function and appreciate the requirements of National Accounts. Problems and issues multiply as National Accounts have to rely on periodic surveys by National Sample Survey Organisation (NSSO) to capture the informal and unorganised aspect of Indian economy. Besides institutional shortcomings, survey based estimations give rise to sampling and non sampling errors, further problematizing the accuracy of calculations.[3]

Indeed, it's crucial that existing official statistical systems are improved without delay by deploying necessary funds and resources. At the same time, we must pay attention to the new developments suggesting that the gaps in traditional data can be filled by 'unconventional data'[4]. The United Nations (UN) Global Working Group on Big Data for Official Statistics has been active in exploring and promoting the possible use of big data to supplement the National Accounts. As per Development Cooperation Report, 2017 published by OECD, data generated and collected through remote sensors, satellite imaging, anonymized call records, social media are examples of innovative big data that combined with official accounts and survey data can really expand the statistician's horizon (OECD 2017).

It is in the spirit of this discourse around use of novel sources of data to measure socio-economic conditions that satellite recorded nightlights data is chosen to investigate its potential in improving methodologies of estimation of subnational incomes in India.

Review of Literature

The relationship between nightlights and economic activity has been long established by a number of researchers (Croft 1978; Elvidge 1997; Ghosh et al. 2009). Quite early on, Doll (2008) made a perceptive study of the applications of nightlights to a range of subjects such as urban extent, population, greenhouse gas emissions and economic activity. A splendid contribution to this fast-growing literature was made by Nordhaus et al. (2011) when luminosity was used as a proxy for traditional measures of output to estimate GDP for countries with next to no census data available. However, it wasn't until the publication of a seminal paper by Henderson et al. (2012) that a

[3] Sampling errors are known and they occur due to issues pertaining to sample size or non random sampling. Non sampling errors are systematic and random errors in doing the survey and recording the observations. Non sampling errors are hard to identify.

[4] Unconventional data here refers to as big data generated using unconventional means, such as remote sensing, social media etc.

comprehensive statistical framework was developed using nightlight data to augment the conventional techniques of GDP estimation through national accounts.

Not only did Henderson find convincing true income growth rate for bad data low to middle income countries, but also calculated the elasticity of nightlights with respect to income. Thereafter, the term 'Inverse Henderson Elasticity' was bandied about in several subsequent papers trying to 'illuminate' our understanding of the application of nightlight in economic and policy research. A nightlight development index was developed by Elvidge et al. (2012) contrasting the distribution of lights with spatial distribution of human population. Using nightlights data, Ghosh et al. (2013) was able to demonstrate the underestimation of remittances in Mexico by official accounts. The model proposed by Henderson was slightly modified by Beyer et al. (2018) to make it suitable to estimate district level income on a monthly basis.

Different approaches have been used by researchers to put nightlights data to use. Beyer (2018) has categorised the methodologies to incorporate nightlights into existing statistical framework in three ways. He calls them elasticity approach, spatial approach and lead approach. Elasticity approach models the long-term relationship between nightlights and GDP, thus allowing prediction of GDP levels. Spatial approach is concerned with the distribution of aggregate GDP across smaller entities such as states and districts based on nightlight distribution (Bundervoet et al. 2015). This methodology is known to have produced GDP predictions with a very high level of spatial granularity. The lead approach is based on the understanding that certain indicators of the economy like the Industrial Production Index and Manufacturing Index bear a relationship with GDP changes. Once information on lead indicators is released, changes in GDP can be predicted. In this methodology, nightlights data is just another lead indicator of the GDP, among many others. This approach is particularly useful for macroeconomic forecasting in the short run.

All these studies didn't forget to alert us to the measurement errors of the nightlights estimation model, but with unanimous agreement that combined with conventional estimation methods, the hybrid model delivered better results. The confidence in evolving methodology is apparent with the studies being conducted to examine spatial inequalities at district (Dehejia 2017) and village (Mukhopadhyay 2018) level in India. A related but different methodology was also devised by Chodorow-Reich et al. (2018) to compute the effect of demonetisation on India's

GDP using nightlights data. Also, attempts have been made to utilize nightlights data to predict poverty in rural India (Subash 2018).

Most of these studies utilized night time lights processed from the light imaging data collected by the US Air Force Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS). DMSP was launched in 1970 and the data was available for the period between 1992 and 2013, when finally the programme ran its course. However, the follow up to DMSP was already in place as NASA and the National Oceanic and Atmospheric Association (NOAA) jointly flew Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB). These satellites were modern and installed with state of the art low-light imaging technology which provided considerable improvements over DMSP data including finer spatial resolution[5] and 45 times reduced pixel footprint (Elvidge 2017). These new capabilities greatly enrich the quality of data and consequently improve its usability for economic analysis and predictive modelling.

Research Context and Gap

The interest in extensively using nightlights to estimation of income, calculation of spatial inequalities and size of informal economy has grown tremendously in the last decade. Part of the reason why economists turn to nightlights data is because of a) paucity of good and reliable data and b) absence of accounting data due to weak statistical systems and presence of informality. Another important aspect is that this regression modelling is done at national, subnational and district level with extremely robust results. But very few papers try to use satellite data to do modelling at a much finer spatial and administrative entity such as a taluk or a village.

There is no paper in the relevant body of literature that tries to estimate the income of sub district entities. Studies have been conducted to measure the spatial inequalities at village level but no serious attempt has been made to calculate taluk or village level output. Hence, there is scope to explore the possibility of employing nightlights data to estimate highly localized output through statistical modelling.

