REGION
The Journal of ERSA
Powered by WU

Volume 1, Number 1, 2020, 999–999
DOI: 10.18335/region.v??i??.???
for review only

journal homepage: region.ersa.org
ISSN: 2409-5370
journal homepage: region.ersa.org

Regional growth, convergence, and spatial spillovers in India:
A reproducible view from outer space

Author(s): SUPPRESSED

Received: February 5, 2026/Accepted:

Abstract. Using satellite nighttime light data as a proxy for economic activity, Chanda and Kabiraj (2020, World Development) studied regional growth and convergence across 520 districts in India. Based on a reproducible open-science approach, this article builds on their work by extending their main findings on three fronts. First, we illustrate regional convergence patterns using an interactive tool for satellite imagery visualization. Second, we assess the degree of spatial dependence in their main econometric specification. Third, we employ a spatial Durbin model to measure the role of spatial spillovers in the convergence process. Our results indicate that spatial spillovers increase the estimated speed of regional convergence. Overall, the results highlight the role of spatial dependence in regional convergence analyses through the lens of satellite imagery, interactive visualizations, and spillover modeling.

1 Introduction

Regional economic growth and convergence are key concerns in developing countries, particularly in large federal states like India where spatial inequalities can threaten social cohesion and political stability. However, studying regional convergence in developing countries has been historically challenging due to limited availability of consistent economic data at subnational administrative levels. In response to this challenge, the emergence of satellite nighttime light data as a proxy for economic activity has enabled a growing literature about regional growth dynamics at granular geographic scales.

Chanda, Kabiraj (2020) leveraged nighttime light data to document regional convergence across 520 districts in India between 1996 and 2010. Their analysis showed that poorer districts grew faster than richer ones during this period, suggesting a reduction in spatial inequalities. However, their econometric approach did not account for potential spatial spillover effects in the convergence process. Specifically, a district's growth trajectory might be influenced not only by its own initial conditions but also by those of its neighbors.

This article extends the study of Chanda, Kabiraj (2020) in three key methodological directions. First, we develop an interactive visualization tool that allows researchers to explore spatial and temporal patterns of regional convergence using satellite nighttime light data. This tool helps identify converging regions and growth hotspots that may be difficult to detect in static visualizations. Second, we formally test for spatial dependence in both the dependent and independent variables of the convergence equations. Our tests highlight that spatial autocorrelation is a relevant feature of satellite data and the regional convergence process. Third, we employ a spatial Durbin model that explicitly

accounts for spatial spillovers and quantify how neighbors can influence the speed of regional convergence.

Our results contribute to the understanding of regional convergence in India on three fronts. First, interactive visualization tools reveal clear spatial patterns in both the initial distribution and subsequent growth of nighttime lights. Second, formal tests of spatial dependence indicate that district-level economic trajectories are not independent of their neighbors. Third, accounting for spatial spillovers through a spatial Durbin model shows that the total convergence effect is considerably larger than previous non-spatial estimates would suggest. Specifically, spatial spillovers appear to accelerate the convergence process by creating additional channels through which lagging regions can catch up.

These findings also provide implications for methodology and policy. Methodologically, they suggest that conventional non-spatial approaches may underestimate the speed of regional convergence by failing to account for inter-district spillovers. From a policy perspective, they suggest that the benefits of place-based policies may extend beyond target districts through spatial multiplier effects, potentially increasing their cost-effectiveness.

In addition to these methodological contributions, this article adopts a reproducible open-science approach by leveraging Jupyter notebooks and the Quarto publishing system. Jupyter notebooks integrate executable code, narrative text, and computational outputs within a single document, supporting multiple programming languages such as Python, R, and Stata. Quarto is an open-source publishing system that generates multiple output formats—including HTML, PDF, and Word—from a single source file, ensuring consistency across all versions of a document. Together, these tools foster scientific reproducibility and open science by making every analytical step transparent, verifiable, and accessible to other researchers.

The rest of this article is organized as follows. Section 2 provides an overview of the data and methods, describing the use of nighttime light data as a proxy for economic activity. It also introduces the methodological extensions related to reproducible open science, interactive visualizations, spatial dependence testing, and spillover modeling. Section 3 presents our empirical results, beginning with an interactive exploration of regional convergence patterns, followed by formal tests of spatial dependence. The section concludes with estimates of direct and indirect convergence effects from the spatial Durbin model. Finally, Section 4 offers some concluding remarks.

2 Data and methods

2.1 Data: Nighttime lights as a proxy for economic activity

One of the key challenges in studying economic growth in developing countries like India is the limited availability and reliability of data on aggregate economic activity below the state level. To address this issue, a growing body of literature, pioneered by Henderson et al. (2012), has utilized satellite nighttime light data as a proxy for economic activity at subnational levels.

Nighttime light (hereafter NTL) data have been widely used to investigate economic growth and convergence across national and subnational regions in various countries. For instance, Adhikari, Dhital (2020) examine the impact of decentralization on regional convergence using NTL data. Pinkovskiy, Sala-i Martin (2016) use NTL data to adjudicate between national accounts and household surveys. They find that national accounts better capture aggregate economic growth. Similarly, Lessmann, Seidel (2017) leverage NTL data to estimate GDP per capita and spatial inequality globally.

NTL data have been applied to a range of research questions in India. For example, Cook, Shah (2022) use NTL data to analyze the impact of public welfare programs. Jha, Talathi (2021) examine the effects of colonial institutions. Chanda, Cook (2022) investigate the impact of demonetization. Beyer et al. (2021) employ NTL data to study the effects of COVID-19. In the context of convergence studies, Chakravarty, Dehejia (2019) document significant regional disparities in India using NTL data and caution that the goods and services tax may further exacerbate them.

Our study builds on Chanda, Kabiraj (2020). Following their approach, we use the per capita growth in nighttime lights as the dependent variable and the initial nighttime lights

per capita as the primary explanatory variable. For 520 districts in India, these variables are derived from NTL data released by the National Geophysical Data Center (NGDC). The data are based on observations from the DMSP/OLS satellites spanning the period from 1996 to 2010. To mitigate the issue of top-coding in NTL data, the NGDC released "radiance-calibrated" nighttime lights for eight specific years within this period. This dataset employs high magnification settings for low-light regions and low magnification settings for brightly lit areas. For this study, we utilize the "radiance-calibrated" nighttime lights data.[1]

2.2 Jupyter and Quarto for reproducible open science

Jupyter notebooks have become an essential tool for reproducible open science. They allow researchers to integrate executable code, explanatory narrative, and computational outputs within a single self-contained document (Kluyver et al. 2016). This integration ensures that every step of the analytical workflow—from data ingestion and processing to statistical modeling and visualization—is transparently documented and can be independently verified by other researchers. Unlike traditional workflows where code, results, and interpretation are separated across different files and software environments, Jupyter notebooks preserve the complete chain of reasoning. This unified format can be shared, inspected, and re-executed. In this article, we employ Jupyter notebooks with three distinct computational kernels—Python, R, and Stata—to document our data processing, analysis, and visualization steps. This approach enables readers to trace each result back to the code that produced it. Moreover, Python and R notebooks can be executed in the cloud using Google Colaboratory. While Stata is not open-source software, its integration with the Jupyter environment provides a rich interactive interface. This integration brings the same benefits of literate programming and transparent documentation to proprietary statistical software (Knuth 1984).

Quarto is an open-source scientific and technical publishing system that extends the capabilities of computational notebooks into a comprehensive manuscript preparation framework (Allaire et al. 2024). Its single-source publishing approach enables researchers to generate multiple output formats from a single source file. These formats include HTML for interactive web-based dissemination, PDF for formal journal submission, Word documents for collaborative editing, and JATS XML for archival and indexing purposes. This approach ensures consistency across all versions of a document, eliminating discrepancies that can arise when maintaining separate files for different output formats. By providing a unified authoring environment that natively supports cross-references, citations, mathematical notation, and embedded computational results, Quarto lowers the technical barriers to producing publication-quality research.

