Agricultural Economics Research Review 2018, 31 (2), 231-240
DOI: 10.5958/0974-0279.2018.00040.X

Satellite data and machine learning tools for predicting poverty in rural India

S P Subash[a]*, Rajeev Ranjan Kumar[b] and K S Aditya[c]

[a]ICAR-National Institute of Agricultural Economics and Policy Research (NIAP), New Delhi-110012, India
[b]ICAR-Indian Agricultural Statistical Research Institute (IASRI), New Delhi-110012, India
[c]ICAR-Indian Agricultural Research Institute (IARI), New Delhi-110012, India

Abstract Increased availability of satellite imageries and rapid development in algorithms to process imagery data has spurred interest amongst economist to use high frequency imagery data for meaningful economic interpretations. One such application is to use satellite night light data as an indicator of poverty. As poverty statistics in India is released once in five years, high frequency night lights data can be used to predict poverty in the years where official poverty statistics is not available. In this paper, we explored use of satellite night light data and machine learning algorithms (Artificial Neural Network) to predict rural poverty at sub-national level, i.e. state. We compared night light data with per capita domestic product as a predictor for the model. We find night light data as a better predictor of poverty than of per capita domestic product. Such predictions using satellite data can be used as a complement to the existing data-sets. This will facilitate economist for modelling the economic relationships in understanding poverty and provide more frequent and local estimates for policy makers.

Keywords Poverty, Night light, Satellite data, ANN, Machine learning

JEL classification I30, I32, C00

1 Introduction

Eradication of extreme poverty in its all forms is the global concern as has been indicated in the Millennium Development Goals and the Sustainable Development Goals (SDGs) of the United Nations. However, the non-availability of the time-series data on economic livelihoods particularly in the developing countries is a key challenge in devising, implementing and monitoring poverty alleviation programs measuring their outcomes (Jean et al. 2016). Traditionally, data on poverty rates are generated through nationally representative sample surveys on income and expenditure. In India, the National Sample Survey Office (NSSO) of the Ministry of Statistics and Program Implementation of the Government of India undertakes quinquennial surveys on household consumption expenditure, which are used to measure poverty.

*Corresponding author: subashspar@gmail.com

Nevertheless, there are issues of reliability and transparency of the data on account of outdated methods, infrequent estimates and political sensitivity (Devarajan 2013). India's poverty statistics is also prone to such criticisms (Deaton & Kozel 2005). Much of the criticism is regarding the income/expenditure cut-off to decide the poverty line (Pangariya & Mukim 2014). A comparison of the expenditures from the NSS surveys and National Accounts Statistics (NAS) points towards considerable gap between the two estimates (Deaton 2005; Pangariya & Mukim 2014). Further, poverty estimates are available only for the years, corresponding to the consumer expenditure surveys. This limits their use in econometric modelling to assess effectiveness poverty reduction programs.

In India, there is a paucity of reliable and high frequency data, which is required to design economic and development policy interventions (Suraj et al.

2017). The Census data is decadal and lack validation. On another hand, the data from sample surveys that are claimed to be more accurate are infrequent. Use of predicted data from satellite imagery as a viable alternative to real data for development indicators was proposed in various studies (Duan et al. 2017; Dugoua et al. 2018; Ghosh et al 2013; Nischal et al. 2015; Suraj et al. 2017). The studies done so far are mostly limited to cross sectional data at national level. With this background, the objective of this paper is to address the issues of data availability by exploring the alternative means of data generation, such as satellite night light data. We use this data-set to measure and predict the poverty rate at sub-national level (state). We also explore whether other data such as per capita domestic product (which is available on yearly basis) could be used instead of night light data.

2 Measuring poverty

Poverty is measured for making comparisons (Ravallion 1994). A comparison could be ordinal or cardinal. While, an ordinal comparison helps us to understand change in poverty over time and space, cardinal comparison quantifies poverty and measure its extent across programs. Another important conceptual distinction regarding measuring poverty relates to its underlying approaches viz., 'welfarist' and 'non-welfarist' approaches. The welfarist approach relies solely on individual utilities (goods and service consumed), while the non-welfarist approach focuses on elementary achievements (education, health, etc.) (see, Sen 1981). Hence, conceptual clarity is important while designing methods and approaches to measure poverty.

There are various measures used to measure poverty: income-based, consumption-based, nutrition-based, anthropological, and multi-dimensional poverty index (Alkire et al. 2017). Each measure has its own advantages and dis-advantages. In India, we follow consumption-based approach to generate poverty estimates. The incidence of poverty (head-count ratio) is calculated as negative deviation from a threshold level of consumption expenditure or poverty line on the assumption that there is a pre-determined and well-defined level of standard of living. Although this approach is criticized (see, Himanshu 2010), it remains main approach to measure poverty.