Problem Statement

What: Poor national accounts data makes measurement of GDP at national level fraught with pitfalls and challenges (Nagaraj 2016). At subnational and district level,

[5] The spatial resolution is close to 0.5km^2

owing to even poorer statistical systems, estimation of income can't be done accurately. Further, it is almost impossible to calculate the income at sub-district level as the data about production activities is just not available.

How: Poor income estimates are responsible for spurious indicator values, thus leading to erroneous policymaking and allocation of funds. Also, it puts a question mark over the GDP projections released by the government (Subramanian 2019), which erodes investors' confidence in the economy.

Why: Unreliable measurements of income at all levels is a result of poor data or lack of data coupled with questionable methodology, which is a patchwork of statistical manipulations.

Research Questions

- Calculate the output and income per capita at sub-district level with acceptable levels of accuracy using nightlights data
- Compare the methodological approaches of utilizing the nightlights data for income estimation purposes as explained in the literature i.e. elasticity approach, spatial approach and lead approach
- Explore whether a completely novel methodological approach is possible to take that is an improvement over the existing methodologies

Research Design

The design of this research has an interesting feature with respect to the research objective and the methodology and tools adopted to achieve it. Usually, an exploratory research undertaking involves collection of qualitative data. The proposed study here is applied research which is exploratory in nature. It is applied in the sense that it tries to answer a practical question: How does one measure the output/income of small spatial entities like taluk or village without relying on conventional methodologies? It doesn't try to answer a fundamental question of knowledge but wishes to find a solution to a practical problem. It is exploratory as well because it wants to incorporate nightlights data to develop a statistical framework applicable to small regions, which has not been done before. There is virtually no published work that tried to investigate this question. Equally, works that have proposed statistical model with nightlights data have not ruled out the possibility of using it for income estimation at taluk level. So, it is a bit of a pleasant surprise that despite being a thoroughly quantitative study, the research design is exploratory.

The nature of data

The data in question here is satellite recorded details of the lights emitted at night across the globe. It has been made available in the form of raster files that can be processed in a GIS software such as QGIS, ARCGIS or programming languages such as python and R. The raw radiance data has been filtered through several algorithmic processes to establish standardization in the final products put out for meaningful use by researchers. Therefore, we must go through some of the relevant features of this data to realize the advantages and the limitations this dataset affords us while analysing it for different kinds of research.

As mentioned above, the data is provided in the form of raster files produced by The Earth Observations Group (EOG) at NOAA/NCEI. These files are basically the average radiance composite images i.e the lights emitted have been averaged temporally on monthly and annual basis.

Before the averaging process, the data has been processed to exclude stray light impact, illumination due to moonlight and cloud covers. A special product called VIIRS cloud mask (VCM) product is used to determine the extent of cloud cover and make necessary corrections.

Discussing the methods employed to filter out unwanted data is beyond the scope of this report but based on the claims made by the agency it is accurate to say that the raw data is reasonably clean to carry out socio-economic analysis as has been proven by the ever growing literature on this subject.

It is crucial to discuss the limitations of which there are quite a few. It is not easy to collect good quality radiance data in many parts of the globe despite best filtering mechanisms and sensors. Indeed, it is very difficult to do this when the weather itself creates hindrances to the collection of geospatial data. Cloud covers during monsoon season in tropical regions or solar illumination observed near the poles in summer months are examples of seasonal events that seriously affect the collection of data of sufficient quality. Therefore, we must not assume a value of zero when no light is observed in the average radiance raster image without referring to the cloud free observations file. Even after referring to the observations file, one must understand that the radiance values collected may be a little different from the values collected at some other time for the same source of light. It is an intrinsic limitation, which we hope doesn't drastically affect our analysis. Obviously, the grounds for this optimism are the authoritative studies that have employed this data favorably till date.

It should be said, however, that the process of examining raster files for pixel level analysis is a time consuming process. There is no denying that the exercise of data refinement will yield better and appropriate results, but the time frame and the modest objective of this report is not suitable to undertake a task which will take a few months to be properly done. Hence, in this report we have worked with data processed by looking at cloud free composites alone.

Link between economic activity and nightlights

The nightlight radiance data has been of huge interest to economists and other social science researchers since the beginning. Although the link between nightlight and economic activity is intuitively clear, it is still very hard to swallow that a complex exercise such as GDP estimation can very well be predicted using a proxy. A very convincing study in this regard was published in 1997 by Elvidge and others to approximate population, GDP in terms of purchasing power parity (PPP) and electricity usage by analysing the lit areas. This study established for the first time a very significant statistical relationship between the observed nightlights and economic activities on the ground.

The paper by Doll and others published in 2006 took the analysis a step further. They studied the relationship between the gross regional product (GRP) of 11 EU countries and the states in the USA. What they were interested in was the elasticity between nightlights and GRP i.e. with one unit change in nightlight intensity what is the corresponding change in the GRP of given administrative area. The paper also suggested some interesting maneuvering of data in order to avoid sweeping generalizations. One of the things that must be realized and the paper by Doll formalized it is that in different geographies, the elasticity of estimated income and nightlights will be different. Also, in highly populated urban areas, the elasticity will be more as opposed to sparsely populated areas away from city centres. In other words, one consistent elasticity relationship between the two will not apply at national level, or for that matter, at state level. This result is especially pertinent for the subcontinent. The study cautions us to estimating the regional income using nightlights data where official income estimates are either absent or very poor.

Ghosh et al 2010 provided a few more insights into the way nightlights data is correlated with specific economic activities. The activities once broadly divided into three categories i.e services, manufacturing and agriculture, makes it easy to account for different relationships and elasticities between different class of economic activities vis a vis nightlights. One of the primary assumptions of the paper is that agricultural output doesn’t correlate very well with nightlights while the other two categories correlate rather well. This again is a very instructive piece of insight with regards to a methodology to estimate GDP using nightlights data.