The combination of Jupyter notebooks and Quarto's manuscript framework creates a rich infrastructure for reproducible research. In this article, the figures and tables that appear in the manuscript are programmatically embedded from specific Jupyter notebook cells. This ensures that every empirical finding presented in the text is directly traceable to its underlying computation. All data processing and analysis steps are documented in the notebooks, and we provide access to the raw data and code through a public GitHub repository. By adopting this integrated framework, we aim to promote transparency and enable other researchers to build upon our work (Peng 2011).

2.3 Google Earth Engine for interactive spatial visualizations

Interactive visualizations of nighttime lights imagery offer methodological advantages over static representations, particularly when analyzing temporal and spatial heterogeneity in economic activity (Donaldson, Storeygard 2016). The dynamic nature of these visualizations enables researchers to simultaneously examine multiple dimensions of the data.

[1] Following Chanda, Kabiraj (2020), our sample consists of 520 districts (out of a possible 593). There were 593 districts in 2001, which rose to 640 districts in the 2011 census. To match districts across the two census files, we merged newly split districts back with their parent districts in the 2011 census. Eight of the 47 new districts were created by splitting areas from multiple parent districts. Those new districts along with their multiple-origin districts were dropped from our sample. We dropped all the districts in the state of Assam where more than 50% of districts were created in that manner.

These dimensions include temporal variations in light intensity, the spatial distribution of economic activity, and the relationship between nighttime lights and other georeferenced variables. The ability to dynamically adjust visualization parameters allows for more nuanced exploration of economic patterns that might be obscured in static representations.

Google Earth Engine (GEE) lowers the computational and technical barriers to creating such interactive visualizations and provides an accessible platform for analyzing satellite data (Gorelick et al. 2017). The platform's browser-based integrated development environment supports the creation of interactive web applications without requiring extensive infrastructure or specialized software installation. This capability is especially valuable for reproducible research, as it enables the development of web applications that can be shared with other researchers. The platform's ability to handle large-scale geospatial computations significantly streamlines the workflow from raw data to interactive visualization. Its extensive catalog of pre-processed nighttime lights datasets, including the DMSP-OLS and VIIRS collections, further reduces the storage and processing burden (Tamiminia et al. 2020).

Interactive visualizations of nighttime lights data are especially useful for identifying and analyzing patterns of regional convergence in luminosity. Through dynamic visualization tools, researchers can track the evolution of light intensity across regions over time. This approach effectively identifies areas that exhibit catch-up growth patterns—where initially dim regions progressively converge toward the luminosity levels of their brighter counterparts.

2.4 Regional convergence modeling

In neoclassical growth models, the per capita growth rate is predicted to be negatively correlated with a region's initial endowment, primarily due to diminishing returns to capital accumulation (Solow 1956). Specifically, poorer regions, assuming similar technology and preferences, are expected to experience higher growth rates compared to their wealthier counterparts. Thus, over the long run, regions with similar characteristics should converge to a common steady state. We examine this convergence hypothesis employing a growth regressions framework in the tradition of Barro, Sala-i Martin (1992). Equation 1 presents this convergence process in its simplest (unconditional) form.

$$g_t = \beta_1 x_{t-1} + \varepsilon_t \tag{1}$$

where g_t represents an N-by-1 vector of observations on per-capita NTL growth for each of the N regions over the period t. The vector x_{t-1} represents an N-by-1 vector of observations on the initial (log) level of per-capita NTL. The parameter β_1 is a regression coefficient that indicates the direction and strength of regional convergence. A negative value of β_1 would suggest that regions with lower initial NTL levels grow faster, consistent with the convergence hypothesis. Finally, ε_t represents a vector of idiosyncratic error terms.

As there are no control variables, this simple convergence framework implies that districts converge to a common steady state. However, regions may differ in various aspects such as geography, socio-economic conditions, and policy implementation. To account for these differences, we include state fixed effects to control for state-specific institutions and policies that influence the rate of convergence. Additionally, we incorporate a range of geo-climatic controls alongside district-specific conditions related to demographics, human capital, and infrastructure. Equation 2 summarizes this conditional convergence framework.

$$g_t = \beta_1 x_{t-1} + X_t \alpha + \varepsilon_t \tag{2}$$

Here, the matrix X_t is an N-by-k collection of observations on control variables (including state fixed effects) for each district in our sample. The vector α, with dimensions $k \times 1$, captures the regression coefficients for these variables. The full list of control variables and their descriptions are available in Table A.1 of Chanda, Kabiraj (2020).

2.5 *Spatial dependence testing*

The analysis of regional convergence using nighttime light data requires explicit consideration of spatial dependencies across districts. Spatial dependence—the tendency for observations at nearby locations to be more similar than those at distant locations—can arise from spatial spillovers, shared geographic conditions, or inter-regional economic linkages (Anselin 1988). If present and unaccounted for, spatial dependence can lead to biased parameter estimates and invalid inference in standard regression models. To formally test for such spatial relationships, we employ the Global Moran's I statistic and Local Indicators of Spatial Association (LISA), applied to the main variables of Equation 2.

The Global Moran's I statistic, originally proposed by Moran (1950), is the most widely used measure of spatial autocorrelation. It quantifies the overall degree of spatial clustering among geographic units and can be expressed as:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij}\, z_i\, z_j}{\sum_i z_i^2} \tag{3}$$

where z_i represents the deviation of observation i from the mean, w_{ij} denotes the spatial connection (weight) between units i and j, and n is the total number of observations. The Moran's I statistic typically ranges from -1 to $+1$. Positive values indicate positive spatial autocorrelation, where similar values tend to be located near each other. Values near zero suggest spatial randomness, and negative values indicate spatial dispersion, where dissimilar values are neighbors. Statistical significance is assessed through a permutation-based inference approach that compares the observed statistic against a reference distribution generated by randomly reassigning values across locations.

While the Global Moran's I provides a single summary measure of overall spatial dependence, it does not reveal where significant clusters or outliers are located. To address this limitation, Anselin (1995) proposed Local Indicators of Spatial Association (LISA), which decompose the global statistic into contributions from each individual observation. The local Moran's I for observation i is defined as:

$$I_i = z_i \sum_j w_{ij}\, z_j \tag{4}$$

where the summation is over the neighbors of i as defined by the spatial weight matrix. This local statistic classifies each observation into one of four categories based on the relationship between a location's value and those of its neighbors. High-High (HH) indicates a high-value location surrounded by high-value neighbors, while Low-Low (LL) indicates a low-value location surrounded by low-value neighbors. High-Low (HL) is a spatial outlier where a high-value location is surrounded by low-value neighbors, and Low-High (LH) is the opposite spatial outlier. Taken together, the HH and LL categories identify spatial clusters, while the HL and LH categories identify spatial outliers. Statistical significance of each local statistic is assessed through conditional permutation tests, with only statistically significant locations (typically at $p < 0.05$) reported in the LISA cluster maps.

The specification of the spatial weights matrix $\mathbf{W}$ is crucial for capturing the underlying spatial structure. We employ a k-nearest neighbors weight matrix with $k = 6$, whereby for each district, the six geographically closest districts are identified as its neighbors:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & w_{ij} & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \cdots & w_{nn} \end{bmatrix}$$

where $w_{ij} = 1$ if district j is among the six nearest neighbors of district i, and $w_{ij} = 0$ otherwise. Following standard practice, we row-normalize the weights matrix so that each row sums to one. This ensures that the spatial lag of a variable, $\mathbf{Wx}$, represents the average value among a district's neighbors.

The detection of significant spatial dependence would justify the use of spatial econometric techniques. In particular, Ertur, Koch (2007) and Fischer (2011) argue that the spatial Durbin model can appropriately account for spatial dependence in the convergence process. This methodological choice allows us to distinguish between direct effects of district characteristics and indirect effects operating through spatial channels (LeSage, Pace 2009).

