Internalizing the Negative Externality from Urbanization: Evidence from India's Rurban Mission

Somdeep Chatterjee* Manhar Manchanda† Shubham Ojha‡

March 20, 2025

Abstract

Despite several advantages, urbanization essentially results in market failure due to creation of a negative externality through increased pollution levels. This makes policymaking difficult for emerging economies which need to tackle the twin objective of increasing growth and development while also keeping pollution under control. In this paper, we study a unique 'place-based-policy' from India which created rural-urban (Rurban) clusters to adopt sustainable ways of achieving the benefits of urbanization while preserving the essence of rural lifestyle. Employing a synthetic difference-in-differences estimation approach, we show that this helps internalize the negative externality considerably while not compromising on the positive impacts of urbanization, through potential reduction in costs of abatement under a Rurbanization mission, relative to standard urbanization approaches.

JEL: R11, Q52, Q53, O12

Keyowrds: rurban; urbanization; pollution; nighttime lights; economic development

*Indian Institute of Management Calcutta

†Indian Institute of Management Calcutta

‡Indian Institute of Management Calcutta

1 Introduction

Historical arguments in favor of *urbanization* have been that it leads to economic growth and development (Henderson et al., 2003, Scott, 2009, Brückner, 2012, Moomaw and Shatter, 1996, Henderson et al., 2003), is correlated with infrastructure development (Heshmati and Rashidghalam, 2020, Maparu and Mazumder, 2017, Pandey et al., 2022, Pradhan et al., 2021), creates job opportunities (Sato and Zenou, 2015, Finney and Kohlhase, 2008, Henderson et al., 2003, Amar et al., 2018) and helps leverage efficient delivery of services (Kolko, 2010, Mendez et al., 2023, Spence et al., 2008, Quigley, 2009, Schafran et al., 2018). Theoretically, however, urbanization strategies naturally lead to market failures due to creation of negative externality through increased levels of pollution as a consequence of rapid urbanization (Liang et al., 2019, Wang et al., 2020, Zhao et al., 2022, Chen et al., 2024, Martin and Zhang, 2021, Liu et al., 2022). This essentially creates a policy conundrum in terms of whether governments in developing countries should focus on urbanization to ensure that its benefits diffuse into the economy or attempt to reduce the impact of the negative externality by deliberately slowing down the process of urbanization.

Efforts towards adopting sustainable models of urbanization have been made in various developing countries (Tan et al., 2016, Ochoa et al., 2018, Chen et al., 2024, Jiboye, 2011). Many of these efforts rely on creating optimal emission policies and maintaining city sizes (Pflüger, 2021) as well as adoption of green forms of transport (Kahn and Schwartz, 2008). While the literature on the impacts of sustainable urbanization approaches is growing (Zhang et al., 2020), majority of these are focused on urban logistics, infrastructure, construction planning and the synergy between these. The literature on the impact of broader government level programs such as clustered development approaches and *place-based-policies* (PBPs) is fairly limited (Xu and Luo, 2025).

The Shyama Prasad Mukherji Rurban Mission (SPMRM), launched in India in 2016, is such a unique PBP which emphasized on creation of "*rurban*" clusters which are networks

of contiguous villages designed to interact the nuances of rural sustainability with the basic advantages of urbanization. The foundation of the National Rurban Mission (NRuM) focuses on this interplay between clustering and sustainability. The policy design draws upon the classical idea of industrial clusters (Porter et al., 1998) and urban agglomeration economies (Henderson et al., 2003, Marshall, 1890), which suggest that a natural consequence of clustering is being able to leverage scale economies. Additionally, the mission specifically incorporates sustainability by emphasizing the creation of cluster-level public goods and promoting environmentally sustainable projects.

In general, PBPs of this nature have increasingly emerged as a popular instrument to boost regional growth and lower inter-country disparities in developing economies (Chaurey, 2017, Lu et al., 2019, Shenoy, 2018). According to a report by UNCTAD, the number of special economic zones (SEZs)- the most common form of PBP- increased from 79 in 1975 to 5400 in 2018 across the world, with a majority of these being housed in developing economies.[1] These policies typically take the form of targeted tax exemptions and government-led investments in infrastructure in underdeveloped regions (Garg and Shenoy, 2021, Busso et al., 2013, Ehrlich and Seidel, 2018). While much of the literature on place-based policies has focused on special economic zones or development zones with the standard relaxed regulatory environments, the economic impact is commonly documented to be positive (Hasan et al., 2021, Koster and Van Ommeren, 2019, Lu et al., 2019, Gallé et al., 2024, Zhao and Qu, 2024).To our knowledge, the combined impacts on development and pollution of such cluster approach driven PBPs has not been studied to date.

To bridge this gap in the literature, in this paper, we estimate the effects of Rurban Mission on environmental outcomes and economic activity. We exploit plausibly exogenous geographic and temporal variation in the program's implementation. We use satellite-based data on nighttime lights as a proxy for economic growth and PM2.5 levels as an

[1]See the World Investment Report 2019 by UNCTAD here.

indicator of pollution. The institutional features of the implementation of the policy allows us to employ a staggered synthetic difference-in-differences (SDiD) methodology. Using this identification strategy, we find that while the policy led to an increase in nighttime lights, it also resulted in a 0.14 $\mu g/m^3$ reduction in mean PM2.5 levels and a 0.16 $\mu g/m^3$ reduction in maximum PM2.5 exposure. In relative terms, this corresponds to a 0.22% decrease in average PM2.5 exposure.

Importantly, our analysis of heterogeneous effects reveals that the reduction in pollution was much more pronounced in tribal clusters, where the decline in mean PM2.5 was 0.67 $\mu g/m^3$, compared to a smaller 0.14 $\mu g/m^3$ (p ¡ 0.10) in non-tribal clusters. This suggests that pre-existing institutions in tribal areas potentially amplify the environmental benefits of the program. The SDiD estimates indicate that the log of mean nighttime lights increased by 1.6 percentage points. The effects on nighttime luminosity growth were similar across tribal and non-tribal clusters, suggesting that both benefited economically from the intervention. Given an estimated elasticity of GDP to nighttime lights of 0.10-0.15 (Baskaran et al., 2024), this implies a 0.15-0.23 percentage point increase in regional GDP growth.

We examine the provision and types of public goods disbursed under the program to explore potential mechanisms explaining our reduced form findings. We find that treated villages experienced significant improvements in public facilities, including increased access to watershed management and improved drainage systems, which may have contributed to pollution reduction. Further, reports suggest that 10-26% of approved projects under the mission were explicitly environment-friendly, including initiatives in renewable energy (solar, wind), water conservation, and afforestation. This strategic emphasis on sustainability likely helped balance economic gains with environmental preservation. Theoretically, by focusing on shared infrastructure, resource pooling, and improved regulatory coordination, it could be argued that the program lowered the marginal cost of pollution abatement, leading to reduced environmental degradation. This helps partially internalize the negative externality arising out of urbanization, while not compromising on the positive impacts on

economic growth and development.

This paper contributes to two important strands in the growing literature on the effectiveness of place-based policies. First, while place-based policies have been extensively studied in urban contexts—such as enterprise zones (Kline and Moretti, 2014, Neumark and Simpson, 2015) —there is limited causal evidence on their effectiveness in rural settings. By demonstrating that targeted rural development can yield positive economic and environmental returns, our study informs debates on spatial inequality and rural development strategies (Jedwab et al., 2017). Second, the paper contributes to the literature on governance and local capacity by showing that institutional factors mediate the effectiveness of development policies. The stronger pollution reduction effects in tribal clusters suggest that pre-existing governance structures and community-led resource management play a crucial role in amplifying policy benefits. This complements the broader literature on institutional quality and public goods provision, which highlights how decentralized governance can improve policy outcomes (Bardhan and Mookherjee, 2006, **?**).

2 Background

India is predominantly rural, with nearly 830 million (68.8% of the total population) people residing in villages. According to the 2011 Census of India, there were a total of 649,481 villages spread across 5,924 sub-districts, 640 districts, 28 States and 8 Union Territories.[2] On the administrative front, a village is the smallest administrative unit in India, composed of a single or a collection of habitation clusters or sub-villages. Although villages function as independent administrative units, a large fraction of them are spatially contiguous and form clusters of inhabited settlements. This paper examines a unique place-based policy, the *Rurban* Mission, which leverages this natural contiguity of villages to drive de-

[2]In the Indian administrative setup, districts are nested within states; sub-districts are nested within districts, and villages are completely contained within sub-districts. The average population of a village is approximately 2,000 while the average population of a sub-district is nearly 200,000.

velopment. Through large-scale infrastructural investments and pooling access points for existing policies, the policy fosters coordinated, cluster-based growth among spatially contiguous villages. In this section, we provide a brief overview of the policy's key features and the exogenous variation it generates in a village's exposure to the program, enabling us to causally estimate its impact on various outcomes.

2.1 The *Rurban* Mission: Policy Background

Launched in 2016, the Shyama Prasad Mukherji Rurban Mission (SPMRM), or the *Rurban* Mission, was allocated a substantial budget of 51.42 billion INR ($\sim$ 591 million USD) to develop 300 *Rurban* clusters, comprising of multiple contiguous villages, across three phases over five years. These clusters are selected as follows. *First*, the central Ministry of Rural Development provides a list of *selected* sub-districts, based on a range of socioeconomic criteria, to each State within which the clusters are to be identified. *Second*, the relevant state government authorities identify, within each sub-district, a large village or Gram Panchayat (GP) as a potential growth center possessing resources available in the area that could potentially lead the region's economic transformation. *Third*, the clusters are then formed by identifying geographically contiguous villages/GPs within a 5–10 km radius (or a radius appropriate to the population density and geography of the region) around the identified growth center. In the Online Appendix B.1, we present a detailed list of the criteria utilized to select the *Rurban* clusters.

The broad objective of the policy was to facilitate the development of clusters of villages that (i) preserve and nurture the essence of rural community life, focusing on equality and inclusiveness, and (ii) provide facilities that are an essential part of urban lifestyles. In line with this vision, the policy specifies four key outcomes: (i) bridging the rural-urban divide in terms of economic development, technology, and access to essential facilities and services; (ii) stimulating local economic development with an emphasis on reducing poverty and unemployment in rural areas; (iii) spreading development across the region; (iv) attracting

investment in rural areas.

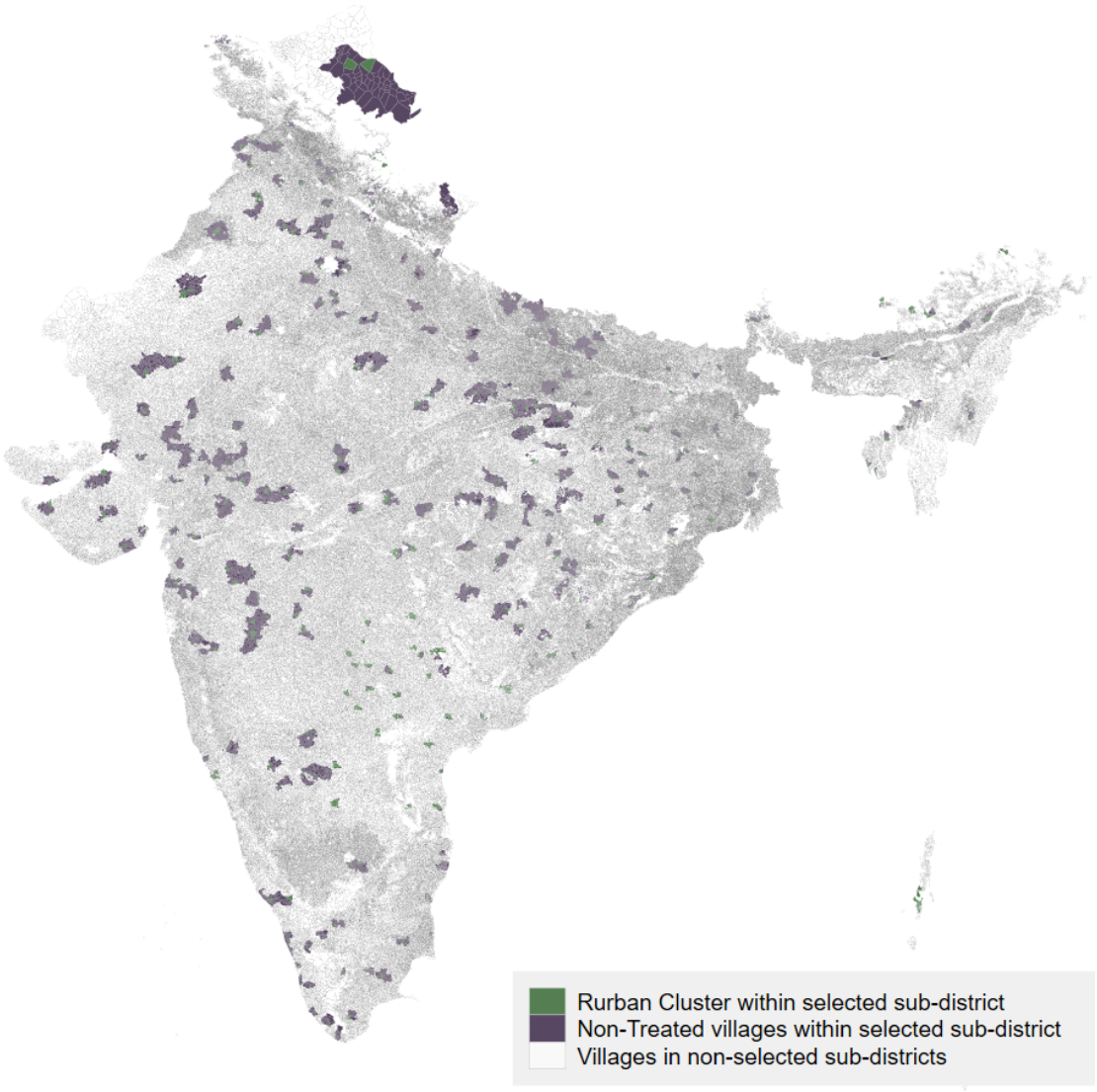
To achieve these holistic outcomes, the *Rurban* mission adopts a multi-faceted approach. While a major component of the program involves investments to create assets and public facilities, the policy in essence focuses on three main aspects for achieving these development goals within these clusters: *first*, pooling access points for existing policies; *second*, creating an integrated planning framework for each cluster with specified project goals and outcomes, and finally, introducing an additional layer of administration within the existing governance hierarchy to facilitate efficient management and implementation of projects. We discuss the interpretation of the policy vis-à-vis common place-based policies in Section. Additionally, in the Online Appendix B.2, we provide administrative details of each of these aspects and their implementation.