The studies we have reviewed thus far in this section actually use single year cross section analysis. For instance, Ghosh et al group together different administrative areas on the basis of similar ratio between nighttime lights and gross value added registered in national accounts. Thereafter, they estimate specific coefficients for each group using data of 2006 nightlight composites. Does it mean we shouldn’t use nightlights data for time series analysis? As mentioned earlier, the recorded radiance value may differ in different points of time due to seasonal factors and sensor settings that vary over time due to ageing equipment. This poses a problem from a time series analysis standpoint as observations may turn out to be egregious and inconsistent. However, the short answer is that one can indeed employ time series analytical techniques with nightlights data.

Henderson et al in 2012 showed a very robust way of carrying out panel data analysis with nightlight radiance to estimate income of administrative entities. The paper shows that by including a term for time-variant effects in the panel regression, the variation caused due to inconsistent observations can be captured by the time variant term, thus overcoming this limitation. The resulting specification takes the following form

$$Y_{it} = \alpha X_{it} + C_i + D_t + e_{it}$$

Where Y and X are GDP and observed radiance. C, D and e denote country specific effects, time variant effects and the error term.

Given the presence of such literature, the purported objective of this study seems to be justified, especially as we don’t have reliable estimates of income at all administrative levels. If we do possess this data, it is quite expensive to obtain. Even if we do obtain it, it comes with a considerable lag. Hence from multiple points of view, using nightlights data for estimation purposes seems like a legitimate exercise.

The present study takes the case of the state of Karnataka to see how appropriately can we employ nightlights data to estimate the income of small administrative entities such as district and taluk. We choose taluks of Bellary district as the test case for this study.

There are quite a few limitations of this study. Firstly, the data processed, on which there is more in the next section, has not been cleaned to a level sufficient to claim absolutely robust results. Secondly, most of the studies involving nightlights are carried out at a much bigger scale where many districts are used for regression models to figure out a range of elasticities and correlations with multiple demographic and economic factors. However, the time frame and the scope of this study are limited. The focus, hence, is to analyse a small dataset to throw interesting light on local issues.

The processing of data

The data files are hosted and produced by Earth Observations Group at National Oceanic and Atmospheric Administration, the link to which is given below.
https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html

The data collected by the VIIRS Day/Night band can be downloaded from the aforementioned site. It is available from April 2012 to April 2019. It is available in two forms

1) Average monthly composites
2) Average annual composites

The annual composites are available only for the years 2015 and 2016. For the rest of the years, one will have to process the data of each month separately and annually aggregate depending on the purpose and analysis. In fact, the datafiles are huge and require considerable computational facility and time to be processed.

The zip files are available in the form of six tiles that refer to different parts of the globe. We are interested in tile 3 which contains the raster image of India and the countries of East and South Asia. From the zip file, we obtain two raster images. One of the files contains the average radiance data and the other contains the number of cloud free observations recorded by the VIIRS in the given period.

The two files are viewed and processed in QGIS. However, the radiance data in these files are stored on a grid represented by pixels. Each pixel denotes a resolution of about 750*750 m^2, which roughly translates to about 0.5 Km^2. This is a plausible resolution to do point source analysis as well.

We have to process the data in a way that can give us the cumulative radiance values of an area i.e India, Karnataka, Bellary, Hospet, in that order. For this purpose, we need vector files of administrative boundaries. It is challenging as administrative boundaries are redrawn after every census or at any other point. Otherwise, the raster data with input of outdated administrative boundaries may lead to egregious results.

GADM, the database of global administrative areas is one such resource which makes available the shapefiles (format used in GIS) of administrative areas at all levels i.e country, state, district and taluk. The only shortcoming of this data is that the administrative boundaries of India are drawn on the basis of the 2001 census boundaries. Hence, the changes in the administrative areas done after that are not reflected in these shapefiles. Any other source that can provide data of such comprehensive detail has not been identified yet. Nevertheless, preliminary research suggests that this data is indeed quite robust as not many changes have been made to administrative areas post 2001 census and without any manipulation we can use it for this report. For refinement, it is advisable that we use the latest shapefiles. One such shapefile of district Bellary as viewed in QGIS is shown below.

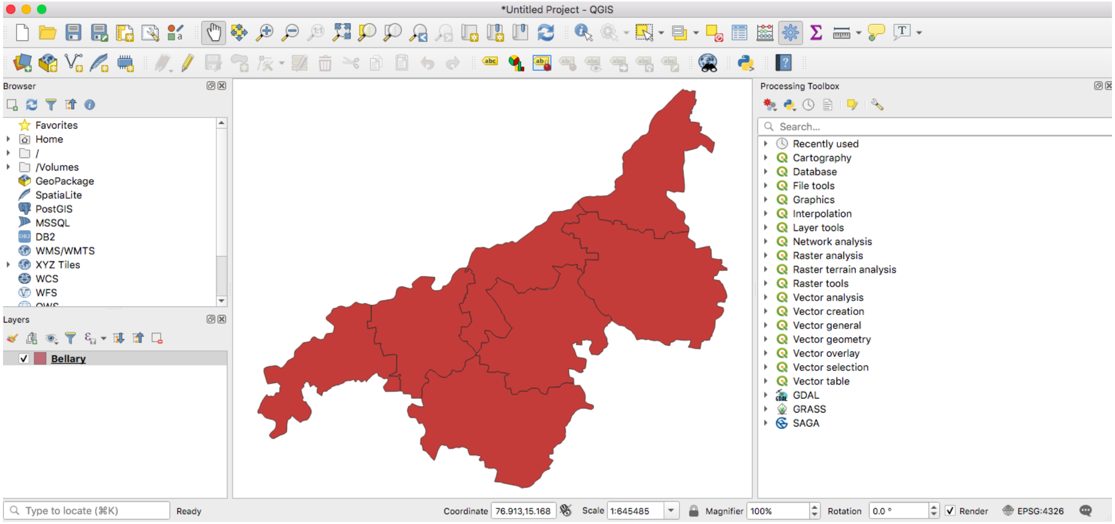


Image 1: District Bellary as visualized in QGIS
Source: Author's own

Once we acquire this data, using corresponding shapefiles, the raster files are cut to obtain images of India, Karnataka and other districts of Karnataka. For demonstration purposes, these images are shown below (Image 2 and image 3).