2.6 Spatial spillover modeling

Our spatial spillover modeling builds upon the spatial Solow growth model developed by Ertur, Koch (2007) and Fischer (2011). Their model extends the traditional Solow framework to account for technological interdependence across regions. The model considers an economy of N subnational regions, each characterized by a Cobb-Douglas production function with constant returns to scale:

$$Y_{it} = A_{it} K_{it}^{\alpha_K} H_{it}^{\alpha_H} L_{it}^{1-\alpha_K-\alpha_H} \tag{5}$$

where Y_{it} represents output, K_{it} physical capital, H_{it} human capital, L_{it} labor force, and A_{it} the level of technological knowledge for region i at time t. The parameters α_K and α_H denote the output elasticities with respect to physical and human capital, respectively. A key innovation of this framework is the modeling of technological knowledge, which incorporates both internal and external factors:

$$A_{it} = \Omega_t k_{it}^{\theta} h_{it}^{\phi} \prod_{j \neq i}^{N} A_{jt}^{\rho W_{ij}} \tag{6}$$

This specification captures three distinct components of technological progress:

- An exogenous component (Ω_t) representing the common stock of knowledge across regions
- An embodied component ($k_{it}^{\theta} h_{it}^{\phi}$) reflecting technology embedded in physical and human capital per worker
- A spatial component ($\prod_{j \neq i}^{N} A_{jt}^{\rho W_{ij}}$) capturing technological interdependence between regions

Based on this theoretical framework, Ertur, Koch (2007) derived a spatial Durbin model that accounts for regional spillovers in the convergence process. Their model can be compactly written in matrix notation as:

$$g_t = \beta_1 x_{t-1} + X_t \alpha + \beta_2 W x_{t-1} + W X_t \gamma + \lambda W g_t + \varepsilon_t \tag{7}$$

In this model, g_t represents an N-by-1 vector of observations on per-capita NTL growth for each of the N regions over the period t. The vector x_{t-1} represents an N-by-1 vector of observations on the initial (log) level of per-capita NTL. The parameter β_1 is a regression coefficient that indicates the direction and strength of regional convergence. The matrix X_t is an N-by-k collection of observations on control variables for each region. The vector α, with dimensions $k \times 1$, captures the regression coefficients for these variables. Additionally, WX_t denotes an $N \times k$ matrix of spatially lagged observations, composed of a linear combination of neighboring values for the variables of interest in each region. The vector γ represents the regression coefficients associated with these spatial lags. The terms Wx_{t-1} and Wg_t refer to N-by-1 vectors capturing the spatial lags of initial (log) per-capita NTL, and the per-capita NTL growth, respectively. Finally, ε_t represents a vector of idiosyncratic error terms.

3 Results

3.1 Regional convergence: An interactive exploration from outer space

Before presenting the regression results, we visually illustrate the concepts of growth and convergence in nighttime lights. First, we create an interactive map of India displaying

regional luminosity in 1996 and 2010.[2] Second, we examine absolute convergence by constructing a convergence scatterplot. This scatterplot depicts the relationship between per-capita growth rates in nighttime lights (1996–2010) and initial per-capita nighttime light levels (1996) for each district. Finally, we present some case studies to illustrate the nighttime light growth that occurred during our study period. Focusing on some of the poorest regions in the country, we show an increase in nighttime lights that visually aligns with the convergence hypothesis.

(a) Luminosity in 1996

(b) Luminosity in 2010

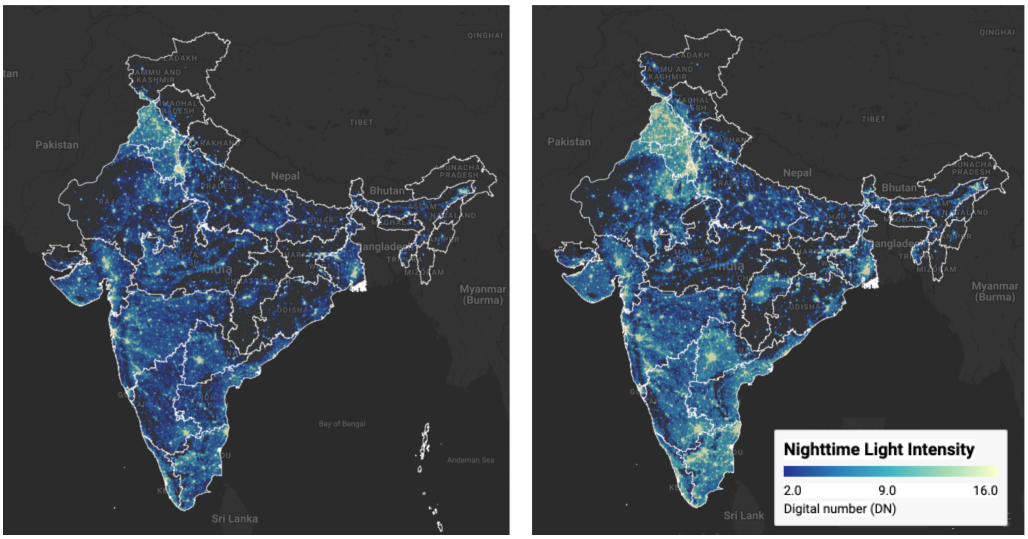

Figure 1: Regional luminosity in India: 1996 vs 2010 Notes: Luminosity is measured in radiance-calibrated digital number (DN) values from DMSP-OLS satellites. Interactive web application available at https://bit.ly/india-rc-ntl. Source: Authors' visualization using pre-processed luminosity images from the Earth Observation Group (NOAA/NCEI). See View from outer space notebook for source code.

Figure 1 presents static maps (captured from our interactive application) of luminosity for the initial and final years (1996 and 2010). The maps show a noticeable increase in brightness across most parts of India in 2010. Since nighttime lights serve as a proxy for economic activity, this increase in luminosity reflects economic growth over the period. Figure 2 illustrates the relationship between per-capita growth in nighttime lights and initial per-capita nighttime light. The scatterplot shows an inverse relationship, suggesting a β-convergence rate of approximately 2%, consistent with the seminal finding of Barro, Sala-i Martin (1992) about convergence.

[2]The interactive web application is available at https://bit.ly/india-rc-ntl. It was developed using Google Earth Engine and the source code is available in the View from outer space notebook.

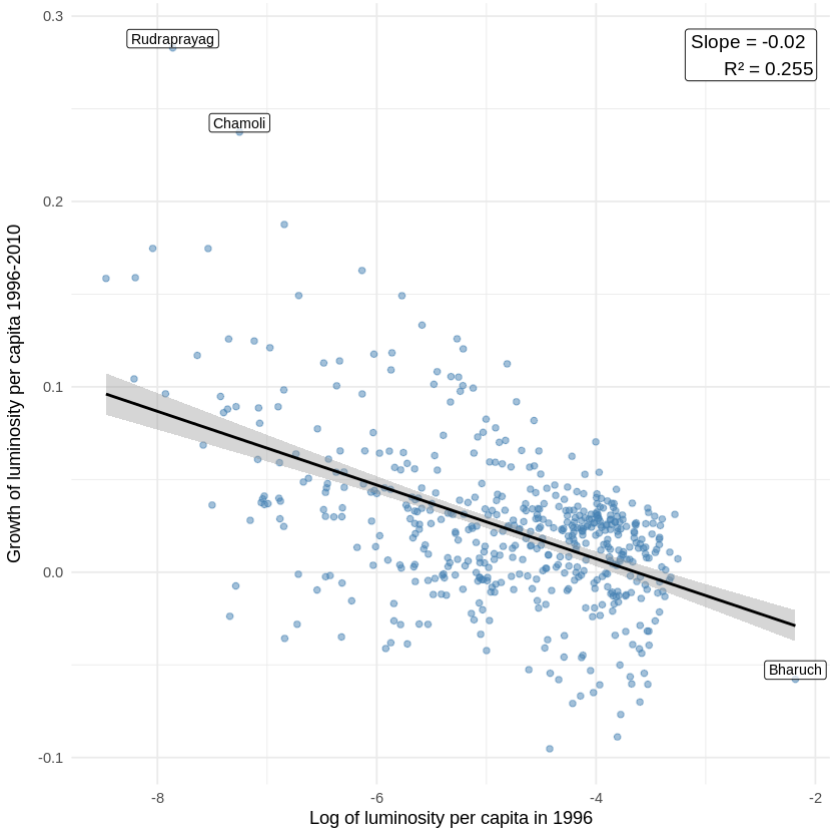

Figure 2: Regional luminosity convergence across districts in India Notes: Each point represents one of the 520 districts. The regression line shows the estimated beta-convergence relationship. Outlier districts are labeled. Source: Data from Chanda and Kabiraj (2020). See Regional convergence notebook for source code.