In recent years, new approaches to poverty estimation have been proposed to overcome data limitations and measurement issues. The judgment regarding the best method of measuring poverty relies on "data availability" (Ravillion 2001). There are alternatives methods of using passively collected data for measuring economic outcomes, such as poverty (Jean et al. 2016). A popular approach is to use satellite night light data (Henderson et al. 2012; Michalopoulos & Papaioannou 2013; Pinkovskiy & Sala-i-Martin 2016). Many studies have explored the linkages between night light and poverty (see, table 1). Literature suggests that the satellite night light data combined with the machine learning approach can be used to predict economic indicators (Jean et al. 2017, Suraj et al. 2017).

We find most studies predicting poverty at national level using night light data provided by the Defense Meteorological Satellite Program - Operational Linescane System (DMSP-OLS) and using approaches like convolution neural networks for mining data and correlation, regression (instrumental variable, panel, ridge) to explore the relationships. Some studies have also used data mined from google maps and other satellite data (Duan et al., 2017; Jean et al. 2017). These studies mined data using deep learning tools and used statistical models for prediction.

3 Data

The data for this study have been collected from an open access night light data provided by the University of Michigan. The dataset named 'India Lights API' provides rural night light data for a period of 20 years, from 1993 to 2013 for about 6,00,000 villages across India (http://india.nightlights.io/#/about). The data was extracted from the satellite pictures of the earth for every night under the Defense Meteorological Satellite Program (DMSP) of the U.S. Department of Defense.

The DMSP provides raster images with a resolution of 30 arc-seconds (roughly 1 square kilometer at the equator). The DMSP-OLS satellite sensors record relative brightness of night light on a 0-63 DN scale. These values are not calibrated and thus don't reflect an absolute measure of luminosity. The luminosity recorded is converted into a single measure by extracting the raster image for a specific time for a region corresponding latitude and longitude coordinates. The luminosity recorded on individual

Table 1. Studies that have used satellite data to predicting poverty

Author	Country	Satellite data	Duration/period of study	Model/Analysis
Akiyama (2012)	Japan	Night light data of the DMSP/OLS	NA	Correlation, Multiple regression
Chen and Nordhaus (2011)	Multiple	Yale G-Econ dataset	1992-2008	Correlation, IV regression (log forms)
Duan et al. (2017)	Bangladesh, India	Landsat-8, Sentinel-1, DMSP-OLS, VIIRS	2013, 2015	Convolutional neural network (CNN) models, Ridge regression
Dugoua et al. (2018)	India	Night light data of the DMSP/OLS	1992 -2013 (2011analysis)	Correlation, Multivariate regression (log transformation)
Elvidge et al. (2012)	Global	Night light data of the DMSP/OLS	2006	Regression
Ghosh et al. (2013)	Global	Night light data of the DMSP/OLS	2006	Regression
Henderson et al. (2009)	Africa	Night light data of the DMSP/OLS	1992-2013	Log linear panel regression
Jean et al. (2016)	Africa	Google, Night light data of the DMSP/OLS	2012	Convolutional Neural Network (CNN) models, Ridge regression
Mellander et al. (2015)	Sweden	Night light data of the DMSP/OLS	2000-12	Correlation and geographically weighted regression
Nischal et al. (2015)	India	Night light data of the DMSP/OLS	2000-12	Multivariate regression and ARIMA
Noor et al. (2008)	Africa	Night light data of the DMSP/OLS	Multiple years	Pearson correlation, spearman rank correlation, Kappa coefficient
Suraj et al. (2017)	India	Night light data of the DMSP/OLS, Google static maps	Multiple years	Deep CNN model, transfer learning

Source: Compiled by authors

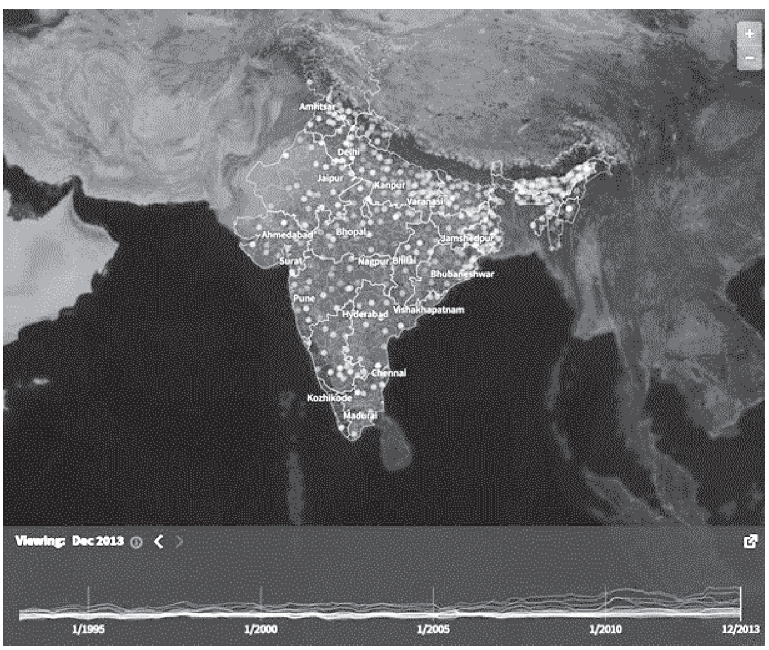

Figure 1. Night light data of India

Note: The image is at a specific point of time (December 2013).

