Nighttime Lights as a Proxy for Economic Development on a Local Level – The Case of the US

Extended Abstract

Timely and spatially detailed information on regional economic activity is essential for policymakers, investors, and researchers, yet official statistics at subnational levels are typically published with substantial delays. In the United States, state- and county-level GDP and personal income are often released many months after the end of the reference period, constraining the ability of public authorities to respond to shocks and limiting private-sector capacity to adjust investment and employment decisions in real time. Satellite-derived nighttime lights imagery (NTLI) offers a near-real-time, globally consistent proxy for human activity and has become an important tool for measuring and forecasting economic development, particularly in data-scarce environments. However, most existing research focuses on developing economies, uses relatively simple light-based indicators such as the sum of lights, and relies on linear models applied to short panels or cross-sectional snapshots.

This paper revisits the NTLI-economy relationship in the context of a highly developed country, the United States, over a long time span and at multiple spatial scales. We ask whether economic well-being – proxied by real GDP and personal income – can be accurately and consistently predicted for both states and counties using only NTLI-based features. We further examine whether richer distributional and spatial metrics derived from nighttime lights provide additional explanatory and predictive power beyond total brightness, and whether modern non-linear machine-learning methods offer performance gains relative to traditional linear regression.

Research questions and contribution

Building on a vast literature documenting strong correlations between nighttime lights and economic activity at national and regional levels, we identify five main limitations in existing work, particularly for advanced economies. First, many studies focus on less developed countries where statistical systems are weak; their findings may not generalize to contexts where official data are of high quality and economic structures differ. Second, most analyses are cross-sectional or cover short time windows, limiting insight into the temporal stability of the lights–economy relationship. Third, the sum of lights is frequently used as the sole NTLI-based predictor, neglecting distributional and spatial information embedded in the imagery. Fourth, empirical strategies often emphasize explanatory correlations rather than out-of-sample predictive performance, and they rely heavily on linear regression despite potential non-linearities. Finally, separate use of DMSP-OLS or VIIRS datasets restricts the temporal coverage and complicates comparisons across periods.

We address these gaps by focusing on a highly developed country (the US), using a harmonized NTLI series that combines DMSP and VIIRS over nearly three decades, and adopting a multi-model, multi-scale framework with explicit evaluation of both explanatory and predictive performance. Specifically, we formulate four hypotheses: (1) the NTLI–economic well-being relationship is stronger at the regional (state) level than at the local (county) level; (2) the spatial distribution of lights within a territory has additional explanatory power beyond the sum of lights; (3) NTLI can be used to accurately predict next-year GDP and income at both scales; and (4) non-linear machine-learning algorithms outperform traditional OLS in terms of explanatory and predictive power.

Data

Our empirical analysis combines satellite-derived nighttime lights with official economic statistics for US states and counties. For NTLI, we use the Harmonized Global Nighttime Lights series – consistent annual panel from the early 1990s onward. This harmonization preserves the DMSP-OLS spatial resolution while leveraging the improved sensitivity and dynamic range of VIIRS, thereby enabling long-run analysis of light–economy linkages.

From the harmonized imagery, we compute a wide range of indicators for each state and county. In addition to conventional measures such as the sum of lights, we construct features that describe intensity distributions (e.g. higher-order moments, quantiles), concentration and inequality of light (e.g. Gini-type measures, shares of light in the brightest pixels), and spatial patterns (e.g. clustering metrics, spatial dispersion indices). These features are designed to capture both the magnitude and the spatial organization of human activity as reflected in nighttime illumination, without relying on external covariates such as land cover or population grids.

Economic well-being is proxied by (i) real GDP and (ii) personal income, obtained from official US statistical sources at annual frequency for all states and counties. To reduce skewness and approximate linearity in baseline models, we apply logarithmic transformations to GDP, income, and selected NTLI measures, and we align the temporal coverage with the availability of both lights and economic data.

Methods

We implement a modelling framework that separates explanatory analysis from predictive analysis and applies the same procedures at two spatial levels: states and counties. For each year in the study period, we link NTLI-derived features to contemporaneous economic outcomes and estimate a set of models to assess in-sample fit (explanatory power) and one-year-ahead forecasting accuracy (predictive power).

The baseline econometric specification is a log–log Ordinary Least Squares (OLS) regression using the sum of lights as the sole predictor of GDP or income. This mirrors much of the earlier literature and serves as a benchmark for evaluating the added value of more complex NTLI features and machine-learning methods. We then extend the OLS model by including the full set of NTLI-derived indicators, allowing us to test whether spatial and distributional metrics provide incremental explanatory power beyond total brightness.

To capture potential non-linearities and interactions, we estimate several machine-learning models: LASSO (penalized linear regression), Support Vector Regression (SVR), k-Nearest Neighbours (KNN), Random Forest (RF), and XGBoost. These algorithms represent diverse modelling philosophies, from sparse linear models (LASSO) to kernel-based methods (SVR) and ensemble tree techniques (RF, XGBoost) that can flexibly approximate complex functional relationships. All models are trained using consistent feature sets and tuning procedures, with hyperparameters selected via cross-validation.