Next, we examine case studies of three economically disadvantaged states in India to illustrate their growth patterns over the study period. Among these, Bihar is the poorest, with a per-capita income at 39.2% of the national average, while Uttar Pradesh and Chhattisgarh have per-capita incomes of 43.8% and 52.3% of the national average, respectively.[3] Figure 3 displays the change in luminosity in these three states during our study period. Although these states remain among the poorest, there is a noticeable increase in luminosity over the course of the study.

3.2 Spatial dependence is a feature of the convergence process

Before proceeding with formal econometric analysis, it is instructive to examine the spatial distribution of the variables under study. Choropleth maps provide a natural tool for this purpose, as they allow researchers to visualize how a variable of interest varies across geographic units. They also help identify spatial patterns that may not be apparent from summary statistics alone. Figure 4 presents the spatial distribution of initial luminosity in 1996 and the subsequent growth rate of luminosity over the 1996–2010 period across the 520 districts in our sample.

[3]The data are taken from a report by the Economic Advisory Council to the Prime Minister (EAC-PM), released on September 18, 2024.

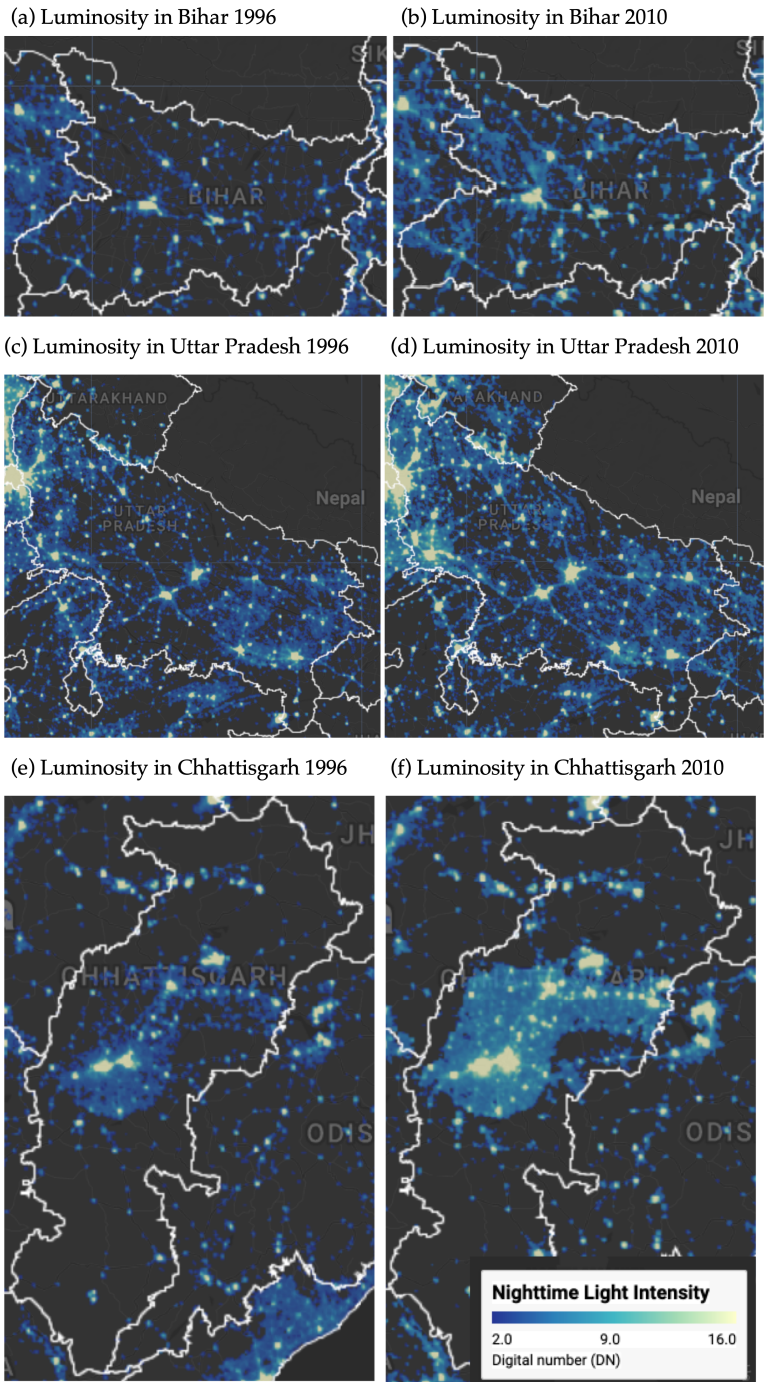

Figure 3: Some illustrative examples of regional convergence Notes: Luminosity is measured in radiance-calibrated digital number (DN) values from DMSP-OLS satellites. Interactive web application available at https://bit.ly/india-rc-ntl. Source: Authors' visualization using pre-processed luminosity images from the Earth Observation Group (NOAA/NCEI). See View from outer space notebook for source code.

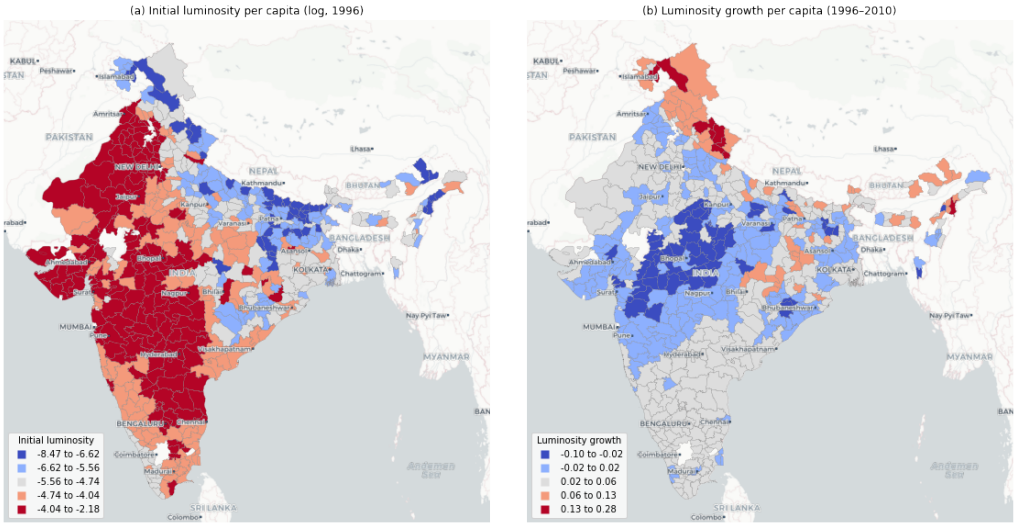

Figure 4: Spatial distribution of initial luminosity and luminosity growth Notes: Districts are classified into five categories using Fisher-Jenks natural breaks. Panel (a) shows log of luminosity per capita in 1996. Panel (b) shows luminosity growth per capita over 1996–2010. Source: Data from Chanda and Kabiraj (2020). See Spatial dependence notebook for source code.