Source: Night Light India API

images are often affected by noise and atmospheric disturbances, which are removed using algorithms. Note, cloud cover and solar glare (according to recommendations from the National Oceanic and Atmospheric Administration) occasionally result in negative values, which can be interpreted as implying statistical confidence that the areas are dark. These are areas where any brightness is below the detection sensitivity of the satellite sensors. There are about 4.4 billion data points (each day for different regions over 20 years) which were aggregated by taking the monthly median measurement for each village, district, and state. The differences in satellite are controlled using a stepwise calibration approach (Li & Zhou 2017). A detailed description on the data, extraction and aggregation is provided in the site mentioned above. India Light API provides data in JSON format which we have converted into an excel format file. The monthly data are aggregated as annual data by taking median values of months in that year.

We have also collected available data on rural poverty estimates at state level (GoI 2014). We have also used data on GDP to predict poverty at state level using the same algorithm used for predicting poverty with night light data. We calculated per capita income by dividing the gross domestic product of state by the population of the state. The per capita income was used as benchmarking data. There is a negative correlation relation between income and poverty, hence it could also be used as a predictor of poverty.

4 Empirical model

We build two models, one with night light as a predictor and another with per capita income. The structural relationship between poverty and predictor can be written as:

$$P_{it} = \alpha + \beta_1 X_{it} + \beta_2 S_t + \beta_3 T_i + \varepsilon_{it} \qquad \ldots(1)$$

Where, P_{it} is the measure of poverty in state I (i= 1,2,3…N) at time t (t=1,2,3…T), X_{it} is the predictor variable (night light or per capita income). S_t is the state dummy for t^{th} time, and T_i is the time dummy for i^{th} state.

4.1 Machine learning techniques

For predicting poverty, we used machine learning (supervised) based regression approach. The application of machine learning techniques with satellite data for econometric analysis is of recent origin (Suraj et al. 2017). Unlike statistical modeling that formalizes relationship between variables based on mathematical equations and rely on the rule based programming, machine learning uses algorithms to learn from data to understand and predict the data. Standard statistical models (like OLS) are not optimized for prediction because these are more focused on unbiasedness (Kelienberg et al. 2015). Machine learning techniques maximize prediction performance by accounting bias-variance trade-offs. Machine learning is stated to be a superior approach in prediction compared to the statistical ones (Makridakis et al. 2018).

Machine learning (ML) techniques are flexible and can handle complex and non-linear relationships. A frequently used ML algorithm in empirical research is the Artificial Neural Network (ANN). ANN have been applied for modelling and prediction in several fields of science for example, mathematics, economics, engineering, medicine, hydrology, meteorology and psychology (see Mair et al. 2000; Cruz & Wishart 2006). The popularity of ANN has grown since their inception in 1943 (McCulloch & Pitts 1943), mainly to solve prediction problems with variables of stochastic nature, nonlinear or unknown variations. As compared to the traditional model-based methods, ANN are data-driven self-adaptive methods in which there are few a priori assumptions about the models (Zhang et al. 1998). ANN is a powerful and versatile technique for capturing and representing complex input and output relationships. These learn from examples and capture suitable functional relationships among the data even if the underlying relationships are unknown or hard to describe. Thus, ANNs are well suited for problems whose solutions require knowledge that is difficult to specify but for which there are enough data or observations.

Artificial neural network model can be classified in terms of its architecture and learning algorithm. The architecture or topology describes the neural connections, and the learning algorithm provides information on how the ANN adapts its weight for every training vector. It is a computational paradigm composed of non-linear elements (neurons) operating in parallel and massively connected by networks characterized by different weights. A single neuron computes the sum of its inputs, adds a bias term, and drives the result through a generally nonlinear activation function to produce a single output. ANN models are specified by network topology, neuron characteristics, and training or learning rules with inputs, output(s) and hidden layers with interconnections.

Instead of testing the sensitivity of each predictor input in the training algorithm, hidden transfer function and output transfer function are varied systematically. As there is no mathematical formula to determine the optimum number of hidden layers, the number of neuron in hidden layer was decided by trial and error method. In order to determine the optimum architecture, combinations of the input, hidden layer and output were tried one by one. All the analysis of ANN has been performed in "R" software.

The schematic representation of how we used machine learning approaches for predicting poverty data is given in Figure 2. From the data set on poverty (Y) and predictor variables (X) [either night light or per capita GSDP], we used a sub-set of data for which the poverty data is known to train and develop a model using machine learning algorithm. The developed model is then used on another subset of data, for which the poverty is not known, as an input data to predict poverty.