One can observe many such grey scale images and countless such images are processed this way, of different districts and states for different months. However, the real data is obtained when we process these images for pixel values. The pixel values quantify the bright cores and not so bright cores.

Using zonal statistics function in QGIS, one can do pixel level calculations. This is fairly intuitive to understand. If we take the district of Bellary as an example, when mapped with the nightlights image, it occupies a certain number of pixels on the grid. Every pixel has an observed radiance value. If you add all the pixels that denote Bellary, you get the radiance emitted from bellary in a particular month.

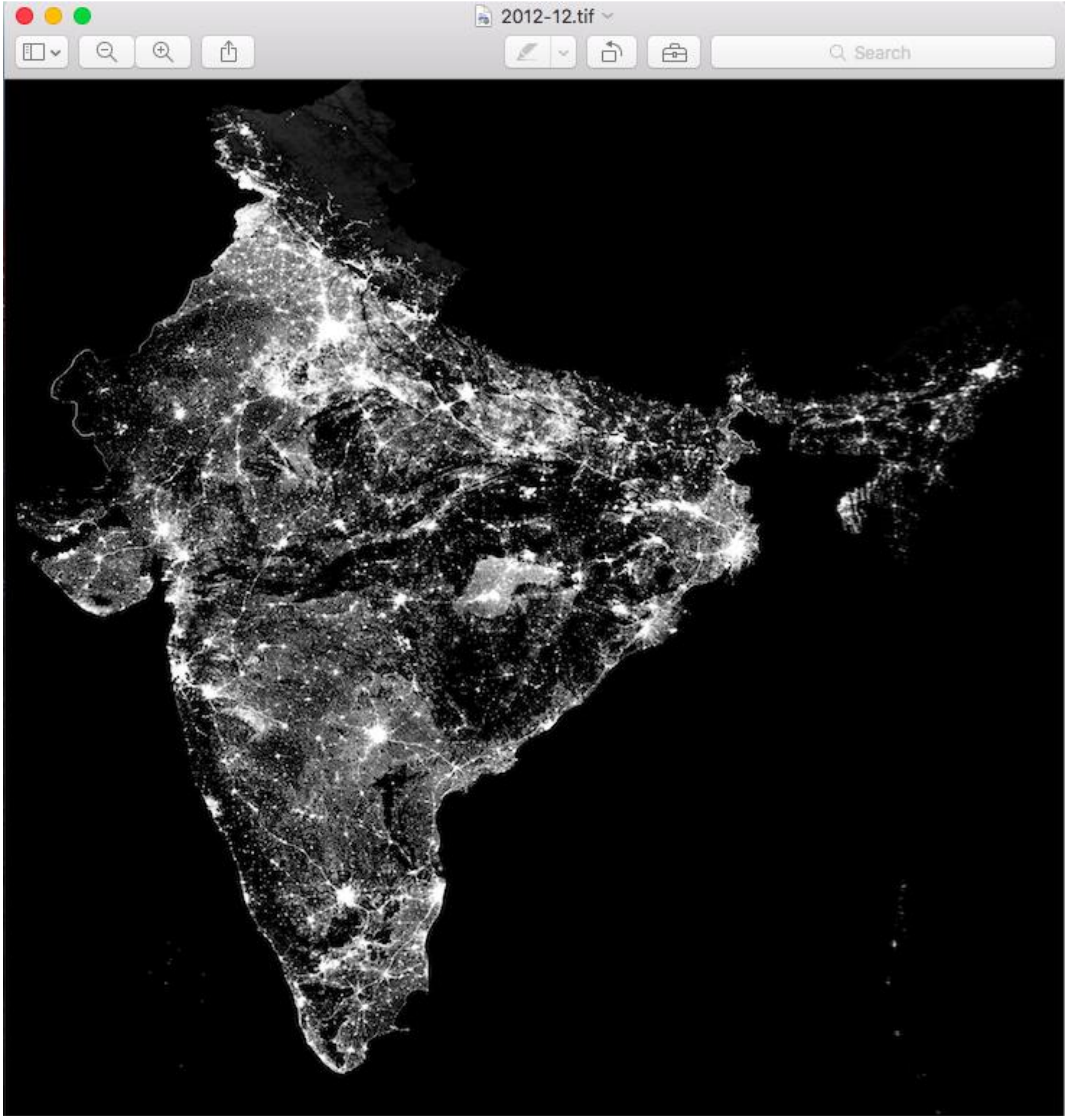


Image 2: Nightlight composites of India for December, 2012 as obtained from the corresponding raster file.