The choropleth maps in Figure 4 reveal a distinct spatial pattern. Panel (a) shows that initial luminosity levels are concentrated in specific geographic corridors, with higher values clustered along western coastal areas while large portions of central and eastern India exhibit markedly lower luminosity. Panel (b), which displays the growth rate of luminosity per capita over the study period, presents a pattern that is largely the inverse of the initial distribution. Districts that were initially bright tend to exhibit lower growth rates, whereas districts that were initially dim tend to grow at a faster rate. This spatial inversion provides a first visual indication that regional convergence in India may have an important spatial dimension.

While the choropleth maps visually suggest the presence of spatial structures in both variables, a formal analysis of spatial dependence requires the definition of a neighborhood for each district. This neighborhood structure is specified through a spatial weight matrix, which encodes the connectivity between geographic units. In this study, we adopt a six nearest neighbors weight matrix, whereby for each of the 520 districts, the six geographically closest districts are identified as its neighbors. This connectivity structure is visualized as a network in Figure 5, where each node represents a district centroid and each edge connects a district to one of its six nearest neighbors.

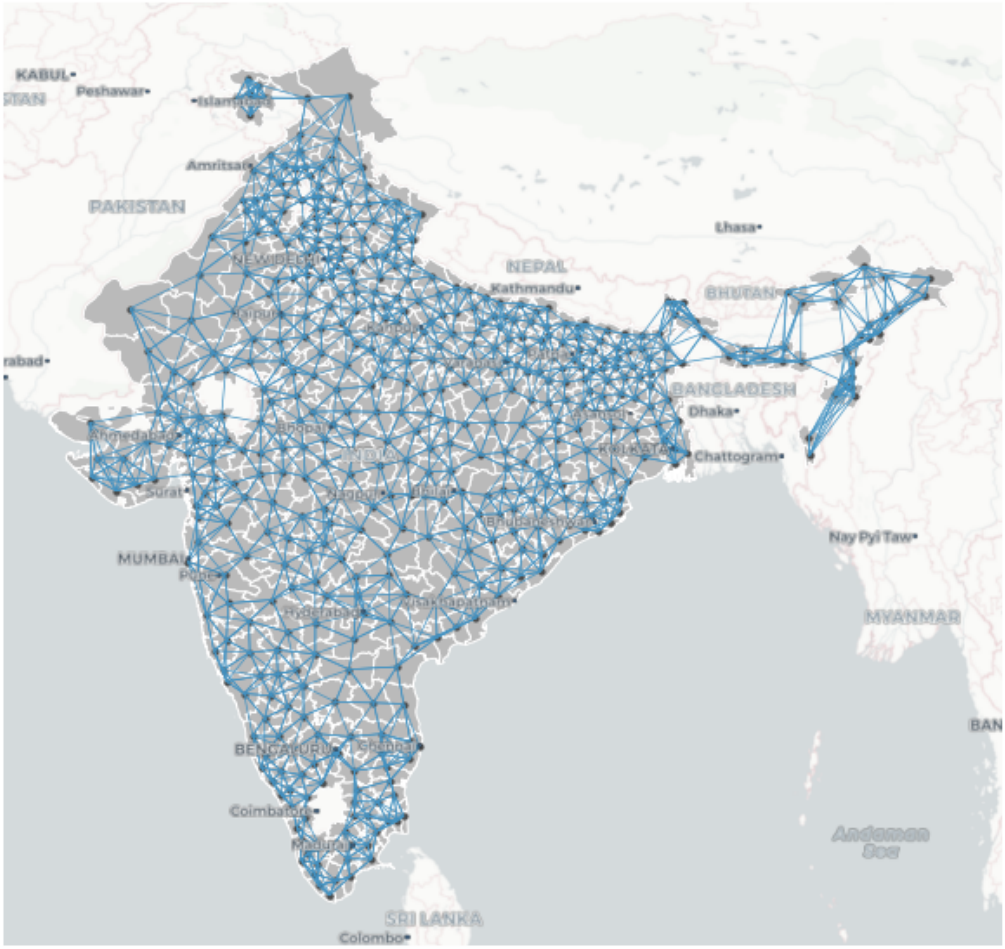

Figure 5: Spatial connectivity structure based on six nearest neighbors Notes: Each node represents a district centroid. Each edge connects a district to one of its six geographically closest neighbors. The weight matrix is row-standardized. Source: Data from Chanda and Kabiraj (2020). See Spatial dependence notebook for source code.

With the neighborhood of each district defined, we can formally assess the degree of spatial dependence in the variables. The notion of spatial dependence can be understood through two complementary concepts. The first is the overall degree of spatial clustering in the data, and the second is the specific locations of statistically significant clusters and spatial outliers. The Global Moran's I statistic captures the first of these concepts. It typically ranges from -1 (indicating perfect spatial dispersion) to $+1$ (indicating perfect spatial clustering), with values near zero suggesting spatial randomness. In the context of our data, the Moran's I for initial luminosity per capita is 0.73 ($p = 0.001$), indicating a strong degree of positive spatial autocorrelation. Districts with high (low) initial luminosity tend to be surrounded by districts that also exhibit high (low) initial luminosity. Similarly, the Moran's I for luminosity growth is 0.60 ($p = 0.001$), confirming that the growth rates of neighboring districts are also significantly correlated. Together, these statistics provide further evidence that both the initial level and subsequent growth of luminosity exhibit substantial spatial clustering.

While the Global Moran's I confirms the overall presence of spatial dependence, it does not reveal the location of statistically significant clusters and outliers. To address this limitation, we employ Local Indicators of Spatial Association (LISA). These indicators decompose the global statistic into district-level contributions and classify each district into one of four categories. High-High (HH) indicates a district with a high value surrounded by neighbors with similarly high values, while Low-Low (LL) indicates a low-value district surrounded by low-value neighbors. High-Low (HL) is a spatial outlier where a high-

value district is surrounded by low-value neighbors, and Low-High (LH) is the opposite spatial outlier. Figure 6 presents the Moran scatterplot and LISA cluster map for initial luminosity per capita. The cluster map identifies distinct geographic concentrations: HH clusters mark the most luminous regions and their bright neighbors, while LL clusters highlight contiguous areas of low luminosity, predominantly in central and eastern India.

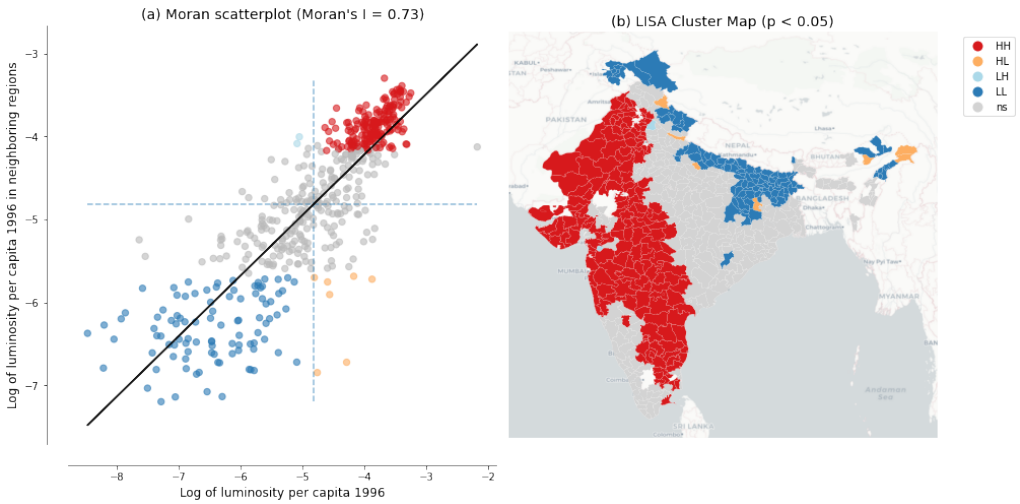

Figure 6: Spatial dependence in the initial level of luminosity Notes: Panel (a) shows the Moran scatterplot with Global Moran's I statistic. Panel (b) shows the LISA cluster map with statistically significant clusters at $p < 0.05$ based on 999 permutations. Source: Data from Chanda and Kabiraj (2020). See Spatial dependence notebook for source code.