4.2 Comparing night light and income data for predicting poverty

There are no standard approaches to compare data predicted from different techniques. So, we follow graphical approaches and measures used in comparing the imputed values. One important measure is the descriptive statistics of the predicted values and to plot these values (Kdenisty plots, box plot and violin plot) (Nguyen et al. 2017). These graphical and descriptive statistics of the predicted values are indicative of whether the predicted data are reasonable. Such judgements about the reasonableness of the predicted data are regarded as 'external checks' (Abayomi et al. 2008). However, studies have shown that the predicted values may or may not fall within plausible or possible ranges (von Hippel 2013). Another approach is to compare observed and predicted data as an 'internal check' (Abayomi et al. 2008). The commonly used internal check is graphical analysis; boxplots (White et al. 2011), density plots (Abayomi et al. 2008) to compare observed and predicted values (Abayomi et al. 2008, Stuart et al. 2009, White et al. 2011, van

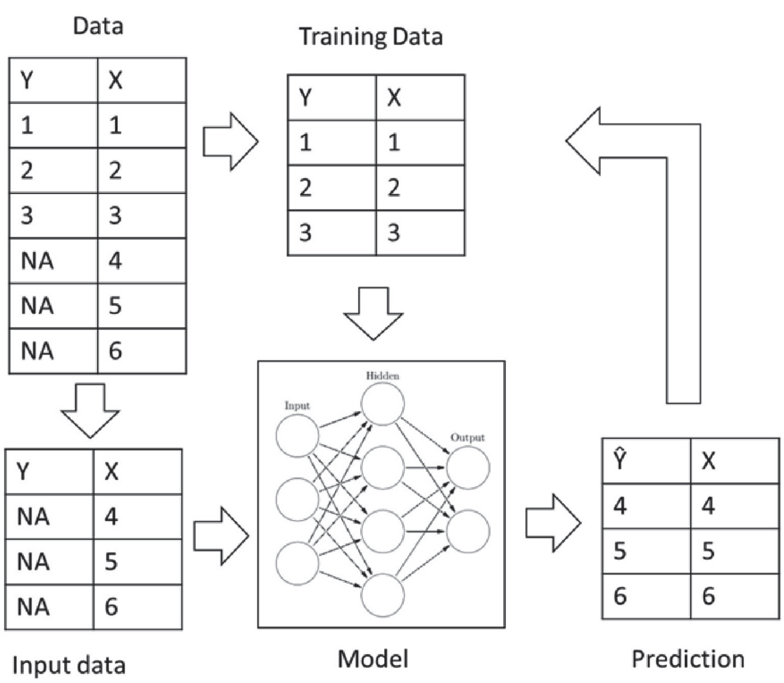

Figure 2. Schematic representation of Machine learning and Artificial Neural Network algorithm for prediction
Note: NA is data points for which data is not available. The values of Y and X are symbolic and to be interpreted for visualization only.

Buuren 2012). In this study, we have used violin plot which is a combination of both box and density plot to compare across models. Another class of methods comprises of the statistical tests, for example Kolmogorov–Smirnov test, that compare the empirical distributions of the observed and predicted data (Abayomi et al. 2008). But, the results are difficult to interpret, as the magnitude of the p-values depends on the proportion of missing values and sample size (Nguyen et al. 2013). It is also to be noted that discrepancies between observed and imputed data should not be taken that there is a problem in the imputed values, as such discrepancies are ought to occur.

Apart from these, evaluation metrics such as Root Mean Square Error (RMSE) is also useful to compare the prediction performance of the model (Schmitt et al. 2015). RMSE measures the difference between imputed and true values. It is the sample standard deviation of that difference. The formula for calculating RMSE is:

$$RSME = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}}$$

We use both RMSE and R-square as measure to compare the prediction performance of satellite night light data and per capita GDP data (Mullainathan and Spiess 2017).

5 Results and discussion

The relationship between night-light and poverty is complex (see table 1). Studies have plotted a linear relationship between night-light and poverty. We plot night-light and poverty as linear and non-linear relationship, and find that the relationship is more or less non-linear and downward slopping or negatively sloped (figure 3). Similarly, a non-linear relationship is also seen between per capita GSDP and poverty (figure 4). In both cases the non-linearity is a result of few outlier observations. We used machine learning techniques to predict the missing data.