2.2 Treated Clusters and Investments

Between 2016 and 2018, 297 clusters comprising approximately 4,500 villages, across 254 administrative districts, received funds in three phases from the Ministry of Rural Development under the initiatives related to infrastructural investments, convergence of schemes and critical gap funding for existing policies. The first-phase of the program, beginning in 2016, included investments in 1,647 villages spread across 108 sub-districts. The second-phase (2017) extended the program to 1,559 additional villages in 101 sub-districts, while the final third phase introduced in 2018 covered 1,461 villages in 120 sub-districts. Note that *Rurban* villages comprise a small fraction of the total number of villages in India. This is documented in Figure 1 where we plot the (i) *Rurban* villages within selected sub-districts, (ii) non-treated villages within selected sub-districts, and (iii) villages in non-selected sub-districts. To get a better idea of the spatial distribution of these villages, we illustrate the distribution of treated villages within selected sub-districts in the northern state of Punjab as an example in Figure 2. Of the state's 77 sub-districts, the Ministry of Rural Development selected 8 for the Rurban Mission, wherein the state government in-

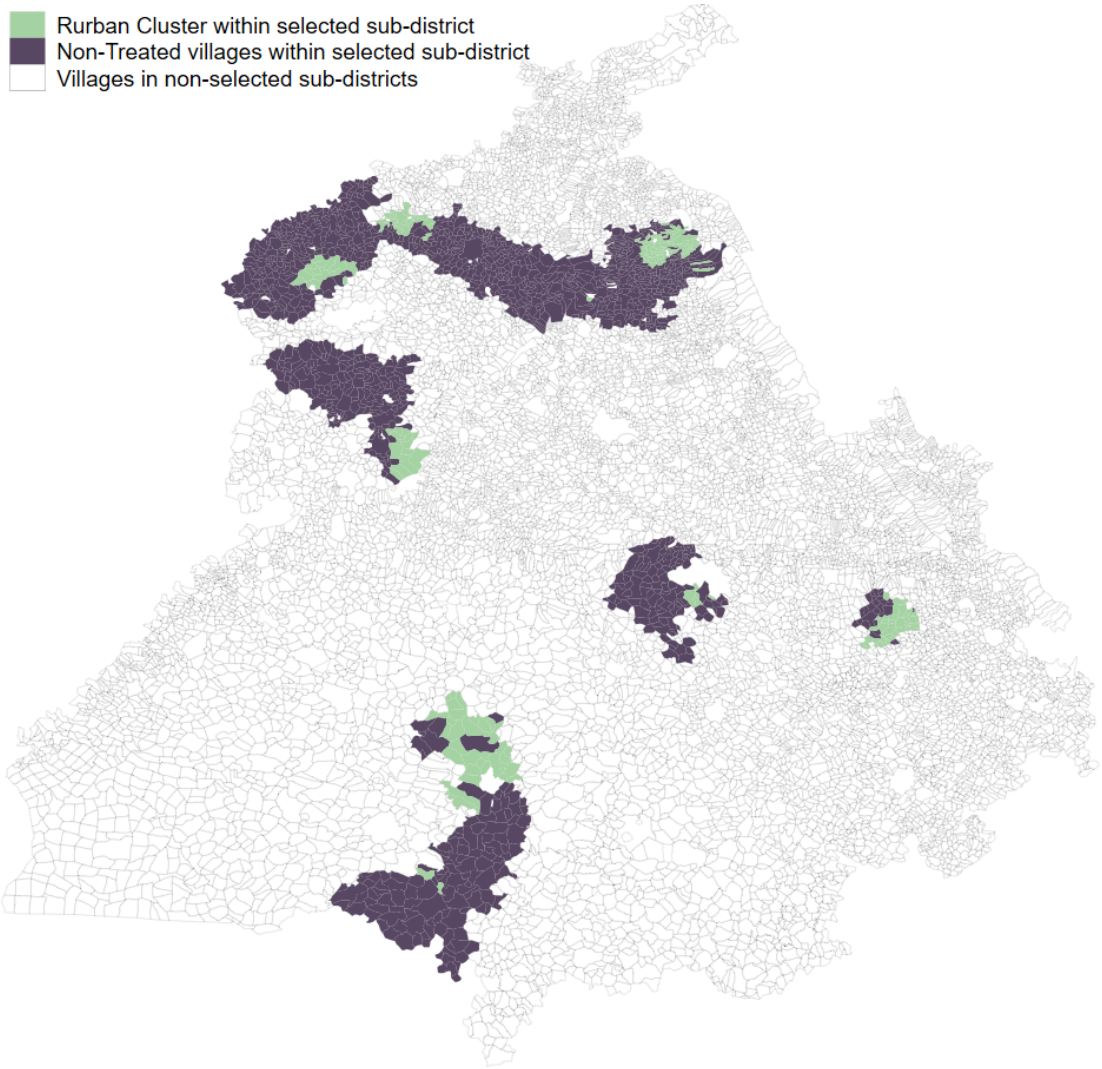
corporated 216 villages— 15.3% of all villages in these sub-districts— into Rurban clusters across three phases. This number represents an even smaller fraction when compared to Punjab's total villages (12,379). Nonetheless, for our empirical design, we use non-treated villages within the selected sub-districts as the comparison group for Rurban villages. This approach ensures sufficient statistical power while ensuring credibility of our comparisons. Further details on the empirical strategy are provided in Section.

Figure 1: *Rurban* Clusters across India



The map shows the spatial distribution of *Rurban* clusters across India. The green-colored villages correspond to the *Rurban* clusters, while the purple-colored regions correspond to non-treated villages in *selected* sub-districts. The gray areas comprise all those villages that are not part of the selected sub-districts under the Rurban Mission

Figure 2: Punjab: *Rurban* Clusters



The map shows the spatial distribution of *Rurban* clusters in the state of Punjab. The green colored villages correspond to the *Rurban* clusters, while the purple-colored regions correspond to non-treated villages in *selected* sub-districts. The gray areas comprise all those villages that are not part of the selected sub-districts under the Rurban Mission

As of 2024, the total expenditure incurred on public works sanctioned under the program amounted to 182.17 billion INR, with 44,231 public works completed (71.8% of the total approved works). These projects encompass a broad range of facilities, from 'basic amenities' like street lighting, waste management for households and villages, and intra- and inter-village roads to major 'economic amenities' such as green technology-based public transport, agricultural extension services, and skill training programs. Notably, most of

these public works are a product of consolidating access points for existing policies and aligning schemes across various government departments to enhance the provision of these amenities.

2.3 Rurban Mission: A Unique Place-Based Policy?

3 Data and Variable Construction

In this section, we describe the data sources and construction of the primary variables of interest to our analysis, and present summary statistics disaggregated by treated and non-treated villages.

3.1 Rurban Treatment Data

We scrape information on the treated clusters from the official online portal of the Rurban Mission.[3] Using administrative documents from the Ministry of Rural Development, we are able to identify the specific villages and/or Gram Panchayats (GPs) treated under the policy, along with their corresponding districts and the phase of implementation. These documents additionally include unique identifying codes for villages and GPs, enabling us to match them with other datasets that use the standardized codes from the 2011 Census. To match these codes to the Census codes, we utilize the online Government of India repository, *Local Government Directory* (LGD), and create a consolidated dataset linking each LGD village/GP to its Census counterpart, achieving an accurate match for approximately 93% of the total villages.[4]

[3] All relevant documents are available here: https://rurban.gov.in/.

[4] The Local Government Directory (LGD) - https://lgdirectory.gov.in is one of the 11 applications developed as part of Panchayat Enterprise Suite (PES) under e-Panchayat Mission Mode project (MMP) by the Government of India with the Ministry of Panchayati Raj. The adoption of LGD codes as the Standard location code in the e-Governance applications of Ministries/Departments/States/UTs was mandated by CabinetSecretariat vide letter dated 04th November 2016. However, most available datasets continue to use the 2011 Census codes for identification of locations.

3.2 Air Pollution

To assess the impact of the Rurban mission on the environmental quality in villages, we utilize annual global satellite-based estimates of fine particulate matter (PM2.5) that are widely used to assess the environmental quality at highly granular levels (Van Donkelaar et al., 2021). Data on surface-level air pollution, i.e., PM2.5 levels, is sourced from the Socioeconomic High-Resolution Urban-Rural Geographic (SHRUG) data portal (Asher et al., 2021), that provides estimated annual data on PM2.5 levels at the village level for the period 1998–2020. These estimates are computed by combining Aerosol Optical Depth (AOD) retrievals from the NASA MODIS, MISR, and SeaWiFS instruments with the GEOS-Chem chemical transport model. The methodology integrates satellite retrievals of aerosol optical depth, chemical transport modeling, and ground-based measurements to enhance seasonal and episodic exposure characterization.

For our analysis, we use two measures of PM2.5- (i) the mean level of PM2.5 in a given geographical unit, and (ii) the maximum level of PM2.5 recorded in a given geographical unit. Since our treatment is assigned at the village level, we calculate the average (or maximum) level of PM2.5 at the village level. We use the raw averages as well as the natural logarithmic of these variables to assess the sensitivity of our empirical results to the choice of functional forms. To address the potential concern that PM2.5 exposure may not be limited to a particular village and may spill over to neighboring villages, we additionally aggregate the aforementioned measures to the GP level (i.e., clusters of contiguous villages), and employ the village-level analysis as a robustness exercise. Finally, to ensure parity with our analysis utilizing satellite nighttime lights imagery as a proxy for economic activity (described in detail in Section 3.3), we restrict our sample to the period 2012-20, which also provides us with a balanced number of pre-and post-treatment years.

3.3 Night-Time Lights

We are additionally interested in evaluating the impact of the Rurban Mission on the level of economic activity in the treated villages. However, credible measures of economic activity at highly disaggregated levels are virtually absent in the context of India, especially spanning our sample period. To overcome this data challenge, we use satellite imagery data on nighttime lights (NL), which have increasingly become popular as proxies for capturing economic activity at highly granular levels when more standard economic data are unavailable (Gibson et al., 2021, Henderson et al., 2012). NL levels have been estimated to be highly significant proxies for population, employment, per capita consumption, and electrification at very local levels (Asher et al., 2021).

We use data from NASA's Visible Infrared Imaging Radiometer Suite Day/Night Band (VIIRS-DNB), which reports NL data for India from 2012 onwards. While a longer time series of NL data for India is recorded under the U.S. Department of Defense's Defense Meteorological Satellite Program- Operational Linescan System (DMSP-OLS), however, it ends in 2014, which precludes our intervention years. Our VIIRS-DNB NL data comes from the SHRUG Data Project, which compiles 15×15-arcsecond NL data disaggregated at the village level for India from 2012–2021. Our primary variable of interest is mean luminosity, defined as the sum of nighttime lights (pixel values) detected within the polygons that constitute a village in a year divided by the total area of the village (number of pixel cells in the polygon). For our primary outcome, we calculate the annual growth rate of mean luminosity, which is defined as the natural log difference in the NL recorded at t and $t+1$. Therefore, each observation corresponds to the year-on-year growth in NL within the village boundaries.