The same LISA analysis applied to luminosity growth rates reveals a geographic pattern that is largely the inverse of the initial luminosity clusters, as shown in Figure 7. Regions that were classified as HH clusters in initial luminosity—the brightest districts and their neighbors—tend to appear as LL clusters in luminosity growth. This indicates that these initially prosperous areas experienced relatively slower growth over the study period. Conversely, districts that formed LL clusters in initial luminosity—the dimmest regions—tend to emerge as HH clusters in growth, reflecting faster catch-up growth in initially lagging areas. This spatial inversion between the initial level and subsequent growth of luminosity is the spatial signature of the convergence process: red regions in the initial luminosity map become blue regions in the growth map, and vice versa. These local spatial patterns provide visual evidence that spatial dependence is a prominent feature of the regional convergence process observed in India. This finding motivates the use of spatial econometric methods to formally account for these strong spatial interdependencies.

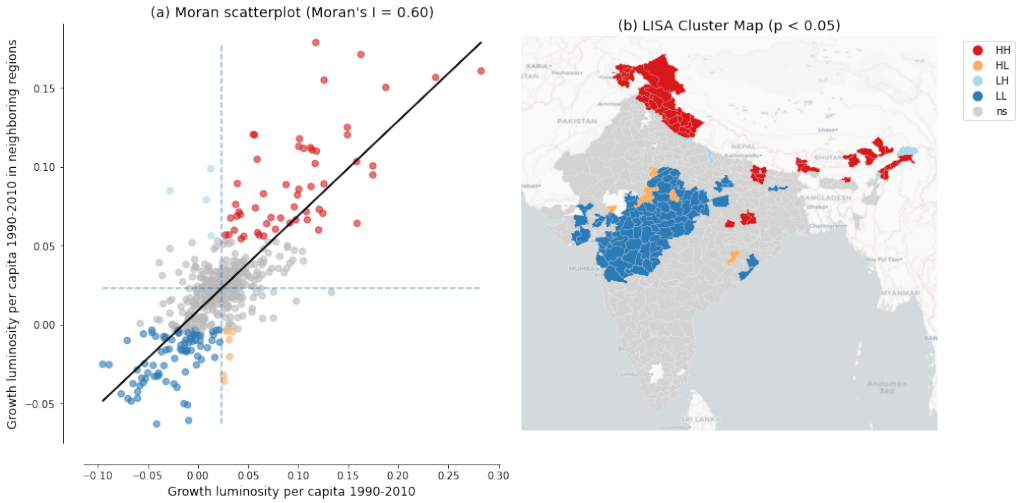

Figure 7: Spatial dependence in the growth rate of luminosity Notes: Panel (a) shows the Moran scatterplot with Global Moran's I statistic. Panel (b) shows the LISA cluster map with statistically significant clusters at $p < 0.05$ based on 999 permutations. Source: Data from Chanda and Kabiraj (2020). See Spatial dependence notebook for source code.

3.3 Evidence of spatial spillovers in regional convergence

The regression results of Table 1 provide evidence of both unconditional and conditional convergence across districts in India. Both conventional OLS and spatial econometric approaches indicate significant negative relationships between initial luminosity levels and subsequent growth rates. The direct effects, representing within-district convergence, remain stable across specifications, ranging from -0.020 to -0.026. This consistency across different model specifications and estimation methods suggests that poorer districts are catching up to their wealthier counterparts, even after controlling for various district characteristics and state-level fixed effects.

Table 1: Unconditional and conditional convergence across districts.

	Model 1		Model 2		Model 3		Model 4	
	OLS	SDM	OLS	SDM	OLS	SDM	OLS	SDM
Direct	-0.020***	-0.021***	-0.022***	-0.021***	-0.025***	-0.026***	-0.025***	-0.025***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Indirect	–	-0.001	–	-0.001	–	-0.015*	–	-0.012*
		(0.006)		(0.005)		(0.008)		(0.007)
Total	-0.020***	-0.022***	-0.022***	-0.022***	-0.025***	-0.041***	-0.025***	-0.037***
	(0.002)	(0.006)	(0.003)	(0.005)	(0.003)	(0.008)	(0.003)	(0.007)
Controls	No	No	No	No	Yes	Yes	Yes	Yes
State FE	No	No	Yes	Yes	No	No	Yes	Yes
AIC	-1945	-2290	-2413	-2466	-2211	-2356	-2469	-2499

The progression from unconditional to conditional specifications reveals how the estimated convergence process is shaped by the inclusion of additional covariates. In the simplest unconditional models (Models 1 and 2), direct effects range from -0.020 to -0.022. In the conditional models that incorporate district-level controls (Models 3 and 4), direct effects increase in magnitude to -0.025 and -0.026. This pattern suggests that omitting

district characteristics attenuates the estimated speed of convergence. Once we account for structural differences across districts—such as population density, urbanization, or sectoral composition—the underlying tendency for poorer districts to catch up becomes more pronounced. The inclusion of state fixed effects, which capture unobserved state-level heterogeneity such as differences in governance and institutional quality, further sharpens the estimates by absorbing variation that might otherwise confound the convergence relationship.

The spatial Durbin model also reveals spatial spillover effects that are not captured by traditional OLS estimations. These indirect effects, which capture the influence of neighboring districts' initial conditions on a district's growth rate, follow a notable pattern across specifications. In the unconditional models (Models 1 and 2), the estimated indirect effects are negligible and statistically insignificant, suggesting that without proper controls, spatial spillovers are confounded with omitted variables. However, once district-level controls are introduced (Models 3 and 4), the indirect effects become both larger in magnitude and statistically significant, reaching -0.015 in Model 3 and -0.012 in our most comprehensive Model 4. This emergence of significant spillover effects in the conditional specifications indicates that the spatial channels of convergence become discernible only after accounting for district-specific characteristics.

The total impact of initial conditions on growth, combining both direct and spillover effects, is substantially larger when we account for spatial dependence. In our fully specified model (Model 4), the total convergence effect in the spatial Durbin model (-0.037) is approximately 48% larger than the OLS estimate (-0.025). This gap is even more pronounced in Model 3, where the SDM total effect (-0.041) exceeds the corresponding OLS estimate (-0.025) by 64%, highlighting that spatial spillovers can constitute a considerable share of the overall convergence process. These differences indicate that conventional non-spatial approaches may significantly underestimate the speed of regional convergence by failing to capture the additional convergence channels created through spatial spillovers.

The model fit statistics further support the spatial econometric approach. The Akaike Information Criterion (AIC) consistently favors the spatial Durbin model over OLS across all four specifications. The most comprehensive model (Model 4 SDM) achieves the lowest AIC value of -2499 compared to -2469 for the corresponding OLS. Notably, the improvement from incorporating spatial structure is most pronounced in the unconditional specification (Model 1). The AIC drops by 345 units when moving from OLS to SDM, reflecting the large amount of spatial dependence left unmodeled by ordinary regressions. Even in the fully specified model, where state fixed effects already absorb much of the spatial heterogeneity, the SDM retains a meaningful advantage. Overall, these results suggest that regional convergence in India operates not only through district-specific factors but also through spatial interactions between neighboring districts. Ignoring these interactions leads to both underestimated convergence speeds and inferior model performance.