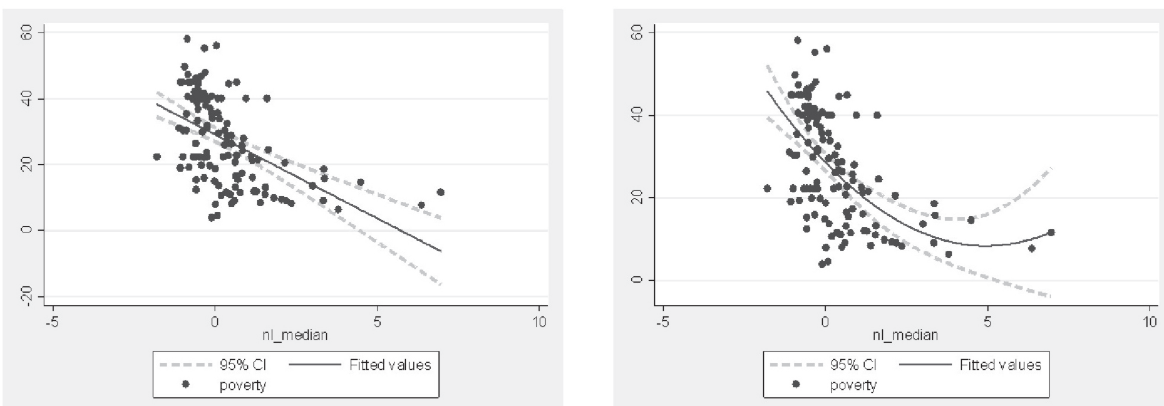

Figure 3. Relationship between night light and poverty

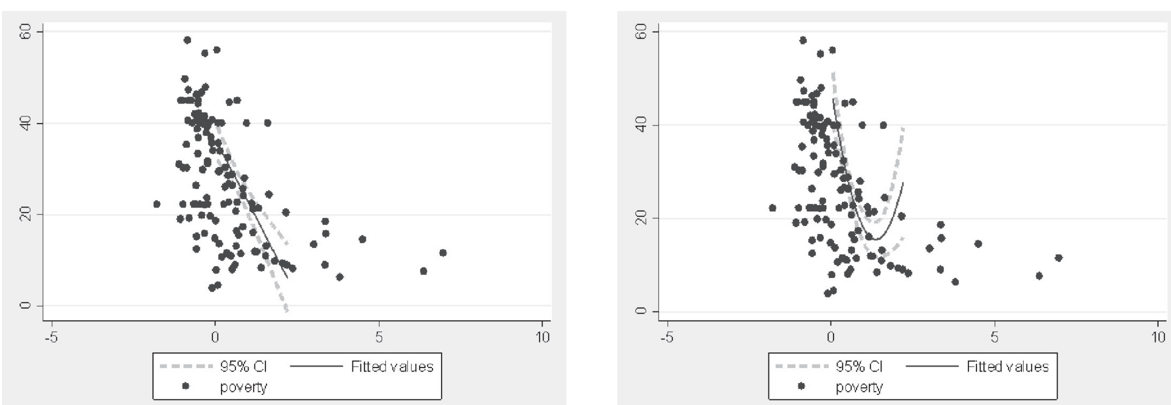

Figure 4. Relationship between per capita GDP and poverty

Table 2. Comparing predicted values of rural poverty with observed values

S.No.	Models	Mean	Std. Dev.	Min	Max
1	Measured poverty values	27.59	14.35	1.05	63.91
2	Machine learning (ANN)- night light	26.28	11.20	2.95	65.92
3	Machine learning (ANN)- per capita GDP	28.53	10.59	7.71	56.82

Note: Mizoram state is excluded from the analysis because of missing data on per capita GDP.

5.1 Comparing night light and per capita GDP data with observed values

The descriptive statistics of the predicted poverty value based on machine learning technique (ANN) and poverty values of the years and states for which data is available (training data) are provided in table 2. As discussed, the predicted values are compared with measured values for internal check and judged for its reasonableness as external check. Statistical measures such as mean, standard deviation, range (minimum and maximum) values of the observed and imputed values are compared. The predicted values from night light data are closer to that of the actual measured data.

These values can also be visualized through, Kdenisty plot, box plot and violin plot (figures 5 to 7). The plot/ line named as poverty is the measured value and other are the predicted values using night light and per capita GDP. From these figures we see that the night light data follow the closest pattern to the measured poverty compared to per capita GDP data. We also find existence of outliers among the predicted values which have resulted in peaks of the predicted values.

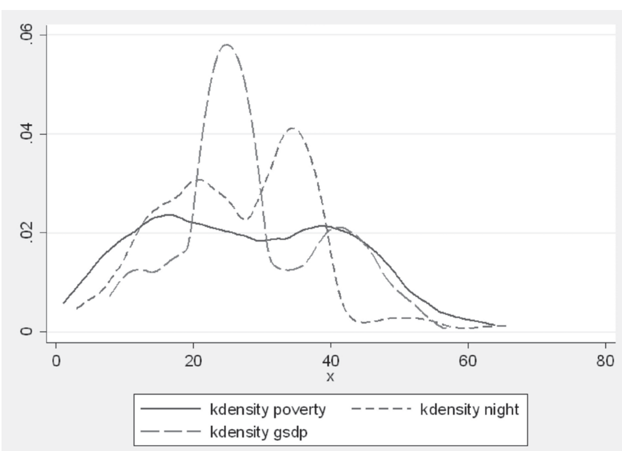

Figure 5. Kdenisty plot comparing predicted values of rural poverty with observed values
Note: Predicted values of night light (night) and per capita GDP