Source: Author's own

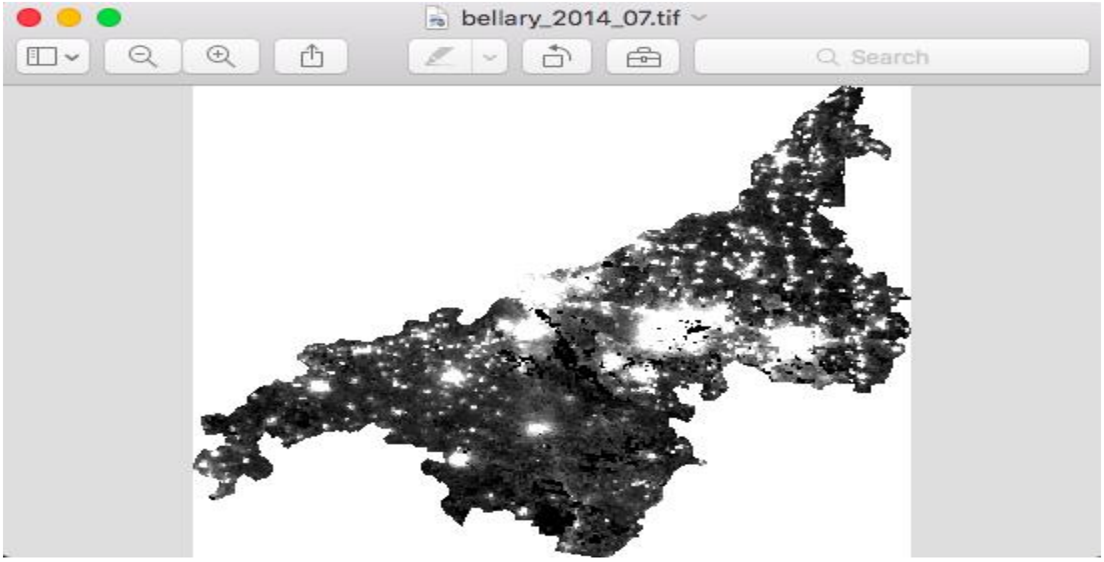


Image 3: Bellary in July, 2014

Source: Author's own

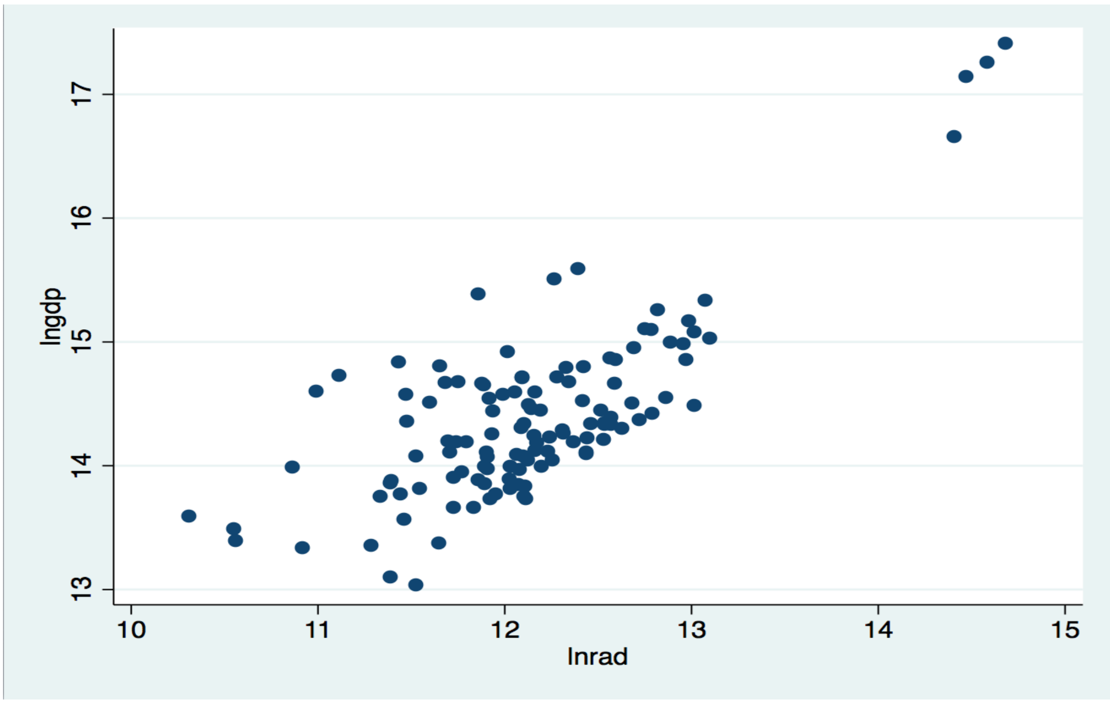
The results of zonal statistics functions are reflected in the attributes of the shape file. These attributes can be exported as CSV dump or excel file. This process is repeated for all the months for the years 2012, 2013, 2014 and 2015 for all the districts of Karnataka, thus giving us a sizable cross sectional data across a time period of 4 years. This data can indeed be expanded up to 7 years. A longer time series can also be constructed by using the data collected by earlier satellites, namely, DMSP-OLS, for radiance from the year 1992 to 2011.

Nightlights and district level GDP

The first thing we need to establish is whether or not gross district domestic product shares any correspondence with the nightlights. For this purpose, we process nightlights data from year 2012 to year 2016. Since the data for the first 3 months of 2012 is not available, we impute the missing values by average radiance values for harmonizing the dataset. The district domestic product is collected from various annual economic surveys published by the Government of Karnataka.

A simple scatter plot (figure 1) for the dataset including values from 2012 to 2015 gives us the initial clue if the two share a strong correlation. The log value of GDDP and radiance are plotted to check graphically if there is a possibility to fit a linear model or otherwise to the data points. Except for a few outliers, the two variables show a strong linear relationship. One can imagine a straight line passing through the cluster of dots which will explain a large part of the data.

Figure 1: Log values of GDP Karnataka districts as plotted against respective radiance values



Source: Author's own

Having confirmed a graphical relationship between the two variables, we next check for correlation between the two, at level and log values.