4 Discussion

4.1 Better luminosity data from VIIRS

Our analysis, following Chanda, Kabiraj (2020), relies on radiance-calibrated DMSP-OLS nighttime lights data covering the period 1996 to 2010. This dataset has been widely used as a proxy for economic activity (Henderson et al. 2012, Chen, Nordhaus 2011). However, DMSP-OLS data are subject to well-documented limitations, including top-coding in bright urban cores and lack of on-board calibration leading to inter-satellite inconsistencies. Blooming artifacts that spatially blur light sources beyond their true boundaries represent an additional concern (Abrahams et al. 2018). These measurement issues can attenuate the precision of convergence estimates, particularly in rapidly urbanizing districts.

The Visible Infrared Imaging Radiometer Suite (VIIRS), operational since 2012, represents a marked improvement over DMSP-OLS along multiple dimensions. VIIRS offers finer spatial resolution (approximately 750 meters versus 2.7 kilometers) and on-board radiometric calibration that provides consistent quantitative measurements. It also

features a wider dynamic range that avoids saturation in urban areas while detecting dim lights in rural settlements (Elvidge et al. 2017). Systematic assessments recommend VIIRS as the preferred product for cross-sectional and recent time-series studies (Gibson et al. 2021). Future extensions of our convergence analysis using VIIRS data could yield more precise estimates of both direct and indirect effects, particularly in districts where blooming effects may have distorted the true spillover effects.

A practical challenge for extending long-run convergence studies is the discontinuity between DMSP (1992–2013) and VIIRS (2012–present) sensor eras. Recent harmonization efforts, notably the global harmonized nighttime light dataset by Li et al. (2020), have created consistent long-run time series by calibrating VIIRS observations to DMSP-equivalent units during the overlap period. Such harmonized datasets could enable the extension of our spatial Durbin analysis to more recent periods. This would allow researchers to examine whether the convergence patterns and spatial spillovers documented here have persisted, accelerated, or changed in character as India's economy has continued to transform.

4.2 New research directions

While our analysis documents the average convergence effect and its spatial spillover component, the cross-sectional regression framework does not capture potential heterogeneity in convergence patterns across the income distribution. Distribution dynamics approaches, as pioneered by Quah (1996), could reveal whether Indian districts are converging to a single steady state or forming distinct convergence clubs where districts converge within groups but diverge across them. The regression tree methods developed by Durlauf, Johnson (1995) for identifying multiple growth regimes could be combined with spatial econometric techniques. Such an approach could examine whether geographic clusters of districts follow distinct convergence trajectories, potentially revealing spatial poverty traps or growth poles that are not visible in average convergence estimates (Rey, Montouri 1999). Furthermore, analysis of harmonized luminosity data across Chinese provinces has uncovered complex inequality dynamics that differ markedly across cross-sectional and temporal dimensions (Glawe, Mendez 2024). Extending these insights to the Indian district context could help determine whether the convergence patterns we document reflect a single equilibrium or mask the formation of distinct spatial clubs.

Another important direction concerns the causal identification of the spillover channels that our spatial Durbin model captures in reduced form. While our estimates document significant indirect effects, the model does not identify whether these spillovers operate through infrastructure linkages, labor migration, technology diffusion, or market access channels. Quasi-experimental approaches, such as those employed by Asher, Novosad (2020) to study the causal effects of rural road construction on structural transformation in India, could be embedded within spatial econometric frameworks to isolate specific spillover mechanisms. Understanding which channels drive the indirect convergence effects is important for designing spatially targeted policies that may amplify positive spillovers.

Comparative evidence from other developing economies suggests that spatial dependence in regional convergence is a widespread phenomenon rather than an India-specific feature. Studies have documented significant spatial spillover effects in convergence processes across Thailand (Tipayalai, Mendez 2024), Turkey (Ursavas, Mendez 2023), and Indonesian districts (Miranti, Mendez 2023), with neighbor effects and spatial conditioning factors playing important roles in shaping convergence trajectories. This cross-country regularity strengthens the case for systematic investigation of the specific mechanisms driving spatial spillovers, as the channels may differ across institutional and geographic contexts.

The growing availability of diverse satellite products and machine learning methods opens possibilities for richer measurement of regional economic activity. Jean et al. (2016) showed that combining high-resolution daytime imagery with nighttime lights through deep learning can considerably improve poverty prediction in data-scarce settings. Similarly, Keola et al. (2015) showed that integrating nighttime lights with land cover data improves economic measurement in agricultural areas where lights alone provide weak signals. These alternative data sources enable the construction of multi-dimensional

proxies for regional economic activity that go beyond what nighttime lights alone can capture. Recent applications demonstrate the practical potential of these advances for subnational economic measurement. Chen et al. (2024) show that higher-quality VIIRS nighttime lights can predict sectoral GDP composition across Turkish provinces, distinguishing between urban service-oriented and rural agricultural regions in ways that DMSP data cannot. Hussein et al. (2025) employ machine learning methods with multiple remote sensing indicators to predict subnational GDP in Vietnam, achieving accuracy improvements over traditional luminosity-based approaches. At a broader scale, Khoun et al. (2025) combine big data sources, socioeconomic surveys, and machine learning to map multidimensional poverty in Cambodia, illustrating how satellite-derived features can complement conventional survey instruments. Applying similar multi-source approaches to Indian districts could yield richer proxies for economic activity that overcome the well-known limitations of nighttime lights in agricultural and low-density areas.

4.3 Research reproducibility and open science

The complexity of satellite-based economic research—involving multi-step data processing pipelines, spatial econometric estimation, and geographic visualization—makes reproducibility both challenging and essential. As Donaldson, Storeygard (2016) note, the processing of satellite data requires careful documentation and sharing of code to ensure that results can be replicated and extended. Each methodological choice in the pipeline, from sensor inter-calibration to spatial weight matrix construction, can affect empirical conclusions. This underscores the need for transparent computational workflows that allow other researchers to verify and build upon published findings.

This article adopts a reproducible research approach through Jupyter notebooks and the Quarto publishing framework. All computational analyses are documented in embedded notebooks that readers can inspect alongside the results they produce. The interactive visualization tool, built on Google Earth Engine, allows researchers to explore the spatial and temporal patterns in satellite nighttime light data.

Open-source tools and cloud computing platforms are rapidly lowering the barriers to reproducible spatial economic research. Cloud-based computational notebooks, such as those developed by Mendez, Patnaik (2024) for processing nighttime lights data, eliminate the need for specialized local software and provide accessible workflows for data ingestion, preprocessing, spatial analysis, and visualization. The combination of open data repositories, version-controlled code, and cloud computing infrastructure creates an ecosystem where the full analytical pipeline—from satellite imagery to econometric results—can be made transparent, verifiable, and extensible by the broader research community.

5 Concluding remarks

This article re-examines the regional convergence hypothesis across Indian districts using satellite nighttime light data, interactive visualizations, and spatial econometric modeling. Building on the work of Chanda, Kabiraj (2020), we developed an interactive web-based visualization tool that illustrates spatial and convergence patterns across Indian districts. Spatial autocorrelation analyses indicate that spatial dependence is a notable characteristic of satellite data and the regional convergence process in India. Estimates from our spatial Durbin model indicate that incorporating spatial spillovers increases the estimated speed of regional convergence. The total convergence effect in our fully specified model is approximately 48% larger than conventional non-spatial estimates. This finding suggests that non-spatial convergence models may underestimate the speed of regional convergence. Additionally, it suggests that place-based development interventions may have broader impacts, as their benefits can extend to neighboring districts through spatial spillover effects.

Our results also illustrate the usefulness of satellite nighttime lights for studying economic dynamics in countries where subnational data are scarce, infrequent, or unreliable. Conventional economic statistics at the district level are often unavailable or inconsistent

across administrative boundaries, particularly in large developing economies. Radiance-calibrated nighttime light data allowed us to analyze 520 Indian districts at a spatial granularity that would be difficult to achieve with traditional national accounts. This data-driven approach is potentially transferable to other data-poor contexts across the developing world. The ongoing transition from DMSP-OLS to the higher-resolution VIIRS sensor should yield more precise measurement of subnational economic activity in future studies.