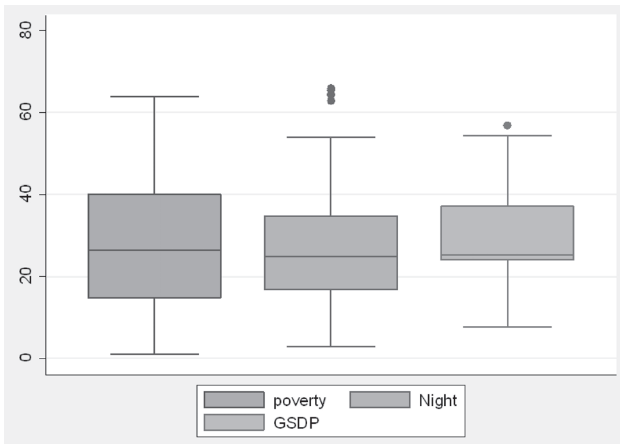

Figure 6. Box plot to comparing predicted values of rural poverty with observed values
Note: Predicted values of night light (Night) and per capita GDP

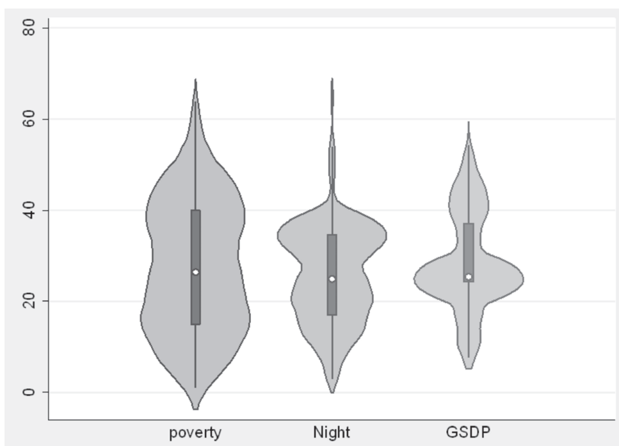

Figure 7. Violin plot comparing predicted values of rural poverty with observed values
Note: Predicted values of night light (Night) and per capita GDP

5.2 Comparing night light and per capita GDP models

We pool the predicted and measured poverty rates and plot these using Kdenisty plot, box plot and violin plot

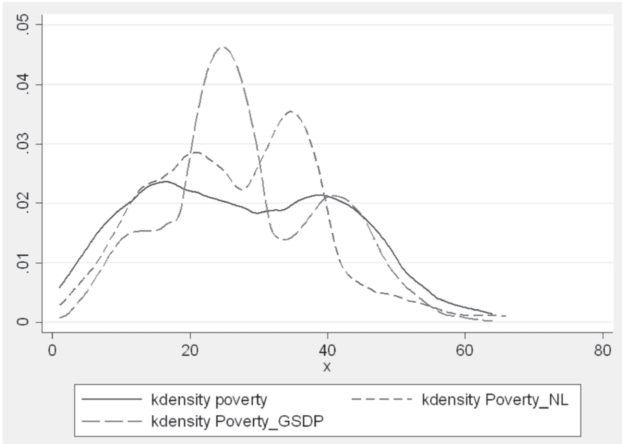

Figure 8. Kdenisty graphs to comparing different models
Note: Pooled poverty data; with night light (Poverty_NL) and per capita GDP (Poverty_GSDP)

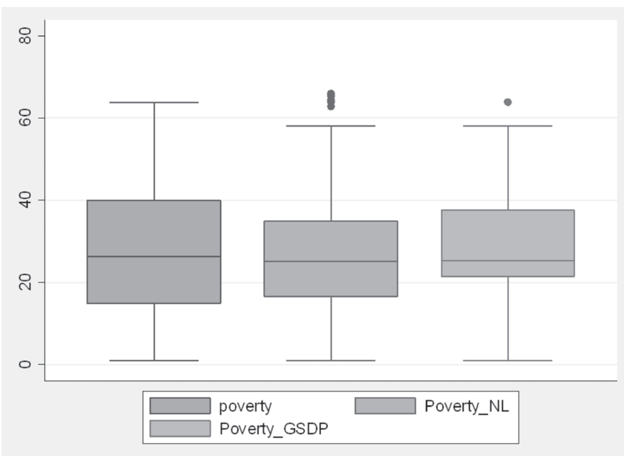

Figure 9. Box plot to comparing different models
Note: Pooled poverty data; with night light (Poverty_NL) and per capita GSDP (Poverty_GSDP)

(figure 8 to 10). The observed patterns are similar to those in figure 5 to 7, but the peaks are smoothened. These graphs show that there is a skewedness in the predicted values obtained through the machine learning approach, which can be attributed to the outliers.

To assess the prediction performance of the models we compute RMSE from the two models. The RMSE value is the lowest for machine learning models using nightlight compared to that for per capita GDP (table 3). Similar inference is drawn from comparing the R^2 values. We have used aggregated state-level satellite night light data. The night light data can be used at disaggregated levels, i.e. districts, taluks and villages.