Table 1: Correlation coefficients

Correlation	GDDP	Radiance
GDDP	1.000	
Radiance	0.9582	1.000

Correlation	Lngdp	Lnrad
Lngdp	1.000	
Lnrad	0.7687	1.000

Source: Author's own

We see a very high correlation between the two variables, GDDP and radiance. The log transformation of the variables also show high correlation, as shown in the tables below.

Having run the initial checks on the dataset, our next objective is to find a model that can formalize the relationship between the two variables in a mathematical form. In this regard, we will first check the results using the three models. The first one is a pooled OLS model, second is a fixed effects, time invariant panel model and the third is a random effects model. The difference between the fixed effects and random effects model is that we assume the error term is uncorrelated with the independent variable for random effects.

The specification of the three models are given below:

Fixed effects & Random effects model:

$(\text{lngdp})_{it} = \beta_1(\text{lnrad})_{it} + \alpha_i + u_{it}$

where

– α_i (i=1,2,3...,n) is the unknown intercept for each state (n district-specific intercepts)

– lngdp_{it} is the dependent variable where i = state and t = year

– lnrad_{it} represents one independent variable

– u_{it} is the error term

Pooled OLS model

$(\text{Lngdp})_{it} = a + b(\text{Lnrad})_{it} + e_{it}$

Where lngdp and lnrad are log values of the variables gddp and radiance respectively, e denotes the error component and a denotes the constant component. The results as obtained in stata are given below in table 2.

Table 2: Results from the three models

	Pooled OLS model	Fixed Effect Model	Random effects model
Coefficients	Lnrad = 0.7831* (0.599905) Cons = 4.8847* (0.7298)	Lnrad = 1.0603* (0.0982) Cons = 1.5179** (1.1936)	Lnrad = 0.9376* (.076) Cons = 3.0082* (.9294)
Prob > F	0.0000	0.0000	NA
Prob > Chi2	NA	NA	0.0000

MSE	0.45847	NA	NA
Rho value	NA	0.8501	0.8190

* = significant at 99% CI, ** = not significant at 90% CI

The reasons underlying the choice to use these two models in place of other models are well documented in literature. One reason being the small size of the sample, where T is 4 and N is 30. The other is that using models such as LSDV will be problematic from the point of view of interpretation. Using district dummies will not only complicate the model by introducing too many regressors, it will also not improve the predictive ability of the model. The same goes for year dummies.

What we have done by using fixed effects model is make the model simple and assume that one single slope can explain the data sufficiently. Had we constructed a much longer series (observations across more number of years) with the same cross section of districts, we could have grouped similar districts together. That would have given us the room and flexibility to test assumptions such as different regression coefficients and intercepts for different groups, same coefficient but different intercepts for different groups and other such assumptions about the error structure. Because our sample is not so big, the literature suggests that pooled OLS and fixed effects are the most appropriate approaches to model the data.

Interpreting the fixed effects model, the coefficient and average value of intercept are 1.060324 and 1.517951 respectively. The xtreg, fe estimates by OLS and hence its reported R^2 within has all the properties of the usual R^2, which is the ratio of explained sum of squares(ESS) to the total sum of squares (TSS).

Using the estimation of coefficients provided by the two models, we do and out of sample prediction for the year 2016 since we do have values to compare the prediction with.

Comparison between observed and predicted values

District	Year	Lngdp	FE prediction	Pooled OLS prediction
Bagalkot	2016	14.89242242	14.47638526	14.45568081
Bangalore	2016	17.55412497	17.07896911	16.37791315

Bangalore Rural	2016	14.43458054	14.89665631	14.76608719
Belgaum	2016	15.4455351	15.26321958	15.03682573
Bellary	2016	15.16586681	15.28334319	15.05168875
Bidar	2016	14.33662309	14.41527974	14.41054912
Bijapur	2016	14.60900069	14.69623226	14.61805676
Chamrajnagar	2016	14.10339261	13.76889598	13.93313905
Chikballapura	2016	14.28607653	14.28886849	14.31718353
Chikmagalur	2016	14.76508132	13.75135739	13.92018529
Chitradurga	2016	14.38230055	14.60042329	14.54729358
Dakshina Kannada	2016	15.72049454	14.48863472	14.4647281
Davanagere	2016	14.59097478	14.6259747	14.5661655
Dharwad	2016	14.84194634	14.29084987	14.31864695
Gadag	2016	14.04224764	13.6725571	13.86198449
Gulbarga	2016	14.73208125	15.09845781	14.91513497
Hassan	2016	14.77405586	14.48231725	14.4600621
Haveri	2016	14.37701028	13.94171385	14.06077993
Kodagu	2016	13.46495968	12.33189287	12.87178848
Kolar	2016	14.52191941	14.74765853	14.65603949
Koppal	2016	14.10008782	14.24786044	14.28689555
Mandya	2016	14.8970106	14.54638136	14.50737896
Mysore	2016	15.18116844	15.21086607	14.99815816
Raichur	2016	14.52009212	14.8682548	14.74511022
Ramanagara	2016	14.42587378	14.4348737	14.42502095
Shimoga	2016	14.9780049	13.94170836	14.06077587
Tumkur	2016	15.32165249	15.14629008	14.95046322
Udupi	2016	14.94581529	13.361738	13.63241781
Uttara Kannada	2016	14.5625302	13.71131285	13.89060894
Yadgir	2016	13.92400615	14.15757378	14.22021107

Table 3

The choice between fixed and random effects model will be dictated by Hausman test. However, firstly we have to decide between fixed/random and pooled OLS. For this purpose, we use Breusch-Pagan Lagrange multiplier (LM) test. The results are provided in table 3 below.