3.4 Public Facilities

We also use village-level data on the availability of a host of public facilities and infrastructure from the Mission Antyodaya (MA) Facilities portal. The MA initiative is an effort of the Government of India that surveyed nearly 650,000 villages across India to conduct gap analyses and provide inputs to participatory plans of village councils (Gram Panchayats or GPs), the lowest level of government in the country. This survey is routinely utilized as a replacement for the village-level directory component of the national-level Census, the last round of which was conducted in 2011. All information from MA surveys is cross-sectional and is available at the village level, which allows us to use the dataset to evaluate the Rurban Mission's effect on the availability of different public facilities. To avoid cherry-picking outcomes, we use the information on *all* recorded public facilities. Additionally, we utilize the village-level directory of the 2011 round of the Census of India publication to control for the baseline levels of public facilities in our sample of villages.

4 Empirical Analysis

4.1 Empirical Strategy

We are interested in identifying the causal impact of the *Rurban* Mission on the environment quality in a cluster of villages. Essentially, we estimate the following functional relation:

$$Y_{c,s,t} = f(Rurban, X_{\text{obs}}, X_{\text{unobs}}) \tag{1}$$

where $Y_{c,s,t}$ is a variable capturing the environment quality in cluster c, sub-district s, and year t; *Rurban* is a binary variable indicating whether the cluster was incorporated into the Rurban program described in Section 2.1; X_{obs} and X_{unobs} are vectors of observed and unobserved covariates of environmental quality, respectively. A naive OLS estimation of the aforementioned relation would arguably be mired in endogeneity issues due to several

concerns. For instance, the fact that villages or clusters of villages that were, in effect, potentially superior in terms of their economic conditions (captured by X_{obs}) were treated under the program can be correlated with their baseline levels of environmental quality, which would lead to a problem of reverse causality. On the other hand, certain unobserved characteristics, such as idiosyncratic weather shocks or productivity differentials (captured by X_{unobs}), could be significant correlates of both treatment status as well as environmental quality levels, which would translate into an omitted variable bias.

Identification Strategy. To alleviate these concerns and obtain estimates that can be more credibly interpreted as causal, we leverage two key features of the implementation protocol under the Rurban mission. *First*, we exploit the staggered rollout of the program across clusters between 2016 and 2018. Across the three years, 300 clusters comprising nearly 4,600 villages and 2,800 Gram Panchayats were covered under the program and received funds in a phased manner (see Section 2.1). *Second*, we make use of the fact that clusters were chosen in a select set of sub-districts, which allows us to have a tighter comparison group for treated villages, i.e., using non-treated villages in selected sub-districts as the comparison group. Using these features in a quasi-experimental design, we employ a generalized difference-in-difference strategy to evaluate the impact of the Rurban Mission on environment quality. That is, our empirical strategy compares the before- and after differences in pollution levels between GPs where Rurban was introduced and the rest of the villages/GPs within the *selected* sub-districts that did not change their treatment status between the two periods.

As a baseline specification, we estimate the following two-way fixed-effects (TWFE) model:

$$Y_{c,s,t} = \alpha_c + \delta_t + \beta Rurban_{c,s,t} + \epsilon_{c,s,t} \tag{2}$$

where $Y_{c,s,t}$ represents the environment quality for GP/cluster c in sub-district s in year t; α_c indicates GP fixed effects; δ_t indicates year fixed effects; and, finally, $Rurban_{c,s,t}$ is an indicator variable for whether, in year t, GP c was covered under the Rurban Mission. We

cluster the standard errors at the GP/cluster level.

Although the TWFE model, similar to equation (2), is the workhorse model for staggered adoption research designs, the recent DiD literature has pointed to its inconsistency in settings where the relatively strong assumptions about the homogeneity of treatment effects are violated (Roth et al., 2023). Therefore, to address concerns about the reliability of the TWFE estimator, we re-estimate equation (2) using the robust estimators introduced in Callaway and Sant'Anna (2021) (CSDiD). By shutting down the 2×2 difference-in-differences comparisons between newly treated and already treated units, the CSDiD estimators deliver consistent estimates even in the presence of heterogeneous treatment effects across time and/or treated units.

For both these estimation protocols, the key identifying assumption is that the pollution trends in treated and non-treated GPs would have followed parallel trends in the absence of the policy. That is, while the identification strategy allows for level differences in the outcomes of the treated and non-treated GPs, it requires that their trends not diverge in the absence of an intervention. We test this assumption by estimating a dynamic event study-like version of equation (2) wherein we interact our treatment variable with year-dummies, the coefficient of which represents the year-on-year impact of the treatment status.

4.2 DiD Results

In Table 1, we present the results from the TWFE (odd-numbered columns) and the robust CSDiD estimators (even-numbered columns) using four outcomes- (i) raw annual mean of PM2.5, (ii) logarithmized annual PM2.5 mean, (iii) annual maximum exposure to PM2.5, and (iv) the log-transformed PM2.5 maximum exposure in a particular GP. The TWFE estimates indicate a modest but significant negative association between treatment under the *Rurban* Mission and PM2.5 exposure. Taken at face value, these results suggest that investments through the program contributed to a shift toward more sustainable develop-

ment, leading to reduced particulate matter emissions. However, as outlined in Section 4.1, TWFE estimates are known to be consistent only under a set of strong assumptions about treatment effect heterogeneity. One particular issue pointed out in the literature is that of negative weights assigned to 2×2 DiD comparisons between groups of units treated at different points in time, which would invariably lead to inconsistent estimates (Baker et al., 2022, Callaway and Sant'Anna, 2021). We employ the decomposition method of Goodman-Bacon (2021) to test whether the problem of heterogeneous treatment effects applies to our setting. Essentially, since each group can serve as a control at some point in time, in the decomposition for the TWFE estimator with staggered treatment, the main effect is a combination of four key comparisons: (i) early implementers versus untreated groups, (ii) late implementers versus untreated groups, (iii) early implementers versus late implementers before treatment, and (iv) early implementers versus late implementers after treatment. We run the Goodman-Bacon (2021) decomposition test for the four outcome variables presented in Table 1.

Figure A1 in the Appendix presents graphical evidence from the Goodman-Bacon (2021) decomposition tests. We find that the problem of heterogeneous treatment effects, as evidenced by the deviation of the differentially-timed estimated effects from the 2×2 DiD estimate, is indeed a concern in our settings, which deems the TWFE estimates inconsistent for our case. To assuage these concerns, we employ the robust estimators proposed by Callaway and Sant'Anna (2021) (CSDiD) that generate consistent estimates in the presence of heterogeneous treatment effects in staggered treatment setups. The even-numbered columns present the CSDiD estimates for our key outcome variables. In contrast to the TWFE estimates, we find that the treatment effect of the *Rurban* Mission, using the CS-DiD procedure, is a tight null and, in some cases, statistically significantly positive but not economically meaningful. Before we attempt to uncover the stark and puzzling difference between the two types of estimators, we address the validity of the key identifying assumption of our setting, as outlined in Section 4.1.

Table 1: Impact of the *Rurban* Mission on Air Pollution: TWFE and CSDiD Estimates

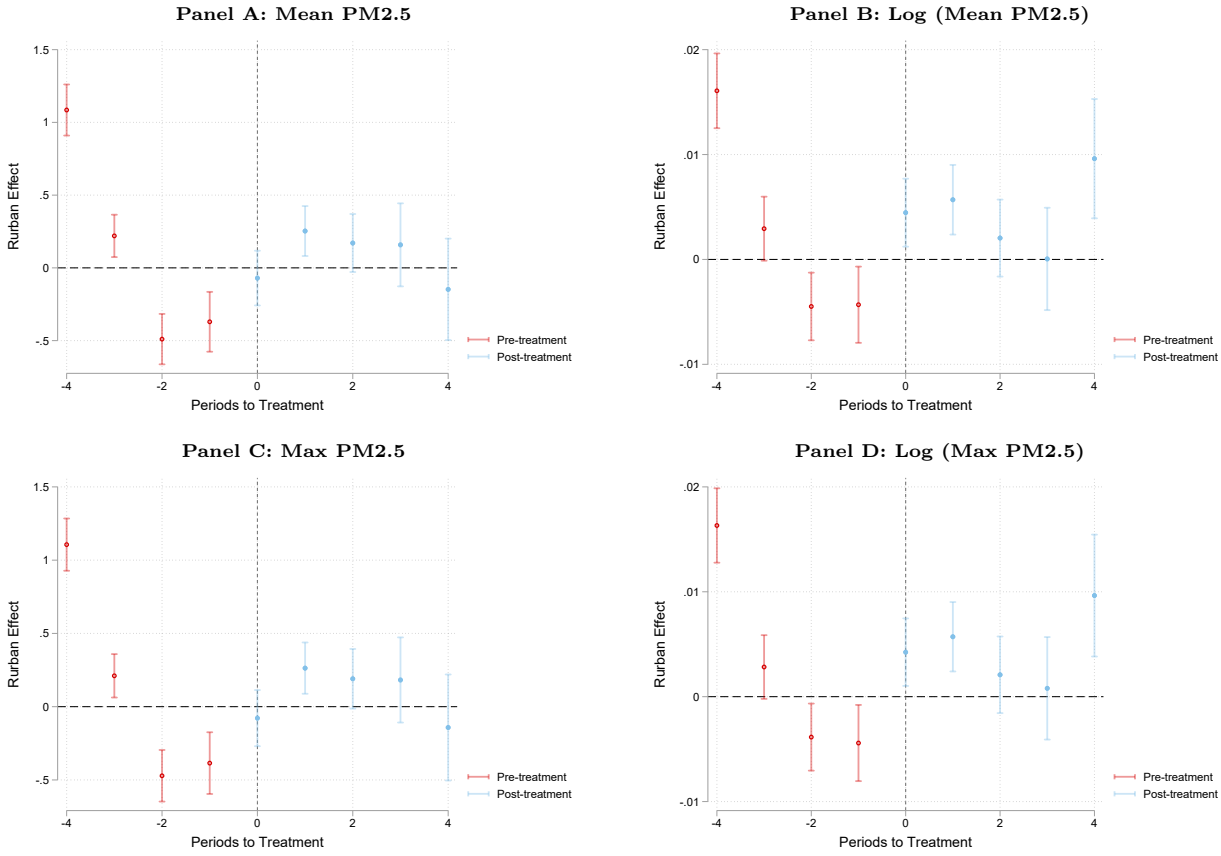
	Air Pollution							
	Mean PM2.5		Log (Mean PM2.5)		Max PM2.5		Log (Max PM2.5)	
	TWFE (1)	CSDiD (2)	TWFE (3)	CSDiD (4)	TWFE (5)	CSDiD (6)	TWFE (7)	CSDiD (8)
Rurban	-.1798***	.1024	-.0015	.0038**	-.1626***	.1124	-.0010	.0039***
	(.0559)	(0.0847)	(.0010)	(.0015)	(.0572)	(.0863)	(0.0010)	(0.0015)
GP fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,737	190,737	190,737	190,737	190,737	190,737	190,737	190,737
Number of GPs	21,193	21,193	21,193	21,193	21,193	21,193	21,193	21,193
Mean Dependent Var.	63.4577	63.4577	4.0605	4.0605	64.7102	64.7102	4.0836	4.0836

Notes: Standard errors clustered at the GP level are presented in parentheses. The dependent variable in Columns (1) and (2) is the annual average of PM2.5 measured within the boundaries of the GP; in columns (3) and (4), the outcome variable is the logarithmic transformation of this annualized average; in columns (5) and (6), the outcome is the maximum PM2.5 annual level detected within the boundaries of the GP; in columns (7) and (8), the outcome is the logarithmic transformation of this annualized maximum level. The odd-numbered columns present estimates from the standard two-way fixed effects (TWFE) procedure with GP and year fixed effects, while in the even-numbered columns, we employ the robust procedure formulated by Callaway and Sant'Anna (2021) to incorporate heterogeneous treatment effects in staggered treatment setups. In all regressions, our sample period runs from 2012 to 2020.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Credibility of the Parallel Trends Condition. One of the crucial assumptions of generalized difference-in-differences is that of parallel trends— that is, in the absence of the Rurban rollout, the environmental outcome trends of the treated and non-treated clusters in different Rurban expansion groups would have continued to evolve at the same rate. We use an event study-like approach within the CSDiD protocol to assess the plausibility of this assumption in our setting. That is, if the Rurban treatment status had significant effects on environmental outcomes between treated and non-treated clusters prior to the program's actual introduction, this would suggest the existence of divergent pre-trends, which, in essence, would violate our parallel trends assumption.