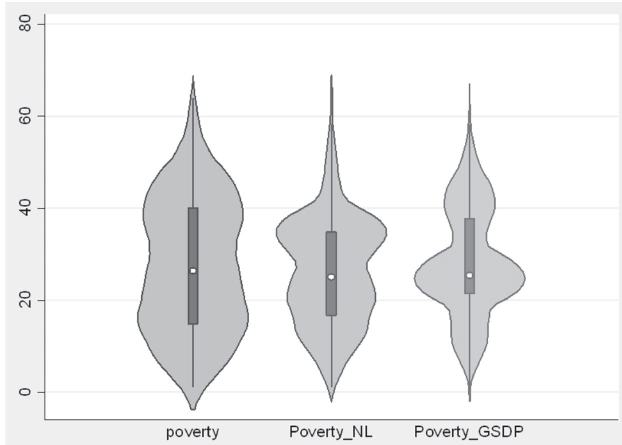

Figure 10. Violin plot comparing different models
Note: Pooled poverty data; with night light (Poverty_NL) and per capita GDP (Poverty_GSDP)

However, at this level of disaggregation data on GDP are not available. So, the potential of satellite night light data at disaggregated level needs to be explored. The R^2 values are closer to those obtained by others, for example Jean et al. (2016) and Newhouse (2018). The challenge to get estimates of R^2 of more than 0.5 to predicting poverty at large scale. Some studies had shown that adding of land use indicators from satellite data could increase the R^2 to 0.64 (Newhouse 2018).

Table 3. Estimates of RMSE and R^2 values from different models

Model	RSME	R^2
Machine learning (ANN)-night light	7.38	0.59
Machine learning (ANN)- per capita GDP	8.29	0.56

Newhouse (2018) raises some other issues on applicability of such approaches in predicting poverty. Prediction of poverty using such approaches are attenuated; it overestimates welfare for the poorest clusters and underestimates for the richest. The Convolutional Neural Network approach could distinguish between dark, medium and bright clusters of night light, but does not distinguish well within each category. There are several other alternatives such as using Lasso regression to prevent overfitting, using flexible prediction methods such as random forest for extrapolation outside the survey sample, and using day light satellite imagery (google static maps); high resolution satellite imagery. These techniques have their own strengths and weaknesses. Nevertheless, the satellite imagery data can still contribute information when survey based estimates of poverty are not available. The bottom line is that these approaches can be used to fill the data gap at a moderate additional cost.

6 Conclusions

The rapid progress in application of satellite imagery in predicting economic indicators are a result of increased availability of satellite imagery and advancement in algorithms to process such data. There is a strong demand by the policymakers for more frequent and local estimates of economic indicators. To address the issue of non-availability of a time series data for poverty, we have used machine learning techniques and satellite night light data to predict poverty. The predicted poverty data can be used to model various economic relationships to answer many relevant policy questions, which otherwise are limited by the unavailability of continuous series on poverty rates. However, these predictions are better suited for modelling purpose with suitable adjustments rather than as a proxy for actual data. Further studies could be undertaken using data at a more disaggregated levels, say districts, at which official data are not available on a time-scale.

Acknowledgement

The paper is part of research project titled Farm-Nonfarm Linkages and Rural Poverty in India conducted at ICAR-National Institute of Agricultural Economics and Policy Research (NIAP), New Delhi. A previous version of the draft was presented at 26th Annual Conference of Agricultural Economics Research Association (India) held at ICAR-National Dairy Research Institute, Karnal, India. We also thank the AERR anonymous reviewers and the editor, Dr. P.S. Birthal for their constructive comments which helped us in refining the manuscript.

References

Abayomi, K., Gelman, A., & Levy, M. (2008). Diagnostics for multivariate imputations. Journal of the Royal Statistical Society. Series C, Applied statistics, 57, 273–91.

Akiyama, Y. (2012). Analysis of light intensity data by the DMSP / OLS satellite image using existing spatial data for monitoring human activity in Japan. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, I-2, XXII ISPRS Congress, Melbourne, Australia, 25 August – 01 September.

Alkire, S., Jindra, Christoph, Aguilar, R., Gisela, & Vaz, A. (2017). Multidimensional poverty reduction among countries in Sub-Saharan Africa. OPHI Working paper 112. Oxford Poverty & Human Development Initiative (OPHI), Oxford University. Oxford, UK.

Chen, Xi., & Nordhaus, W.D. (2010). The value of luminosity data as a proxy for economic statistics. Proceedings of National Academy of Sciences, 108(21), 8589–8594.

Cruz, J.A. & Wishart, D.S. (2006). Applications of machine learning in cancer prediction and prognosis. Cancer Informatics, 2, 59–77.

Deaton, A. (2005). Measuring poverty in growing world (or measuring growth in a poor world). The Review of Economics and Statistics, 87(1), 1-18.

Deaton, A & Kozel, V. (2005). Data and dogma: The great Indian poverty debate. World Bank Research Observer, 20(2), 177-200.

Devarajan, S (2013). Africa's statistical tragedy. The Review of Income and Wealth, 59(1), S9-S15

Duan, T., Chartock, E., Ishfaq, H., Novosad, P., Asher, S., Burke, M., Lobell, D., & Ermon, S. (2017). Predicting poverty with satellite imagery in Bangladesh and India. https://web.stanford.edu/~hmishfaq/cs325b.pdf.