This study also illustrates how reproducible open science practices can strengthen the credibility and extendability of scientific research. The entire analytical pipeline is documented in computational Jupyter notebooks using Python, R, and Stata code. Research results are automatically embedded within the manuscript through Quarto's publishing framework. Moreover, a single manuscript source generates multiple output formats: HTML, PDF, Microsoft Word, among others. The HTML version is particularly useful as it allows readers to engage with interactive visualizations, easily inspect the code of multiple computational notebooks, and reevaluate the results in light of their source code. By hosting the complete codebase and data in a public GitHub repository and making the Python and R notebooks executable in the cloud through Google Colaboratory, we aim to encourage other researchers to further adopt reproducible open science practices that can strengthen the credibility and extendability of scientific research.

6 Acknowledgments

This research project was supported by JSPS KAKENHI Grant Number 24K04884. During the preparation of this manuscript, the authors acknowledge the use of Claude Code (Anthropic) to assist with manuscript editing, computational notebook development, and research infrastructure setup. After using this AI tool, the authors reviewed the outputs, confirmed their accuracy, and take full responsibility for the content of this publication.

7 Conflict of Interest

The authors declare no conflict of interest.

8 Data and Code Availability

All data and computational code used in this study are available in the project repository: [Repository URL removed for blind review]. Also, the interactive HTML version of this manuscript (https://quarcs-lab.github.io/project2025s/) embeds the computational notebooks, allowing readers to inspect the complete analytical pipeline from raw data to published results.

References

Abrahams A, Oram C, Lozano-Gracia N (2018) Deblurring dmsp nighttime lights: A new method using gaussian filters and frequencies of illumination. *Remote Sensing of Environment* 210: 242–258. CrossRef

Adhikari B, Dhital R (2020) Decentralization and regional convergence: Evidence from nighttime lights. *Journal of Development Economics* 145: 102467. CrossRef

Allaire J, Teague C, Scheidegger C, Xie Y, Dervieux C (2024). Quarto

Anselin L (1988) *Spatial Econometrics: Methods and Models*. Springer. CrossRef

Anselin L (1995) Local indicators of spatial association—LISA. *Geographical Analysis* 27: 93–115. CrossRef

Asher S, Novosad P (2020) Rural roads and local economic development. *American Economic Review* 110: 797–823. CrossRef

Barro RJ, Sala-i Martin X (1992) Convergence. *Journal of Political Economy* 100: 223–251. CrossRef

Beyer RCM, Jain T, Sinha BN (2021) Lights out? covid-19 containment policies and economic activity. *Journal of Asian Economics* 74: 101318. CrossRef

Chakravarty S, Dehejia R (2019) The gst and regional divergence in india: A nightlights perspective. *Economic and Political Weekly* 54: 47–54

Chanda A, Cook J (2022) Demonetization and its redistributive effects: Evidence from india. *Journal of Development Economics* 156: 102835. CrossRef

Chanda A, Kabiraj S (2020) Shedding light on regional growth and convergence in india. *World Development* 133: 104961. CrossRef

Chen X, Nordhaus WD (2011) Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences* 108: 8589–8594. CrossRef

Chen Y, Ursavas U, Mendez C (2024) Can higher-quality nighttime lights predict sectoral GDP across subnational regions? Urban and rural luminosity across provinces in Türkiye. *Letters in Spatial and Resource Sciences* 17: 1–21. CrossRef

Cook J, Shah M (2022) Aggregate effects from public works: Evidence from india. *Review of Economics and Statistics* 104: 201–217. CrossRef

Donaldson D, Storeygard A (2016) The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives* 30: 171–198. CrossRef

Durlauf SN, Johnson PA (1995) Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics* 10: 365–384. CrossRef

Elvidge CD, Baugh KE, Zhizhin M, Hsu FC, Ghosh T (2017) Viirs night-time lights. *International Journal of Remote Sensing* 38: 5860–5879. CrossRef

Ertur C, Koch W (2007) Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics* 22: 1033–1062. CrossRef

Fischer MM (2011) A spatial mankiw-romer-weil model: Theory and evidence. *Annals of Regional Science* 47: 419–436. CrossRef

Gibson J, Olivia S, Boe-Gibson G, Li C (2021) Which night lights data should we use in economics, and where? *Journal of Development Economics* 149: 102602. CrossRef

Glawe L, Mendez C (2024) Harmonized luminosity and economic activity across provinces in China: Cross-sectional differences, regional time series, and inequality dynamics. *Applied Economics*. CrossRef

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202: 18–27. CrossRef

Henderson JV, Storeygard A, Weil DN (2012) Measuring economic growth from outer space. *American Economic Review* 102: 994–1028. CrossRef

Hussein S, Nguyen MTT, Mendez C (2025) Predicting subnational GDP in Vietnam with remote sensing data: A machine learning approach. *Letters in Spatial and Resource Sciences* 18: 1–12. CrossRef

Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353: 790–794. CrossRef

Jha S, Talathi A (2021) Colonial institutions and long-run development in india: Evidence from nightlights. *Journal of Development Economics* 153: 102726. CrossRef

Keola S, Andersson M, Hall O (2015) Monitoring economic development from space: Using nighttime light and land cover data to measure economic growth. *World Development* 66: 322–334. CrossRef

Khoun T, Poortinga A, Thwal NS, Gonzalez de Alba J, McMahon A, Mendez C (2025) Mapping the dimensions of poverty through big data, socioeconomic surveys and machine learning in Cambodia. *Social Indicators Research* 180: 1593–1618. CrossRef

Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S et al. (2016) Jupyter notebooks—a publishing format for reproducible computational workflows. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87–90. IOS Press

Knuth DE (1984) Literate programming. *The Computer Journal* 27: 97–111. CrossRef

LeSage JP, Pace RK (2009) *Introduction to Spatial Econometrics*. Chapman and Hall/CRC. CrossRef

Lessmann C, Seidel A (2017) Regional inequality, convergence, and its determinants: A view from outer space. *European Economic Review* 92: 110–132. CrossRef

Li X, Zhou Y, Zhao M, Zhao X (2020) A harmonized global nighttime light dataset 1992–2018. *Scientific Data* 7: 168. CrossRef

Mendez C, Patnaik A (2024). A python notebook for processing nighttime lights data: Methods and applications. GitHub Repository. Accessed: 2026-02-11

Miranti C, Mendez C (2023) Social and economic convergence across districts in Indonesia: A spatial econometric approach. *Bulletin of Indonesian Economic Studies* 59: 421–445. CrossRef

Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23. CrossRef

Peng RD (2011) Reproducible research in computational science. *Science* 334: 1226–1227. CrossRef

Pinkovskiy M, Sala-i Martin X (2016) Lights, camera, income! illuminating the national accounts-household surveys debate. *Quarterly Journal of Economics* 131: 579–631. CrossRef

Quah DT (1996) Twin peaks: Growth and convergence in models of distribution dynamics. *Economic Journal* 106: 1045–1055. CrossRef

Rey SJ, Montouri BD (1999) Us regional income convergence: A spatial econometric perspective. *Regional Studies* 33: 143–156. CrossRef

Solow RM (1956) A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70: 65–94. CrossRef

Tamiminia H, Salehi B, Mahdianpari M, Quackenbush LJ, Adeli S, Brisco B (2020) Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing* 164: 152–170. CrossRef

Tipayalai K, Mendez C (2024) Regional convergence and spatial dependence in Thailand: Global and local assessments. *Journal of the Asia Pacific Economy* 29: 693–720. CrossRef

Ursavas U, Mendez C (2023) Regional income convergence and conditioning factors in Turkey: Revisiting the role of spatial dependence and neighbor effects. *Annals of Regional Science* 71: 363–389. CrossRef