Figure 3 presents the event study plot from the CSDiD estimates for our analysis's four primary outcome variables. We use a window of 4 years on either side of the treatment

Figure 3: Dynamic CSDiD Estimates: PM2.5 Exposure



assignment year to create a balanced plot of pre-and post-treatment effects. Across the panels in Figure 3, we find that there is a statistically significant effect of the treatment status under the Rurban program on different measures of PM2.5 exposure in the pre-treatment period. That is, the estimated year-on-year coefficients for the years prior to the implementation of Rurban in a given cluster are statistically significantly different from zero. This graphical evidence indicates the existence of divergent pre-trends, which invariably violates our key identifying assumption required to generate consistent estimates that can be interpreted as causal. Table A1 in the Appendix documents the pre-treatment and post-treatment estimates separately to shed more light on the potential violation of the parallel trends condition in our setup.

The evidence in favor of the existence of divergent pre-trends is, however, unsurprising in

our case. This is primarily due to the fact that the selection protocol under the Rurban Mission utilized pre-existing 'trends' instead of levels of development variables in villages as an input to the selection procedure (see Section B.1 in the Online Appendix). Nonetheless, even if we discount the existence of pre-trends, the post-treatment quadrant of all the panels in Figure 3 suggests that the Rurban mission either had tight null or, in some cases, modest positive effects on fine particulate matter exposure in villages.[5] Coupled with the TWFE negative estimates in Table 1- albeit inconsistent-, this suggests that the Rurban mission, in effect, did not lead to higher pollution levels in treated villages- a result that is fairly uncommon when evaluating clustered place-based policies (see, for example, Zhao et al. (2022)). However, these estimates cannot be interpreted to be causal and must be viewed as robust evidence with caution. To alleviate the concerns related to the parallel trends condition, we employ the synthetic difference-in-differences (SDiD) technique (Arkhangelsky et al., 2021) to arrive at credible causal estimates in our setting. In the next section, we provide details on this methodology and the results from the SDiD estimation technique.

4.3 Alternative Empirical Strategy: Synthetic Difference-in-Differences

In this section, we employ a synthetic difference-in-differences (SDiD) strategy à la Arkhangelsky et al. (2021) to estimate the effect of the Rurban Mission on fine particulate matter exposure in villages. A key aspect of the SDiD estimation strategy is that, in addition to allowing for pre-treatment differences in the treated and control units, as in the generalized DiD case, it optimally generates a matched control group, similar to the synthetic control (SC) methods. The synthetically matched control group is constructed by computing two sets of weights- the first set of weights aligns pre-exposure trends in the outcome of 'unexposed' units with those for the exposed units, while the second set balances pre-exposure

[5]The fourth year in the post-treatment is imprecisely estimated since our data spans 2012-2020, and only Phase-I clusters (nearly 100 in total) were exposed to the treatment for that long. In the other years, the effects are fairly muted and statistically indistinguishable from zero.

periods with post-exposure ones. Particularly, higher weights are assigned to units that exhibit near-identical pre-treatment trends, thereby generating a treatment and control group that, on average, possess parallel trends.

There are two critical reasons why our setting is ideal for employing the SDiD estimation procedure. First, in our setting, the set of non-treated units is fairly large which increases the precision and accuracy of constructing a synthetic counterfactual within the SDiD procedure. In each post-treatment year, we have nearly 18,934-20,405 control units compared to 763-2,234 treated units, i.e., at all points in the post-intervention phase, our sample of 21,193 GPs consists of nearly 90% control units. Second, the SDiD protocol naturally extends to our staggered treatment assignment setting. In this case, the treatment matrix is broken down into three adoption-date-specific matrices, with each timing-specific matrix comprising simply of pure control units and units that adopt in the specific period. The average treatment effect of the treated (ATT) under the SDiD procedure in staggered settings is then calculated by applying the SDiD estimator to each of the three adoption-specific samples and calculating a weighted average of the three estimators, where weights are assigned based on the relative number of treated units and time periods in each adoption group (see Arkhangelsky et al., 2021, Appendix A).[6]

The estimates from employing the block-bootstrap procedure under the SDiD protocol are presented in Table 2. We find that after correcting for the divergent pre-trends, there is a negative effect of being treated under the Rurban Mission on PM2.5 exposure, across different outcome specifications. In particular, we find that Rurban reduces a treated GP's average fine particulate matter exposure by nearly 0.14 $\mu g/m^3$ (see Column (1)). Relative to the sample mean, this corresponds to a 0.22% reduction in the average PM2.5 exposure. This is corroborated by the estimate for the alternative outcome specification using the logarithmized average PM2.5 exposure, wherein we find that there is a 0.3 percentage

[6]The algorithmic procedure and the STATA package utilized to execute the SDiD estimation procedure within a staggered setting is outlined in detail in Clarke et al. (2024).

Table 2: Impact of the *Rurban* Mission on Air Pollution: SDiD Estimates

	Air Pollution			
	Mean PM2.5	Log (Mean PM2.5)	Max PM2.5	Log (Max PM2.5)
	(1)	(2)	(3)	(4)
Rurban	-.1419***	-.0039***	-.1569***	-.0041***
	(.0277)	(.0009)	(.0444)	(.0012)
GP fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
SDiD Procedure		Block Bootstrap		
Observations	190,737	190,737	190,737	190,737
Number of GPs	21,193	21,193	21,193	21,193
Mean Dependent Var.	63.4577	4.0605	64.7102	4.0836

Notes: Standard errors clustered at the GP-level are presented in parentheses. The dependent variable in Column (1) is the annual average of PM2.5 measured within the boundaries of the GP; in Column (2), the outcome variable is the logarithmic transformation of this annualized average; in Column (3), the outcome is the maximum PM2.5 annual level detected within the boundaries of the GP; in column (4), the outcome is the logarithmic transformation of this annualized maximum level. In all columns, we employ a synthetic difference-in-differences estimation technique à la Arkhangelsky et al. (2021) and present estimates obtained from employing the block bootstrap procedure by running 500 bootstrap replications. In all regressions, our sample period runs from 2012 to 2020.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

point reduction in the average PM2.5 exposure ($\sim$ 1% reduction relative to the mean). In columns (3) and (4), we find similar negative effects on the maximum annual exposure to fine particulate matter detected within the boundaries of the GP.

To situate our estimates within the broader literature on changes in PM2.5 exposure, we benchmark the magnitude of our estimates to existing studies. In general, movements in PM2.5 are observed to be relatively sticky. For instance, Borgschulte et al. (2024) find that an additional day of wildfire smoke- a primary cause of fine particulate matter exposure- leads to an increase in the quarterly average levels of PM2.5 by 0.06 $\mu g/m^3$ at the county level. Analyzing a special economic zone policy in China, Zhao et al. (2022) find that treated cities witnessed an increase of 0.1 percentage point increase in the average annual PM2.5 exposure. Compared to these effects, our SDiD estimates align closely with the average estimates reported in the literature— a noteworthy finding given that we examine

a light-touch intervention. We examine the robustness of these estimates in Section 7.

5 Economic Activity: *Nighttime Lights*

How does the *Rurban* Mission impact economic activity in treated clusters? Recall from Section 2.1 that the core objective of the policy was to leverage the favorable economic conditions of select village clusters to drive broader regional economic transformation. While the previous section suggests that the policy had either positive or neutral effects on air quality, a key question remains, that is, whether this mitigation comes at the cost of economic growth.

As noted in Section 3.3, disaggregated economic data at the village/GP level are absent during our sample period. To address the unavailability of data on economic activity at granular levels, we utilize satellite imagery data on nighttime lights (see Section 3.3 for more details on the data). For our sample period from 2012-21, we use four outcomes for our analysis vis-à-vis economic activity: (i) logarithmic annual nighttime lights weighted by the number of pixel cells of the village polygon, (ii) logarithmic annual maximum of nighttime lights weighted by the number of pixel cells of the village polygon, (iii) growth in the average annual nighttime lights, and (iv) growth in the annual maximum luminosity.[7] All the variables are aggregated up to the GP level, similar to the analysis of pollution exposure. We continue to employ the SDiD procedure for estimating causal estimates of the *Rurban* Mission on nighttime lights.

Table 3 presents the results from this estimation exercise. Across the four outcome specifications, we find that the *Rurban* Mission had a positive impact on the level and growth of nighttime lights in treated clusters, which is both statistically significant and economically meaningful. Our preferred outcome specification is presented in Columns (1) and

[7]The growth variables are calculated as the difference in the natural logarithm of the variable at t and $t+1$. Hence, for our sample period of 2012-21, we have nine observations per GP (excluding the base year, 2012).

Table 3: Impact of the *Rurban* Mission on Economic Activity: SDiD Estimates

	Nighttime Lights (NL)			
	Log (Mean NL)	Growth in Mean NL	Log (Max NL)	Growth in Max NL
	(1)	(2)	(3)	(4)
Rurban	.0154***	1.5993***	-.0023	.3005***
	(.0024)	(.1490)	(.0016)	(.0834)
GP fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
SDiD Procedure		Block Bootstrap		
Observations	211,930	190,737	211,930	190,737
Number of GPs	21,193	21,193	21,193	21,193
Mean Dependent Var.	.5659	2.6065	.1288	.3705

Notes: Standard errors clustered at the GP-level are presented in parentheses. The dependent variable in Column (1) is the logarithmic transformation of the annual sum of nighttime lights divided by the number of pixel cells within the boundaries of the GP; in Column (2), the outcome variable is the year-on-year difference of the outcome in column (1), multiplied by 100; in Column (3), the outcome is annual maximum nighttime lights detected within the GP, divided by the number of pixel cells within the boundaries of the GP; in column (4), the outcome variable is the year-on-year difference of the outcome in column (3), multiplied by 100. In all columns, we employ a synthetic difference-in-differences estimation technique à la Arkhangelsky et al. (2021) and present estimates obtained from employing the block bootstrap procedure by running 500 bootstrap replications. In all regressions, our sample period runs from 2012 to 2021.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(2), where we use different versions of the annual average nighttime lights detected within the GP boundaries. We find that relative to the sample mean, treated GPs witness a 2.7% higher average nighttime lights compared to non-treated GPs. This effect is more pronounced when we consider the growth rate in the average luminosity in Column (2). In particular, we find that Rurban clusters witness nearly 1.6 percentage points higher NL growth compared with non-treated clusters within the same sub-districts. Relative to the mean, this corresponds to a massive 61.3% higher luminosity growth in treated regions. The effect is similar in the case of maximum luminosity weighted by the size of the GP; however, only the growth variable remains statistically significant at the conventional significance levels.[8]

[8]While the maximum PM2.5 exposure served an intuitive outcome to measure the intensity of pollution, the case for using maximum annual luminosity, however, is not straightforward. We nonetheless use it as a robustness exercise, following Burlig and Preonas (2024).

To attribute economic meaning to these estimates, we utilize estimates of the elasticity of GDP to nighttime lights to compute the effect in terms of actual regional GDP growth (see Bickenbach et al., 2016, Baskaran et al., 2023). Using an estimated NL-to-GDP elasticity of 0.10-0.15, our estimates translate to a modest 0.15-0.23 percentage points increase in the regional GDP growth of a treated GP. Furthermore, to calculate the growth premium of the Rurban Mission, we utilize the national GDP growth of India, which averaged around 5.5% during our sample period. Our estimates suggest that the growth premium associated with the *Rurban* Mission was between 2.7 and 4.1%.[9] In summary, we find that the cluster-based development under the Rurban program that sought to preserve the sustainability of villages' economic dynamics led to an invariable increase in the economic activity of the region, as measured by nighttime luminosity. Moreover, and more importantly, these economic gains were not achieved at the cost of environmental disruption and degradation. Unlike studies on cluster-based development programs that highlight a zero-sum tradeoff between economic growth and environmental quality, our empirical analysis finds that the *Rurban* Mission effectively reconciled this tradeoff, achieving positive effects on both objectives simultaneously. This suggests that a light-touch intervention that leverages the existing resource base and governance structure to facilitate regional development may potentially serve as a viable and sustainable alternative to more disruptive area development programs. In the next section, we explore the underlying channels through which the effects documented in this section, along with those in Table 2, operationalize.

6 Potential Mechanisms

In this section, we explore the potential underlying mechanisms through which the effects on pollution and economic growth documented in Sections 4.3 and 5 materialize. Our

[9]This is calculated as follows. We calculate the proportion of the average annual GDP growth rate explained by our estimates, i.e., estimated GDP growth as a result of the treatment status under the Rurban mission (0.11–0.32%) divided by the average annual GDP growth rate in the country (5.5%). The interpretation of this 'premium' is the proportion of the growth rate explained by our intervention.

focus is on three major aspects- (i) public facilities in villages (Section 6.1), (ii) types of clusters (tribal v/s non-tribal) (Section 6.2), and (iii) types of projects undertaken under the program (Section 6.3).