Dugoua, E., Kennedy, R. & Urpelainen, J. (2018). Satellite data for the social sciences: measuring rural electrification with night-time lights. International Journal of Remote Sensing, 39(9), 2690–2701.

Elvidge, C. D., Baugh, K. E., Anderson, S. J., Sutton, P. C., & Ghosh, T. (2012). The night light development index (NLDI): A spatially explicit measure of human development from satellite data. Social Geography, 7(1): 23–35.

Ghosh, T., Anderson, S.J., Elvidge, C.D. & Sutton, P.C. (2013). Using nighttime satellite imagery as a proxy measure of human well-being. Sustainability (Switzerland), 5(12), 4988–5019.

GoI. (2014). Report of the expert group to review the methodology for measurement of poverty. Planning Commission, Government of India.

Henderson, V., Storeygard, A. & Weil, D. (2009). Measuring economic growth from outer space. American Economic Review, 102(2), 994-1028.

Himanshu (2010). Towards new poverty lines for India. Economic and Political Weekly, 45(1), 38-48.

Jean, N., Burke, M., Xie, Davis, M.W., Lobell, D.B. & Ermon, S. (2016). Machine learning to predict poverty. Science, 353(6301), 790–794.

Kleinberg, B.J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. American Economic Review, 105(5), 491–95.

Li, X. & Zhou, Y. (2017). A stepwise calibration of global DMSP/OLS stable night time light data (1992–2013). Remote Sensing, 9(6), 637.

Mair, C., Kadoda, G., Le, M., Phalp, K., Scho, C., Shepperd, M. & Webster, S. (2000). An investigation of machine learning based prediction systems. Journal of Systems and Software, 53 (1), 23–29.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: concerns and ways forward. PLoS ONE, 13 (3), 1–26.

McCulloch, W.S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5, 115-133.

Mellander, C., Lobo, J., Stolarick, K. & Matheson, Z. (2015). Night-time light data: a good proxy measure for economic activity? PLOS ONE, 10(10), 1–18.

Michalopoulos, S & Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary African development. Econometrica, 81(1), 113-152.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. Journal of Economic Perspectives. 31(2), 87-106.

Newhouse, D. (2018). Using Satellite Imagery to predict local poverty and population density. World Bank.

Nguyen, C.D., Carlin, J.B., & Lee, K.J. (2013). Diagnosing problems with imputation models using the Kolmogorov–Smirnov test: a simulation study. BMC Medical Research Methodology, 13(144), 1-9.

Nguyen, C.D., Carlin, J.B. & Lee, K.J. (2017). Model checking in multiple imputation: an overview and case study. Emerging Themes in Epidemiology, 14(8), 1–12. doi:10.1186/s12982-017-0062-6.

Nischal, K. N., Radhakrishnan, R., Mehta, S. & Chandani, S. (2015). Correlating night-time satellite images with poverty and other census data of India and estimating future trends. In: Proceedings of the Second ACM IKDD Conference on Data Sciences – CoDS, 15 (March 2015), 75-79.

Noor, A.M., Alegana, V.A., Gething, P.W., Tatem, A.J. & Snow, R.W. (2008). Using remotely sensed night-time light as a proxy for poverty in Africa. Population Health Metrics, 6(5), 1-13.

Panagariya, A & Mukim, M. (2014). A comprehensive analysis of poverty in India. Asian Development Review, 31(1), 1-52.

Pinkovskiy, M & Sala-i-Martin, X. (2016). Lights, camera … income! illuminating the national accounts-household surveys debate. The Quarterly Journal of Economics, 131(2), 579-631.

Ravallion, M. (1994). Poverty comparison. Harwood Academic Publishers, Switzerland.

Ravillion, M. (2001). Comments on "counting the world's poor", by Angus Deaton. The World Bank Research Observer, 16(2), 149-156.

Sen, A. (1981). Poverty and Famines: An essay on entitlement and deprivation. Oxford University Press. Oxford. UK.

Stuart, E.A., Azur, M., Frangakis, C., & Leaf, P. (2009). Multiple imputation with large data sets: a case study of the children's mental health initiative. American Journal of Epidemiology, 169(9), 1133-1139.

Suraj, P., Guptay, K.A., Sharmay, M., Paulz, S.B., & Banerjee, S. (2017). On monitoring development using high resolution satellite images. http://arxiv.org/abs/1712.02282.

van Buuren, S. (2012). Flexible imputation of missing data. Chapman & Hall/CRC, Boca Raton.

von Hippel, P.T. (2013). Should a normal imputation model be modified to impute skewed variables? Sociological Methods and Research. 42(1),105–138.

White, I.R., Royston, P., & Wood, A.M. (2011). Multiple imputation using chained equations: issues and guidance for practice. Statistics in Medicine, 30(4), 377-99.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting, 14, 35-62.

Received: 30 November 2018; Accepted: 24 December 2018