Table 3: Model diagnostic tests

	Value of the statistic	Inference
LM test	Chibar2 = 113.94	Null hypothesis rejected. Random effects model accepted over pooled OLS model
Hausman test	Chi2(1) = 3.92	Null hypothesis rejected. Fixed effects model accepted over random effects model

Once we have a reliable model, we can use it to predict district level GDP. Indeed, relying on just nightlights to arrive at income estimates of a district is not advisable. For improving the district GDP estimates one will have to use different kinds of data in addition with nightlights. However, nightlights data is available at high frequency (monthly) and with appropriate changes to methodology, can give us very early indication of district GDP estimates.

Taluk level estimates

The objective of the study is to arrive at indicative taluk level income estimates. Since we have been able to arrive at the district GDP using nightlights, how do we approximate taluk GDP?

One of the ways in which we can calculate taluk income is by allocating the district level income into different taluks since income of all the taluks together make up the income of the district. Since we don't have other estimates of taluk level income, we can't employ the methodology of estimating district level income. Also, the intensity of nightlight doesn't translate to equal economic activity in one place or in different measures. Hence, we built our model using official statistics and checked the specific relation between nightlights and income in 30 districts in the state of Karnataka.

One simple allocation formula can distribute district GDP among all the taluks on the basis of nightlight ratios. The formula will take the following mathematical form

$(SOL_{taluk})/(SOL_{district}) = (GDP_{taluk})/(GDP_{district})$

Another allocation strategy can be based on several other considerations such as size of primary sector in the district economy and total cultivated area proxied as total landholdings in the taluk. One such simple formula might look like

$GDP_{taluk} = (SOL_{taluk}) / (SOL_{district}) * (GDP_{district})^{(s+t)} +$
$(Landholdings_{taluk})) / (landholdings_{dist}) * (GDP_{district})^{agri}$

This allocation strategy can be refined to develop a better allocation formula that is more reflective of the economic activities taking place in different taluks and districts. Using the formulas discussed above one can arrive at a rough estimate of taluk level income. There is a possibility here to think out of the box and construct a formula that includes most representative factors of economic activity at the taluk level. Population data can also be used to improve the estimation.

Major Conclusions and Implications

The objective this study set out to achieve was to investigate the potential of nightlights data to accurately capture the extent of economic activity in India, particularly at lower levels of administration such as district and taluk. The study is able to present and summarise the various methodologies developed in the relevant literature so far. By processing the nightlights data first hand, the study has been able to hold forth on the limitations of radiance data, both within itself and its ability to represent economic activity as such. In conclusion, it can be said that nightlights data can be used as an indicator of economic activity and most certainly can be used to provide rough estimates of GDP. The nightlights sample used in this study shows a remarkable correspondence with official estimates. However, the prediction based on just 3 years of nightlights data has not proved sufficient. It is recommended one constructs a longer time series using data captured with DMSP satellites in conjunction with data made available by SUOMI satellite. What is also recommended is that in addition to nightlights data one employs taluk and district level land use data to improve the robustness and precision of models. The successful implementation of these methodologies in practice will mean that the government can publish the provisional income figures at the earliest. Also, it implies that an early discovery about rise and fall income, especially after economic episodes such as demonetisation, COVID-19 etc., can hugely aid the process of resource allocation and planning, and improve the interventions by government prompt and effective.

References

Beyer, Robert C. M., Esha Chhabra, Virgilio Galdo, Martin Rama. 2018. 'Measuring districts' monthly economic activity from outer space.' Policy Research working paper no. WPS 8523. Washington, D.C: World Bank Group. http://documents.worldbank.org/curated/en/835491531401292135/pdf/WPS8523.pdf (*accessed on 14th Jan '20*)

Bundervoet, Tom, Laban Maiyo and Apurva Sanghi. 2015. 'Bright Lights, Big Cities: measuring national and subnational economic growth in Africa from outer space, with an application to Kenya and Rwanda'. Policy Research Working Paper 7461. Washington DC: The World Bank. http://documents.worldbank.org/curated/en/912151468188369841/pdf/WPS7461.pdf (*accessed on 14th Jan '20*)

Chen, Xi, and William D. Nordhaus. 2011. 'Using Luminosity Data as a Proxy for Economic Statis- tics.' *Proceedings of the National Academy of Sciences*, vol. 108, no. 21, p.p. 8589–94. https://www.pnas.org/content/108/21/8589 (*accessed on 14th Jan '20*)

Chodorow-Reich, Gabriel, Gita Gopinath, Prachi Mishra and Abhinav Narayanan. 2019. 'Cash and the Economy: Evidence from India's Demonetization', NBER working paper 25370. The National Bureau of Economic Research. https://www.nber.org/papers/w25370 (*accessed on 14th Jan '20*)

Croft, Thomas A. 1978. 'Night-time Images of the Earth from Space.' *Scientific American*, vol. 239, p.p. 68–79. http://adsabs.harvard.edu/abs/1978SciAm.239a..86C (*accessed on 14th Jan '20*)

Database of Global Administrative Areas. 'GADM database'. https://gadm.org/download_country_v3.html (*accessed on 14th Jan '20*)

Dehejia, Vivek, Praveen Chakravarty. 2017. 'India's income divergence: Governance or development model?' IDFC Institute Briefing Paper 5. IDFC Institute. http://www.idfcinstitute.org/site/assets/files/11015/briefing_paper_april_11_for_web.pdf (*accessed on 14th Jan '20*)

Doll, Christopher N.H. 2008. 'CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications.' Center for International Earth Science Information Network, Colombia University, USA https://pdfs.semanticscholar.org/ff03/200355ce4f6f48a3c03fabc479e7fc46fa0f.pdf (*accessed on 14th Jan '20*)