6.1 Public Facilities

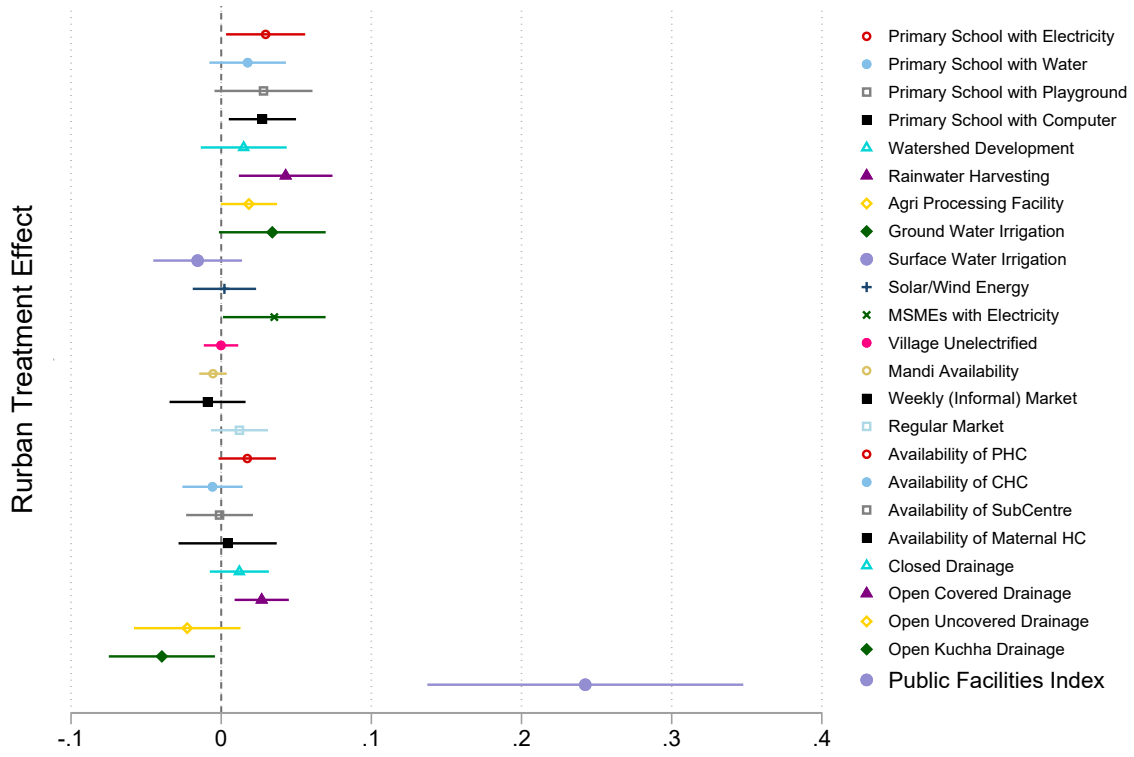
To understand how the Rurban Mission leads to economic gains while minimizing environmental degradation, we analyze the association between treatment under the program and the availability of public facilities within the village. As noted in Section 3.4, we use the Mission Antyodaya (MA) village-level survey on amenities and facilities available in a village. This survey records information in 2020, nearly four years after the Phase-I rollout of the Rurban Mission. To estimate the relationship between *Rurban* and the availability of village-level facilities, our cross-sectional specification is as follows:

$$Y_{v,s,MA} = \alpha + \beta Rurban_{v,s} + \gamma \mathbf{X}_{v,s,2011} + \delta_s + \epsilon_{v,s,MA} \tag{3}$$

In equation (3), $Y_{v,s,MA}$ is a binary variable indicating the availability of a particular amenity or public facility in village v, sub-district s in the Mission Antyodaya survey-year MA. *Rurban* is an indicator equalling 1 if village v was treated in one of the three phases of the Rurban Mission. $\mathbf{X}_{v,s,2011}$ is a vector of controls or baseline characteristics recorded for the village in the 2011 round of the Census of India. The set of controls varies for every outcome variable- for instance, if the outcome variable corresponds to drainage infrastructure, then the controls would include information on the baseline level of drainage systems in the particular village during 2011. We also include a common set of controls in all specifications, which involve the population of the village, the Scheduled Caste (SC) population in the village, the share of non-farm employment in the village, and the phase in which the particular village was treated under the Rurban program, if at all. δ_s represents sub-district fixed effects to control for sub-district-specific heterogeneity. Note that we conduct this analysis at the village level, contrary to our primary analyses in Sections

4.3 and 5, to preserve relevant variations in the extensive margin outcome variables for our cross-sectional sample. Finally, our standard errors are clustered at the GP level.

Figure 4: Public Facilities: Mission Antyodaya (2020)



Notes: The figure plots coefficients and 90% confidence intervals from estimating equation (3) using different binary indicators of public amenities as the outcomes. The data comes from the Mission Antyodaya survey, which records information on village-level facilities in 2020.

In Figure 4, we plot the coefficients from our estimation exercise with the availability of different public facilities as outcomes. After controlling for a host of baseline characteristics, we find that overall, Rurban villages have better availability of public facilities in 2020 compared to non-treated villages. We discuss a few of these which allow us to provide suggestive evidence in favor of the findings in preceding sections. We find that there is a move towards more sustainable agricultural practices- as indicated by a higher prevalence of watershed management services and rainwater harvesting, and a lower reliance on surface water irrigation. Moreover, the increased availability of farm-gate processing facilities

potentially reduces the need to transport large quantities of agricultural produce across villages for processing and marketing. Further, we find that Rurban villages possess an improved drainage system compared to non-treated villages, indicated by a higher availability of covered drainage systems and a lower reliance on uncovered or *kuchha* (unlined) drains. Together, these factors may help explain the modest improvement or stagnation in air quality observed in Section 4.3. While we do not observe significant differences in market infrastructure—such as *mandis*, regular markets, or weekly *haats*—we find a higher availability of electricity supply to Micro, Small, and Medium Enterprises (MSMEs). Coupled with broader improvements in health infrastructure and drainage, this suggests an overall enhancement in economic conditions, which likely contributes to the observed increase in economic activity within these villages (as shown in Table 3). Finally, we additionally construct an index of public facilities using Principal Component Analysis with information on more than 20 facilities, including the outcomes discussed above. We find a positive effect of being treated under the Rurban Mission on the overall supply of public goods- 0.12 standard deviations higher availability of public goods in treated villages.

While these findings suggest positive associations between the Rurban Mission and improved public amenities, we emphasize that this analysis provides suggestive rather than causal evidence. The cross-sectional OLS results cannot fully account for potential confounders that may drive these results. Nonetheless, even if not strictly causal, the observed improvements in public facilities, sustainable agricultural practices, and economic conditions provide us with a potential underlying channel for our main effects on pollution and nighttime lights.

6.2 Heterogeneity by Types of Clusters: Tribal v/s Non-Tribal

Can the type of cluster help explain the findings observed in the preceding sections? Under the *Rurban* Mission, two types of village clusters were treated- *tribal* and *non-tribal* (see Section B.1 in the Online Appendix). Tribal clusters, selected within sub-districts that

were predominantly tribal in terms of their demographic composition, comprised villages that additionally provided slightly higher support for environmental projects within the ambit of the program. In India, tribal regions possess institutional and governance setups that are centered around environmental conservation and preserving self-reliant resource bases. As a potential channel, we examine whether tribal clusters were able to leverage their existing institutional features better to facilitate the improvement in environmental quality while fostering economic gains in the respective villages vis-à-vis non-tribal clusters.

To answer this question, we estimate a different version of equation (2) and conduct a sub-sample analysis. In particular, we employ our preferred two SDiD estimation procedure for two sub-samples of village clusters: (i) comparing *treated* tribal clusters (GPs) with *non-treated* village clusters/GPs in tribal sub-districts, and (ii) comparing *treated* non-tribal clusters (GPs) with *non-treated* village clusters/GPs in non-tribal sub-districts.[10] The results from this sub-sample analysis are presented in Table 4.

The sub-sample analysis reveals that it is indeed the tribal village clusters that witness a larger decline in fine particulate matter exposure following the intervention. While the direction of the estimates is identical for the sub-sample of non-tribal clusters (Panel B), the magnitude is nearly five times smaller than the estimates for tribal clusters. This reveals a complementarity of existing institutions and resource base compositions within the village clusters in generating the observed effects. In Table A2 in the Appendix, we present similar results from a sub-sample analysis using nighttime lights growth as the outcome. We find that the estimates across the two sub-samples are largely similar and in some cases, are mildly higher for non-tribal village clusters. This suggests that tribal clusters are able to produce similar economic gains while inducing a much larger improvement in their environment quality, possibly due to the pre-existing institutional apparatus and conditions.

[10] In the sub-samples, our treated GPs are officially identified under the Rurban Mission as "tribal" clusters. Our comparison group in this sub-sample analysis comprises GPs that were not treated within the sub-districts that were identified by the Ministry of Rural Development as 'tribal' (see Online Appendix Section B.1).

Table 4: Impact of the *Rurban* Mission on Air Pollution: SDiD Estimates

	Air Pollution			
	Mean PM2.5	Log (Mean PM2.5)	Max PM2.5	Log (Max PM2.5)
	(1)	(2)	(3)	(4)
Panel A: Tribal Clusters				
Rurban	-.6701***	-.0098***	-.6922***	-.0102***
	(.0602)	(.0012)	(.0744)	(.0012)
GP fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	87,273	87,273	87,273	87,273
Number of GPs	9,697	9,697	9,697	9,697
Mean Dependent Var.	64.2381	4.0697	65.4434	4.0933
Panel B: Non-Tribal Clusters				
Rurban	-.1361*	-.0028**	-.1421***	-.0029**
	(.0740)	(.0011)	(.0505)	(.0013)
GP fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	98,721	98,721	98,721	98,721
Number of GPs	10,969	10,969	10,969	10,969
Mean Dependent Var.	62.2366	4.0444	63.5134	4.0668

Notes: Standard errors clustered at the GP-level are presented in parentheses. Panel (A) presents the estimates for the sub-sample for the tribal clusters, and Panel (B) presents estimates for the non-tribal sub-sample (see Footnote 10). The dependent variable in Column (1) is the annual average of PM2.5 measured within the boundaries of the GP; in Column (2), the outcome variable is the logarithmic transformation of this annualized average; in Column (3), the outcome is the maximum PM2.5 annual level detected within the boundaries of the GP; in column (4), the outcome is the logarithmic transformation of this annualized maximum level. In all columns, we employ a synthetic difference-in-differences estimation technique à la Arkhangelsky et al. (2021) and present estimates obtained from employing the block bootstrap procedure by running 500 bootstrap replications. In all regressions, our sample period runs from 2012 to 2020.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.3 Projects Undertaken

Do the types of projects matter? As discussed earlier, the *Rurban* Mission leads to an improvement in air quality while simultaneously fostering economic gains. We note that this finding is somewhat puzzling, as a vast body of literature suggests a clear, almost

zero-sum, trade-off between economic development and environmental degradation (Holtz-Eakin and Selden, 1995, Nordhaus, 1977, Yu et al., 2023). Ideally, the first-best for any place-based policy is to foster economic development without negatively impacting the environment. To understand the underlying channel for our somewhat puzzling estimates, we examine the project composition under the program.

To answer this question, we leverage administrative data from the official web-based application launched by the Ministry of Rural Development under the Rurban Mission that tracks project approvals (including the type of project), implementation, and completion. Our descriptive analysis in this section is based on two sets of documents: *first*, documents outlining the components and sub-components that Rurban projects *should* encompass, and *second*, the list of projects actually approved under the program up to December 2024. The use of these documents allows us to examine the intended allocation of resources (ex-ante) versus the actual implementation (ex-post). To carry out this text-based analysis, we adopt a rule-based classification approach, which employs a set of conditional statements (rules) to categorize data points. In our case, projects are bifurcated based on a predefined set of keywords, and each project is categorized as into three categories- (i) *Environmentally Friendly*, (ii) *Less Environmentally Friendly*, or (iii) *Neutral*, based on the presence of these keywords in the project name.[11]

The 'Environmentally Friendly' category included keywords related to renewable energy (solar, wind, renewable, bio-energy), water conservation (rainwater harvesting, watershed management, groundwater recharge), sustainable infrastructure, waste management (recycling, waste-to-energy), and afforestation (tree-planting, etc.). In contrast, the 'Less Environmentally Friendly' category comprised keywords associated with heavy construction and

[11]This classification is based on Natural Language Processing (NLP) techniques, which are widely used in sustainability studies and policy analysis (Chang et al., 2021, Zhong et al., 2021). First, we define keywords based on the literature and policy guidelines (training dataset), developing a comprehensive keyword list informed by policy documents such as India's National Action Plan on Climate Change (See the Action Plan here), and UN Sustainable Development Goals - SDGs (see the list here). These keywords were categorized into three groups: Environmentally Friendly (Positive Contribution to Sustainability), Less Environmentally Friendly (High Environmental Impact), and Neutral (Mixed or Context-Dependent Impact).

infrastructure (factory, mining, thermal, oil) and fossil fuel dependency (coal, petroleum, gas power), among others. The challenging aspect was arriving at a set of keywords that could be classified as 'Neutral'. Keywords in this category included general infrastructure projects such as water pipelines, schools, healthcare facilities, and smart boards, which arguably have a net zero or negligible impact on air quality. Before we present the descriptive evidence from this analysis, we present one example of categorization in Table 5 that illustrates the classification of projects based on the keyword library.