For the explanatory analysis, models are trained and evaluated on the same year's data, with performance summarized by the coefficient of determination R^2, measuring the proportion of variance in GDP or income explained by the predictors. For the predictive analysis, we adopt a rolling one-year-ahead design: models are trained on year *t* and used to forecast GDP or income in year *t+1*. Forecast accuracy is again assessed using R^2, now computed on the held-out next-year observations. This design mimics a realistic "nowcasting" scenario where NTLI data are available earlier than official economic statistics.

To understand which NTLI-derived features drive model performance, we employ permutation-based feature importance, a model-agnostic explainable AI (XAI) technique. For each trained model, we randomly permute the values of a given predictor across observations and measure the resulting drop in performance; predictors whose shuffling leads to large declines in R^2 are deemed more important. This provides an intuitive ranking of the relative contribution of the sum of lights versus more nuanced distributional and spatial indicators, without relying on model-specific coefficients.

Results

Preliminary descriptive analysis confirms a positive association between log sum of lights and log real GDP and income at both state and county levels, but with noticeably tighter relationships for states.

Hypothesis 1 concerns the relative strength of the NTLI–economy relationship at regional versus local scales. Throughout the full period we study, nighttime lights prove to be much more tightly linked to economic outcomes at the state level than at the county level. This pattern is robust across both GDP and personal income, and it appears whether we use a simple measure such as the sum of lights or a richer set of indicators derived from the imagery. Taken together, the results suggest that aggregation to larger regions helps to smooth out local idiosyncrasies and measurement noise, yielding a stronger and more stable link between lights and economic activity at the regional scale.

Hypothesis 2 assumes that the way lights are distributed in space carries additional information beyond their total intensity. When we move from models based solely on the sum of lights to specifications that also include distributional and spatial metrics, the ability of nighttime lights to account for differences in economic well-being improves markedly, especially for states. For these larger units, the extended feature set raises the explained share of variation in both GDP and income well above the single-indicator benchmark, while for counties the improvements, though more modest, are still visible. Analyses of feature importance show that indicators capturing how concentrated or unequal the light is, and how it is spatially dispersed, often emerge as key predictors alongside total brightness. This implies that whether light is clustered in a few bright urban cores or spread more evenly across a territory conveys economically meaningful information.

Hypothesis 3 addresses predictive performance: can NTLI-based models say something useful about next year′s GDP and income? When we use the models to forecast one year ahead, several specifications attain high accuracy for states, even though they rely only on nighttime lights. For counties, forecasts are less precise, reflecting the noisier and weaker underlying relationship, yet some models still manage to capture a non-negligible share of the variation in future outcomes. These findings indicate that harmonized nighttime lights can function as a practical early indicator of subnational economic conditions, even in a context such as the United States where official statistics are relatively rich but slow to arrive.

Hypothesis 4 compares non-linear machine-learning algorithms with conventional linear regression. Across both the contemporaneous fits and the one-year-ahead forecasts, tree-based methods – most notably Random Forest and XGBoost – consistently outperform OLS. Their advantage is greatest when they can draw on the full NTLI feature set, which suggests that they are able to exploit non-linearities and interactions among the various distributional and spatial indicators of light. LASSO, while occasionally improving on simple OLS through variable selection, generally remains less powerful than these more flexible approaches, and SVR and KNN occupy an intermediate position that depends on tuning choices. Overall, the results point to non-linear machine-learning models as better suited than traditional econometric specifications for extracting explanatory and predictive signals from nighttime lights in this setting.

Conclusions and limitations

Our findings confirm that satellite-derived nighttime lights are a powerful proxy for regional economic well-being in a highly developed country, particularly at aggregated spatial scales. At the state level, NTLI-based models – using only remotely sensed data – can explain and predict a large share of the variation in GDP and personal income, making them useful for nowcasting applications where official statistics are delayed. At the county level, the relationship is weaker but still meaningful, suggesting that lights can complement, rather than replace, conventional data sources in local economic analysis.

The results also highlight the value of moving beyond simple sums of lights. Distributional and spatial metrics derived from harmonized imagery contribute significantly to model performance, especially when combined with non-linear machine-learning methods. This underscores the importance of taking full advantage of the rich information embedded in high-resolution NTLI, rather than treating it as a single scalar proxy. Moreover, the use of permutation-based feature importance demonstrates that explainable AI tools can provide transparent insight into how different aspects of nighttime illumination relate to economic outcomes.

From a methodological perspective, the paper demonstrates that harmonized NTLI series can be effectively used for long-run analysis in advanced economies, provided that appropriate modelling choices are made. The multi-model, multi-scale framework, with explicit separation of explanatory and predictive objectives, offers a template for future work in other countries and regions. It also suggests that policy institutions interested in near-real-time monitoring of regional economies may benefit from integrating NTLI-based indicators into their analytical toolkits.

At the same time, limitations remain. The lights–economy relationship is weaker in rural areas and in sectors with low light emissions, such as agriculture and forestry, which means that NTLI-based estimates can understate the economic activity of sparsely lit regions. In addition, even in advanced economies, changes in lighting technology and energy efficiency may gradually decouple nighttime brightness from economic output. Future research could therefore focus on integrating NTLI with complementary remote-sensing products – such as land cover, vegetation indices, or daytime imagery – as well as with high-frequency administrative and mobility data, to further improve coverage of non-illuminated economic activity.