Elvidge, Christopher D., Kimberley E. Baugh, Eric A. Kihn, Herbert W. Kroehl, Ethan R. Davis, and C. W. Davis. 1997. 'Relation between Satellite Observed Visible-Near Infrared Emissions, Popula- tion, and Energy Consumption.' *International Journal of Remote Sensing*, vol. 18, p.p. 1373–79. https://www.tandfonline.com/doi/abs/10.1080/014311697218485?tab=permissions&scroll=top (*accessed on 14th Jan '20*)

Elvidge, Christopher D., Kimberly E. Baugh, Paul C. Sutton, Tilottama Ghosh. 2012. 'The Night Light Development Index (NLDI): a spatially explicit measure of human development from satellite data' *Social Geography*, vol. 7, no. 1, p.p. 23-35. http://www.soc-geogr.net/7/23/2012/sg-7-23-2012.html (*accessed on 14th Jan '20*)

Elvidge, Christopher D, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu and Tilottama Ghosh. 2017. 'VIIRS night-time lights'. *International Journal of Remote Sensing*, vol. 38, no. 21, p.p. 5860-5879 https://www.tandfonline.com/doi/full/10.1080/01431161.2017.1342050 (*accessed on 14th Jan '20*)

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer. 2015. 'The Next Generation of the Penn World Table'. *American Economic Review*, vol. 105, no.10, p.p. 3150-3182. https://www.rug.nl/ggdc/docs/the_next_generation_of_the_penn_world_table.pdf (*accessed on 14th Jan '20*)

Ghosh, Tilottama, Sharolyn Anderson, Rebecca L. Powell, Paul C. Sutton and Christopher D. Elvidge. 2009. 'Estimation of Mexico's Informal Economy and Remittances Using Nighttime Imagery.' *Remote Sensing* vol. 1, no. 3, p.p. 418–44. https://pdfs.semanticscholar.org/fcd4/c6c907fb5f5d4dc0baefc8f2452174103784.pdf (*accessed on 14th Jan '20*)

Ghosh, Tilottama, Sharolyn J. Anderson, Christopher D. Elvidge and Paul C. Sutton. 2013. 'Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being.' *Sustainability*, vol. 5, no. 12, p.p. 1-32. https://pdfs.semanticscholar.org/f87d/16bcd1f3be2827b1736c91bd665ce2575453.pdf?_ga=2.168066913.1710461712.1568507077-1597104448.1567708687 (*accessed on 14th Jan '20*)

Government of India. 'An overview of Indian System of national accounts.' Ministry of Statistics and Programme Implementation. http://www.mospi.gov.in/132-overview-indian-system-national-accounts (*accessed on 14th Jan '20*)

Henderson, J. Vernon, Adam Storeygard and David N. Weil. 2012. 'Measuring Economic Growth from Outer Space'. *American Economic Review*, vol. 102, no. 2, p.p. 994-1028. https://www.jstor.org/stable/pdf/23245442.pdf?refreqid=excelsior%3A75389da410745a960893dd3cdab588f7 (*accessed on 14th Jan '20*)

Tanaka, Kiyoyasu and Souknilanh Keola. 2017. 'Shedding Light on the Shadow Economy: A Nighttime Light Approach'. *Journal of Development Studies*, vol. 53, no. 1, p.p. 32-48. https://ideas.repec.org/a/taf/jdevst/v53y2017i1p32-48.html (*accessed on 14th Jan '20*)

Lequiller, Francois and Derek Blades. 2014. 'Understanding National Accounts.' Second Edition, OECD Publishing. http://www.oecd.org/sdd/UNA-2014.pdf (*accessed on 14th Jan '20*)

Mukhopadhyay, Abhiroop, David G. Urzainqui. 2018. 'The dynamics of spatial and local inequalities in India.' WIDER working paper 182. World Institute for Development Economic Research. https://www.wider.unu.edu/publication/dynamics-spatial-and-local-inequalities-india (*accessed on 14th Jan '20*)

National Centre for Environmental Information. 'Version 1 VIIRS Day/Night band Nighttime Lights'. National Oceanic and Atmospheric Administration. https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html (*accessed on 14th Jan '20*)

OECD. 2017. 'The role of national statistical systems in the data revolution'. *Development Co-operation Report: Data for Development*. OECD Publishing, Paris. https://www.oecd-ilibrary.org/development/development-co-operation-report-2017/the-role-of-national-statistical-systems-in-the-data-revolution_dcr-2017-8-en;jsessionid=egn0ZidqFfhUXnxJPsKwpy4r.ip-10-240-5-174#targetText=National%20statistical%20systems%20are%20ensembles,and%20disseminate%20%E2%80%9Cofficial%20statistics%E2%80%9D. (*accessed on 14th Jan '20*)

Subash S P, Rajeev Ranjan Kumar, K S Aditya. 2018. 'Satellite data and machine learning tools for predicting poverty in rural India.' *Agriculture Economics Research Review*, vol. 231, no.2, p.p. 231-240. http://www.indianjournals.com/ijor.aspx?target=ijor:aerr&volume=31&issue=2&article=007 (*accessed on 14th Jan '20*)

Subramanian, Arvind. 2019. 'India's GDP mis-estimation: Likelihood, magnitudes, mechanisms and implications.'CID Faculty Working Paper no. 354. Centre for International Development. Harvard University. https://www.hks.harvard.edu/centers/cid/publications/faculty-working-papers/india-gdp-overestimate (*accessed on 14th Jan '20*)

World Bank. 2017. 'Growth Out of the Blue'. South Asia Economic Focus. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/28397 (*accessed on 14th Jan '20*)