Table 5: Classification of Projects by Impact

Project	**Keywords**	**Classification**
Solar-Powered Water Pump Installation	solar, water	Positive
Coal-based Power Plant Expansion	coal, power plant	Negative
Smart Board Installation in School	school, installation	Neutral

The first set of documents (intended projects) outlines 347 sub-project components, of which, based on our classification algorithm, 38 (11%) are 'environmentally friendly' projects, 79 (32%) are 'environmentally less friendly', and 230 (57%) are 'neutral'. The ex-ante project compositions indicate the prevalence of environmental sustainability as a key input to the agenda, even at the initial planning stages. The second set of documents, which lists the projects ultimately approved under the program, allows us to assess whether environmentally sustainable projects were prioritized during implementation. The classification analysis of the approved projects reveals that 6,164 projects were 'environmentally friendly' (10-26%, depending on the denominator chosen).

Within the environmentally friendly category, the type of projects undertaken were as follows: 2,346 projects focused on renewable energy (primarily solar-based initiatives), 522 projects were related to afforestation and ecological restoration (plantation, reforestation), 306 projects addressed water conservation and recharge (rainwater harvesting, groundwater recharge), and 2,593 projects fell under other light-touch environmentally friendly initia-

tives. In Figure A2 in the Appendix, we present descriptive evidence supporting the claim that environmental quality was a non-negligible aspect of the projects by illustrating (i) the distribution of the common words used in project descriptions and (ii) distribution of projects approved under the program. In summary, this descriptive analysis offers suggestive evidence that environmental concerns were embedded within the program's framework. It is not that villages naturally generated economic gains sustainably, but rather that a concerted environmental outlook of the policy support was necessary to maintain their environmental conditions.

7 Robustness Exercises

In this section, we examine the robustness of our findings from Section 4.3. In particular, we examine whether (*i*) the results are sensitive to the presence of outliers or extreme observations (Section 7.1), (*ii*) the results are sensitive to the inclusion of certain states (Section 7.2), and (*iii*) the results are robust to the choice of the comparison group.

7.1 Winsorization

Here, we test the sensitivity of our estimates to the presence of outliers. To address this concern, we winsorize our sample by assigning a lower weight to extreme observations on both sides of the tail of the distribution. This allows us to minimize the effect of outliers on our estimates and to check whether our results are driven by the presence of these outliers in our data. For this, we estimate four types of specifications using the staggered SDiD estimation procedure wherein we assign lower weight to (i) the extreme 5% values, (ii) the extreme 10% values, (iii) the extreme 500 values, and (iv) the extreme 1000 values at both tails. The results from this exercise are contained in Table A3.

Across all winsorized samples, the estimates remain largely stable and consistent with the

original estimates presented in Table 2, regardless of the proportion of extreme values winsorized in the sample. Moreover, as the degree of winsorization increases, the precision of the estimates improves significantly. This strengthens our confidence in our empirical design showcasing that the presence of outliers in the sample is unlikely to influence the main findings.

7.2 State-Specific Sensitivity

Given that our analysis covers a large number of states that arguably possess significant heterogeneity in their respective capabilities and resource base to operationalize the projects under the program, we assess whether the inclusion of any particular state drives our results. The effectiveness of policy implementation can vary across states and over time—for instance, northern states such as Bihar and Uttar Pradesh have historically remained laggards in terms of their institutional apparatuses and, therefore, might be unable to leverage the Rurban mission to generate the gains observed in the preceding sections. To rule out the possibility that such differences drive our findings, we conduct a robustness check where we drop one state at a time and estimate a set of equations (with each state excluded once) using the SDiD procedure.

In Figure A3 in the Appendix, we present graphical evidence for this leave-one-out analysis, dropping one state at a time from our estimation sample and re-estimating the baseline staggered SDID effect for air pollution. This exercise is conducted for various measures of PM2.5, similar to the outcomes in Table 2. The vertical axis of the figure represents the dropped state. Our results remain largely stable across all PM2.5 measures, except in some cases where the estimate becomes statistically insignificant but mainly negative.

8 Conclusion

Sustainable urbanization has remained a major challenge for development policymakers. While urbanization leads to growth, development, and associated impacts, it also imposes a negative externality by way of higher pollution, resulting in market failures. This paper explores a unique attempt at internalizing this externality leveraging the idea of lower abatement costs through implementing place-based policies and a cluster urbanization approach.

The basic premise of this paper suggests that place-based policies can potentially retain the core benefits of urbanization while avoiding the potential pitfalls of market failure that standard models of urbanization cause. Consider a partial equilibrium setting for an emissions market with upward-sloping marginal external costs and a downward-sloping marginal abatement cost. The way standard urbanization policies affect development would imply increased emissions, through a potential rightward shift of the marginal abatement costs, i.e., a higher cost of abatement per unit of emission due to increased concentration and population density arising out of urban development. This results in a higher emissions in the resulting equilibrium.

In contrast, the Rurban model with explicit focus on sustainability and preserving rural identity implies one would avoid the higher growth in density of population and therefore a leftward shift of the abatement cost curve, i.e., lower costs of abatement per unit of pollution. The resultant equilibrium emissions are, therefore, lower under such a place-based approach.

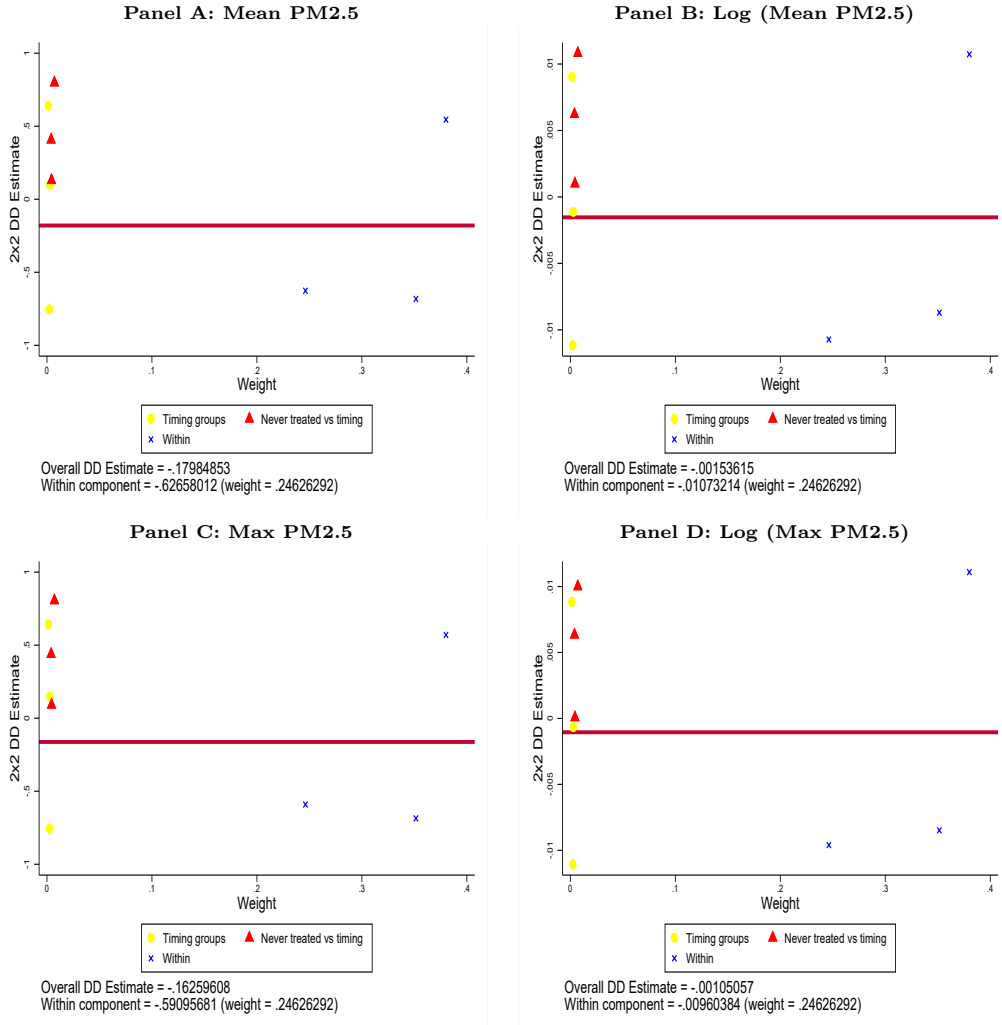
Our paper essentially provides reduced form empirical support to such a theoretical argument and recommends the adoption of policies that contain the advantages of urbanization but avoid the pitfalls of the negative externality to avoid potential market failure.

Appendix (for publication)

A Additional Tables and Figures

A.1 Decomposition Test à la Goodman-Bacon (2021)

Figure A1: Goodman-Bacon Decomposition



A.2 Dynamic CSDiD Estimates: Air Pollution

Table A1: Dynamic CSDiD Estimates

	CSDiD Estimate	Pre-Average	Post-Average	Observations	Number of GPs
	(1)	(2)	(3)	(4)	(5)
Mean PM2.5	.1024	.1111***	.0728	190,737	21,193
	(.0847)	(.0308)	(.0898)		
Log (Mean PM2.5)	.0038**	.0025***	.0043***	190,737	21,193
	(.0015)	(.0005)	(.0015)		
Max PM2.5	.1124	.1150***	.0831	190,737	21,193
	(.0863)	(.0314)	(.0917)		
Log (Max PM2.5)	.0039***	.0027***	.0044***	190,737	21,193
	(.0015)	(.0005)	(.0015)		

Notes: The table presents dynamic estimates from the robust procedure by Callaway and Sant'Anna (2021). Each row is a separate outcome variable measuring pollution (PM2.5) exposure at the Gram Panchayat (GP) level. In Column (1), we reproduce the robust CSDiD estimates from the even-numbered columns of Table 1. Columns (2) and (3) are the pooled treatment effects for the period of pre-treatment years (*Pre-Average*) and the post-treatment years (*Post-Average*) separately. We use a treatment-period window of $(-4, +4)$, i.e., 4 years prior to and post the policy introduction in a particular cluster, to calculate the dynamic estimates under the CSDiD procedure.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.3 Heterogeneity in Nighttime Lights Growth: Tribal v/s Non-Tribal Clusters

Table A2: Impact of the *Rurban* Mission on Economic Activity: Tribal v/s Non-Tribal Clusters

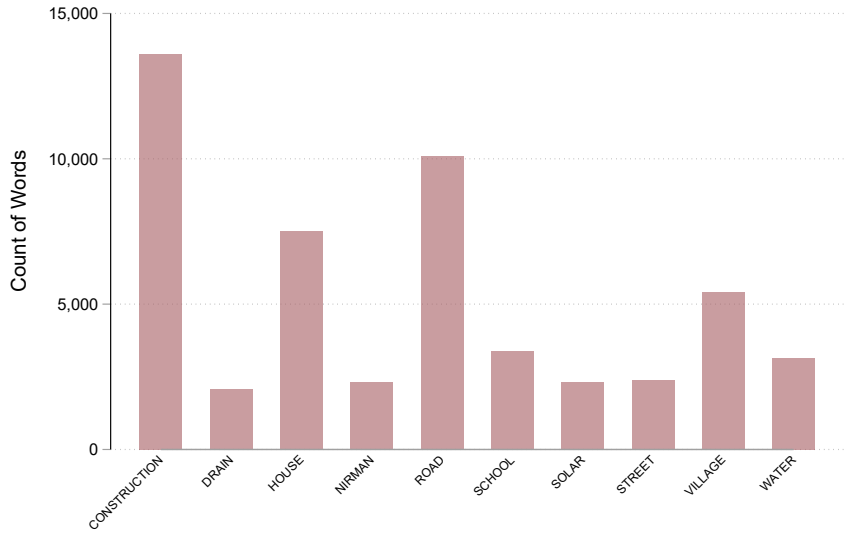
	Nighttime Lights (NL)			
	Tribal		Non-Tribal	
	Growth in Mean NL (1)	Growth in Max NL (2)	Growth in Mean NL (3)	Growth in Max NL (4)
Rurban	1.2428***	1.0092***	.1748	.1938***
	(.3460)	(.0892)	(.1184)	(.0574)
GP fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
SDiD Procedure	Block Bootstrap			
Observations	87,273	87,273	98,721	98,721
Number of GPs	9,697	9,697	10,969	10,969
Mean Dependent Var.	2.5840	.4504	2.5408	.2556

Notes: Standard errors clustered at the GP-level are presented in parentheses. In Columns (1) and (3), the outcome variable is the year-on-year log difference in mean nightlights, multiplied by 100; in Columns (2) and (4), the outcome is the year-on-year log difference in the maximum nightlights weighted by the number of pixels, multiplied by 100. Columns (1)-(2) correspond to the sub-sample of tribal clusters; Columns (3)-(4) present the estimates for the sub-sample of non-tribal clusters. In all columns, we employ a synthetic difference-in-differences estimation technique à la Arkhangelsky et al. (2021) and present estimates obtained from employing the block bootstrap procedure by running 500 bootstrap replications. In all regressions, our sample period runs from 2012 to 2021.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

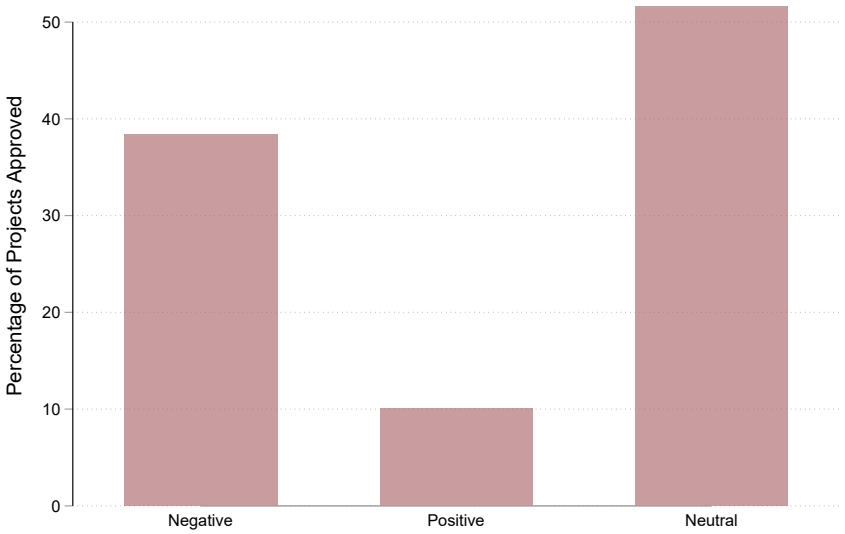
A.4 Project Composition under the Rurban Mission

Figure A2: Project Composition: Common Words and Project Classifications

Panel A: Frequency of Words



Panel B: Rule-Based Classification



Notes: In Panel (A), the figure shows the distribution of words with the maximum frequency in the 'approved projects' documents released by the Ministry of Rural Development. Auxiliary verbs are manually omitted from the list of words to arrive at the final list of words. In Panel (B), the figure shows the proportion of project types using the rule-based classification approach. Here, positive projects correspond to 'environmentally friendly' projects, negative projects refer to those that are relatively less sustainable, and neutral (or uncategorized) projects are those that cannot be easily classified under the aforementioned categories.

A.5 Winsorized Outcomes

Table A3: Impact of the *Rurban* Mission on Air Pollution: Winsorized Outcomes

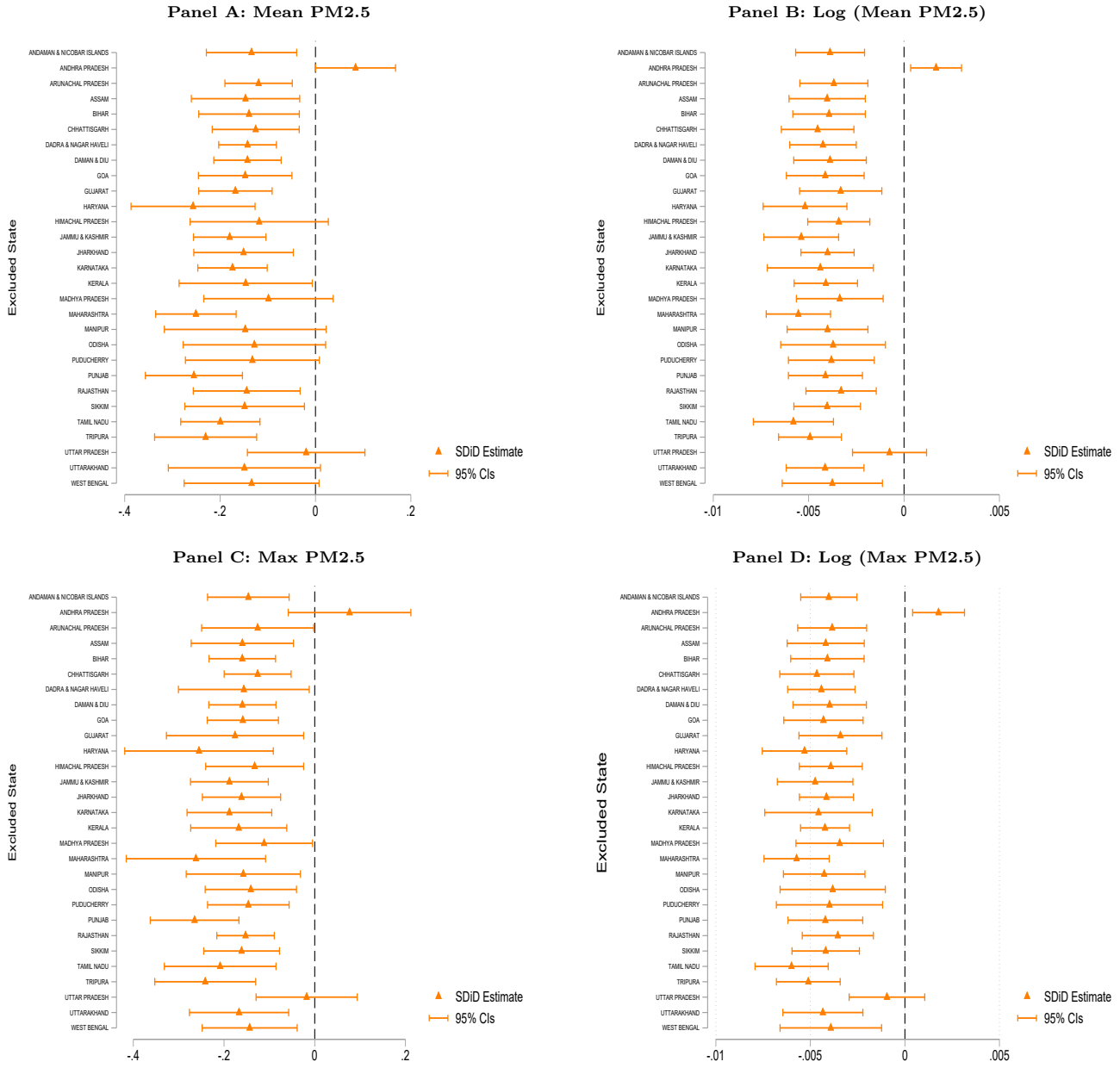
	Air Pollution			
	Mean PM2.5	Log (Mean PM2.5)	Max PM2.5	Log (Max PM2.5)
	(1)	(2)	(3)	(4)
Extreme 5%	-.2789***	-.0044***	-.2896***	-.0046***
	(.0614)	(.0011)	(.0496)	(.0006)
Extreme 10%	-.3382***	-.0049***	-.3548***	-.0048***
	(.0749)	(.0009)	(.0909)	(.0009)
Extreme 500 obs.	-.1535**	-.0043***	-.1645***	-.0043***
	(.0631)	(.0009)	(.0492)	(.0006)
Extreme 1,000 obs.	-.1676***	-.0045***	-.1840***	-.0045***
	(.0504)	(.0014)	(.0594)	(.0014)
GP fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
SDiD Procedure		Block Bootstrap		
Observations	190,737	190,737	190,737	190,737
Number of GPs	21,193	21,193	21,193	21,193
Mean Dependent Var.	63.4577	4.0605	64.7102	4.0836

Notes: Standard errors clustered at the GP-level are presented in parentheses. The dependent variable in Column (1) is the annual average of PM2.5 measured within the boundaries of the GP; in Column (2), the outcome variable is the logarithmic transformation of this annualized average; in Column (3), the outcome is the maximum PM2.5 annual level detected within the boundaries of the GP; in column (4), the outcome is the logarithmic transformation of this annualized maximum level. In all columns, we employ a synthetic difference-in-differences estimation technique à la Arkhangelsky et al. (2021) and present estimates obtained from employing the block bootstrap procedure by running 500 bootstrap replications. Each row represents specifications from a particular winsorized sample, as outlined in Section 7.1. In all regressions, our sample period runs from 2012 to 2020.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.6 Leave-one-state-out Analyis

Figure A3: Robustness Check: State-Specific Sensitivity



References

Amar, S., Ariusni, A., and Satrianto, A. (2018). Pull factors affecting the jobseekers decision to do urbanization. In *First Padang International Conference on Economics Education, Economics, Business and Management, Accounting and Entrepreneurship (PICEEBA 2018)*, pages 154–161. Atlantis Press.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in india using the shrug open data platform. *The World Bank Economic Review*, 35(4):845–871.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

Bardhan, P. and Mookherjee, D. (2006). *Decentralization and Local Governance in Developing Countries: A Comparative Perspective*. MIT Press.

Baskaran, T., Bhalotra, S., Min, B., and Uppal, Y. (2023). Women Legislators and Economic Performance. *Journal of Economic Growth*, pages 1–64.

Baskaran, T., Bhalotra, S., Min, B., and Uppal, Y. (2024). Women legislators and economic performance. *Journal of Economic Growth*, 29(2):151–214.

Bickenbach, F., Bode, E., Nunnenkamp, P., and Söder, M. (2016). Night lights and regional GDP. *Review of World Economics*, 152:425–447.

Borgschulte, M., Molitor, D., and Zou, E. Y. (2024). Air pollution and the labor market: Evidence from wildfire smoke. *Review of Economics and Statistics*, 106(6):1558–1575.

Brückner, M. (2012). Economic growth, size of the agricultural sector, and urbanization in africa. *Journal of Urban Economics*, 71(1):26–36.

Burlig, F. and Preonas, L. (2024). Out of the darkness and into the light? development effects of rural electrification. *Journal of Political Economy*, 132(9):2937–2971.

Busso, M., Gregory, J., and Kline, P. (2013). Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, 103(2):897–947.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.

Chang, I.-C., Yu, T.-K., Chang, Y.-J., and Yu, T.-Y. (2021). Applying text mining, clustering analysis, and latent dirichlet allocation techniques for topic classification of environmental education journals. *Sustainability*, 13(19):10856.

Chaurey, R. (2017). Location-based tax incentives: Evidence from india. *Journal of Public Economics*, 156:101–120.

Chen, J., Li, W., and Zhao, F. (2024). The impact of development zone upgrading on carbon emission reduction: Examining china's regional development policies. *Environment, Development and Sustainability*, 26:1–22.

Clarke, D., Pailañir, D., Athey, S., and Imbens, G. (2024). On synthetic difference-in-differences and related estimation methods in stata. *The Stata Journal*, 24(4):557–598.

Ehrlich, M. v. and Seidel, T. (2018). The persistent effects of place-based policy: Evidence from the west-german zonenrandgebiet. *American Economic Journal: Economic Policy*, 10(4):344–374.

Finney, M. M. and Kohlhase, J. E. (2008). The effect of urbanization on labor turnover. *Journal of Regional Science*, 48(2):311–328.

Gallé, J., Overbeck, D., Riedel, N., and Seidel, T. (2024). Place-based policies, structural change and female labor: Evidence from india's special economic zones. *Journal of Public Economics*, 240:105259.

Garg, T. and Shenoy, A. (2021). The ecological impact of place-based economic policies. *American Journal of Agricultural Economics*, 103(4):1239–1250.

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

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, 225(2):254–277.

Hasan, R., Jiang, Y., and Rafols, R. M. (2021). Place-based preferential tax policy and industrial development: Evidence from india's program on industrially backward districts. *Journal of Development Economics*, 150:102621.

Henderson, J. V. et al. (2003). Urbanization and economic development. *Annals of economics and finance*, 4:275–342.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, 102(2):994–1028.

Heshmati, A. and Rashidghalam, M. (2020). Measurement and analysis of urban infrastructure and its effects on urbanization in china. *Journal of Infrastructure Systems*, 26(1):04019030.

Holtz-Eakin, D. and Selden, T. M. (1995). Stoking the fires? co2 emissions and economic growth. *Journal of public economics*, 57(1):85–101.

Jedwab, R., Christiaensen, L., and Gindelsky, M. (2017). Demography, urbanization, and development: Rural push, urban pull, and ... urban push? *Journal of Urban Economics*, 98:6–16.

Jiboye, A. D. (2011). Sustainable urbanization: Issues and challenges for effective urban governance in nigeria. *Journal of sustainable Development*, 4(6):211.

Kahn, M. E. and Schwartz, J. (2008). Urban air pollution progress despite sprawl: The "greening" of the vehicle fleet. *Journal of Urban Economics*, 63(3):775–787.

Kline, P. and Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *Quarterly Journal of Economics*, 129(1):275–331.

Kolko, J. (2010). Urbanization, agglomeration, and coagglomeration of service industries. In *Agglomeration economics*, pages 151–180. University of Chicago Press.

Koster, H. R. and Van Ommeren, J. (2019). Place-based policies and the housing market. *Review of Economics and Statistics*, 101(3):400–414.

Liang, L., Wang, Z., and Li, J. (2019). The effect of urbanization on environmental pollution in rapidly developing urban agglomerations. *Journal of cleaner production*, 237:117649.

Liu, Q., Wang, X., and Zhang, H. (2022). How does the selection of national development zones affect urban green innovation? evidence from china. *PLOS ONE*, 17(6):e0268111.

Lu, Y., Wang, J., and Zhu, L. (2019). Place-based policies, creation, and agglomeration economies: Evidence from china's economic zone program. *American Economic Journal: Economic Policy*, 11(3):325–360.

Maparu, T. S. and Mazumder, T. N. (2017). Transport infrastructure, economic development and urbanization in india (1990–2011): Is there any causal relationship? *Transportation research part A: policy and practice*, 100:319–336.

Marshall, A. (1890). 1920. principles of economics. *London: Mac-Millan*, pages 1–627.

Martin, R. and Zhang, L. (2021). Changing population exposure to pollution in china's special economic zones. *AEA Papers and Proceedings*, 111:74–78.

Mendez, P., Atienza, M., and Modrego, F. (2023). Urbanization and productivity at a global level: new empirical evidence for the services sector. *Regional Science Policy & Practice*, 15(9):1981–1998.

Moomaw, R. L. and Shatter, A. M. (1996). Urbanization and economic development: a bias toward large cities? *Journal of Urban Economics*, 40(1):13–37.

Neumark, D. and Simpson, H. (2015). Place-based policies. In *Handbook of Regional and Urban Economics*, volume 5, page 1197–1287.

Nordhaus, W. D. (1977). Economic growth and climate: the carbon dioxide problem. *The American Economic Review*, 67(1):341–346.

Ochoa, J. J., Tan, Y., Qian, Q. K., Shen, L., and Moreno, E. L. (2018). Learning from best practices in sustainable urbanization. *Habitat international*, 78:83–95.

Pandey, B., Brelsford, C., and Seto, K. C. (2022). Infrastructure inequality is a characteristic of urbanization. *Proceedings of the National Academy of Sciences*, 119(15):e2119890119.

Pflüger, M. (2021). City size, pollution and emission policies. *Journal of Urban Economics*, 126:103391.

Porter, M. E. et al. (1998). *Clusters and the new economics of competition*, volume 76. Harvard Business Review Boston.

Pradhan, R. P., Arvin, M. B., and Nair, M. (2021). Urbanization, transportation infrastructure, ict, and economic growth: A temporal causal analysis. *Cities*, 115:103213.

Quigley, J. M. (2009). Urbanization, agglomeration, and economic development. *Urbanization and growth*, 115:1–36.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.

Sato, Y. and Zenou, Y. (2015). How urbanization affect employment and social interactions. *European Economic Review*, 75:131–155.

Schafran, A., McDonald, C., Lopez Morales, E., Akyelken, N., and Acuto, M. (2018). Replacing the services sector and three-sector theory: urbanization and control as economic sectors. *Regional Studies*, 52(12):1708–1719.

Scott, A. J. (2009). World development report 2009: reshaping economic geography.

Shenoy, A. (2018). Regional development through place-based policies: Evidence from a spatial discontinuity. *Journal of Development Economics*, 130:173–189.

Spence, M., Annez, P. C., and Buckley, R. M. (2008). *Urbanization and growth*. World Bank Publications.

Tan, Y., Xu, H., and Zhang, X. (2016). Sustainable urbanization in china: A comprehensive literature review. *Cities*, 55:82–93.

Van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C., et al. (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology*, 55(22):15287–15300.

Wang, S., Gao, S., Li, S., and Feng, K. (2020). Strategizing the relation between urbanization and air pollution: Empirical evidence from global countries. *Journal of Cleaner Production*, 243:118615.

Xu, Z. and Luo, J. (2025). Sustainable urbanization: Unpacking the link between urban clusters and environmental protection. *Sustainability*, 17(3):873.

Yu, Y., Li, K., Duan, S., and Song, C. (2023). Economic growth and environmental pollution in china: New evidence from government work reports. *Energy Economics*, 124:106803.

Zhang, D., Xu, J., Zhang, Y., Wang, J., He, S., and Zhou, X. (2020). Study on sustainable urbanization literature based on web of science, scopus, and china national knowledge infrastructure: A scientometric analysis in citespace. *Journal of cleaner production*, 264:121537.

Zhao, C. and Qu, X. (2024). Place-based policies, rural employment, and intra-household resources allocation: Evidence from china's economic zones. *Journal of Development Economics*, 167:103210.

Zhao, C., Xie, R., Ma, C., and Han, F. (2022). Understanding the haze pollution effects of china's development zone program. *Energy Economics*, 111:106078.

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., et al. (2021). Machine learning: new ideas and tools in environmental science and engineering. *Environmental science & technology*, 55(19):12741–12754.

Online Appendix

B Policy Details: The *Rurban* Mission

B.1 Selection Criteria for Rurban Clusters

Herein, we outline the criteria used by the Government of India to select treated clusters under the Rurban Mission.[1] There are two types of clusters under the policy: *tribal* and *non-tribal*; each of these categories differ in their selection criteria.

Selection Criteria for non-Tribal Clusters

For the selection of Non-Tribal clusters, the Ministry of Rural Development would provide a list of leading sub-districts to each State within which the clusters are identified. The selection of these sub-districts by the Ministry is based on parameters such as

1. Decadal Growth in Rural Population.
2. Decadal Growth in non-farm workforce participation.
3. Presence of Economic Clusters.
4. Presence of places of tourism and pilgrimage significance.
5. Proximity to Transport Corridors.

After assigning appropriate weightage to each parameter, the State Governments select the clusters, within each sub-district identified by the Ministry, based on the following additional performance parameters:

1. Decadal Growth in Rural Population. Rise in Land Values.
2. Decadal Growth in Non-Farm work force participation.
3. Percentage Enrollment of girls in secondary schools.
4. Percentage Households with Bank accounts under Pradhan Mantri Jan Dhan Yojana.
5. Performance in Swachh Bharat Mission (Grameen).
6. Good Governance Initiatives by Gram Panchayats.

[1] This information is publicly available at the official website of the Rurban Mission (https://rurban.gov.in/).

Any other factor that the States may consider relevant may also be included. However, a total weightage of 80% would be given for the first four parameters, and the States will have the flexibility to choose the last three parameters, subject to a total weightage of 20%.

Selection Criteria for Tribal Clusters

For the identification of *Tribal* clusters, the Ministry would select the leading sub-districts within the top 100 tribal districts of the country based on the population shares of the Scheduled Tribes (STs). The selection of these sub-districts would be based on parameters similar to the criteria used for Tribal clusters, including:

1. Decadal growth in Tribal Population. Current Tribal Literacy Rate.
2. Decadal Growth in Non-Farm work force participation.
3. Decadal growth in Rural Population and
4. Presence of Economic Clusters.

After according appropriate weightage to each of these parameters, the Ministry selects sub-districts for State Governments to choose clusters from. The selection of clusters by the States would then be based on the following criteria:

1. Decadal growth in Tribal Population.
2. Growth in Tribal Literacy rates.
3. Decadal Growth in Non-Farm work force participation.

Any other factor which the States may consider relevant may be included in addition to the above three parameters, provided the weightage for the above three parameters is not reduced below 80%.

B.2 Components of the *Rurban* Mission

B.2.1 Convergence

To facilitate the convergence of existing policies under the program, the following components were identified as crucial: (a) Household level: Skill development training linked to economic activities, sanitation, provision of piped water supply, solid and liquid waste management, LPG gas connections. (b) Village level: Village streets and drains, street lights.

(c) Cluster level: Agro-processing, agricultural services, storage and warehousing, fully equipped mobile health units, upgrading school/higher education facilities, inter-village road connectivity, public transport, digital literacy, and Citizen Service Centres (CSC) for the electronic delivery of citizen-centric services (e-gram connectivity). For each of the components mentioned, the policy document also outlines a minimum threshold that should be achieved. For instance, under Skill Development Training Linked to Economic Activities, at least 70

The identified Rurban cluster is referred to as a project, to be implemented within a fixed timeframe of three years, through the integration and convergence of project components. The funding for cluster projects follows a 70:30 ratio, where 70% of the funding for proposed interventions comes from converged schemes (covering major capital costs for the projects).

Once the existing policies are identified, the Government of India provides Critical Gap Funding (CGF) to supplement any shortfall in funding through various government schemes, ensuring that the desired outcomes for the cluster are met. The entire CGF is funded by the Ministry of Rural Development. However, CGF is capped at 30

B.2.2 ICAP

The second key feature of the policy is its emphasis on developing an Integrated Cluster Action Plan (ICAP) for each cluster. The rationale behind ICAP is to establish a planned layout with an enforcement mechanism, similar to those found in urban areas, ensuring systematic and sustainable rural development.

Under the Shyama Prasad Mukherji Rurban Mission (SPMRM), the ICAP follows a structured approach. The process begins with the selection of clusters, where suitable village groups are identified based on demographic and economic factors. Once selected, the planning area is delineated and officially notified, ensuring clear administrative boundaries. A detailed cluster profile is then created, covering various aspects such as demographic characteristics, socio-economic conditions, cultural attributes, and administrative structures.

Following this, a deficiency analysis is conducted to identify gaps in infrastructure and economic opportunities. To address these shortcomings, consultations with key stakeholders help in determining the priority mission components for development. The ICAP is then aligned with existing government schemes, ensuring optimal utilization of financial resources.

To facilitate implementation, a structured investment and phasing plan is developed, outlining the financial requirements over a three-year period. Additionally, Critical Gap Funding (CGF) is estimated to bridge any resource shortages. An implementation strategy is formulated, detailing execution and monitoring mechanisms, followed by an operation and maintenance (O&M) plan to ensure the long-term sustainability of the initiatives.

Before finalizing the ICAP, Gram Sabha resolutions are obtained to secure local community approval, after which the plan is submitted to the Ministry of Rural Development (MoRD) for approval. Once approved, the plan undergoes revisions based on Detailed Project Report (DPR) costing. To ensure that the development strategy remains relevant and responsive to evolving needs, the ICAP is iterated every five years, enabling continuous and adaptive rural development.

B.2.3 Institutional Framework

The states are the key implementers under the National Rurban Mission (NRuM). NRuM emphasizes the engagement of key stakeholders at the national, state, district, and Gram Panchayat levels. To facilitate this, a new institutional framework is proposed under NRuM, as illustrated in Figure **??**. At the national level, the National Mission Directorate (NMD) oversees the implementation of NRuM. Another unit, called the National Management Unit, assists the NMD with the selection of clusters, the preparation of ICAPs and spatial plans, and provides guidance to states in these processes.

At the state level, an empowered committee is formed to approve ICAPs and CGF applications before forwarding them to the ministry for approval. Additionally, a State Nodal Agency (SNA) operates under the administrative domain of the State Government's Rural Development and/or Panchayati Raj Department. A technical and operational unit is attached to the SNA, which includes specialists in convergence, planning, engineering, and MIS.

The SNA serves as a bridge between the national agency and the district-level agency, i.e., the District Project Management Unit (DPMU). The DPMU is responsible for: (i) Coordinating with implementing departments/agencies for the notification of planning areas and spatial planning matters. (ii) Ensuring the convergence of schemes planned in the ICAP in an integrated and time-bound manner.

One of the most innovative aspects introduced as part of the institutional framework is the formation of the Cluster Development and Management Unit (CDMU) at the cluster level. Institutions at such a granular level, specifically designed for planning and managing the

program, were previously absent in rural areas. This unit consists of at least two professionals a spatial planning professional and a rural management/development professional. The unit closely monitors the spatial planning aspects and the ICAP preparation for the cluster. Additionally, it tracks the progress of activities within the cluster and provides regular updates to the DPMU/SPMU.