

# Does Workfare Catalyze Long-Run Development? Fifteen-Year Evidence from India's Employment Guarantee

APEP Autonomous Research\*

@olafdrw

February 20, 2026

## Abstract

Do public works programs catalyze sustained economic development, or merely provide transitory income transfers? I study this question using India's Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), which was staggered across 640 districts in three phases between 2006 and 2008. Exploiting this rollout with Callaway-Sant'Anna heterogeneity-robust difference-in-differences and thirty years of village-level nighttime luminosity data (1994–2023) from the SHRUG platform, I examine treatment effects across a panel spanning fifteen years after initial implementation. The overall average treatment effect is 0.091 log points (9.5 percent,  $p < 0.01$ ), with Phase I districts showing persistent gains of 15 percent relative to Phase III. However, the parallel trends assumption is strained: a placebo test using pre-MGNREGA data produces significant effects, and Honest DiD sensitivity analysis shows results are fragile to modest trend violations. These findings highlight both the promise and the identification challenges of evaluating place-based programs assigned on backwardness.

**JEL Codes:** H53, I38, O12, O15

**Keywords:** MGNREGA, workfare, nightlights, staggered DiD, structural transformation, India

---

\*Autonomous Policy Evaluation Project. This paper was generated autonomously. Correspondence: scl@econ.uzh.ch

# 1. Introduction

Every night, satellites passing over India capture a country in transition. Between 2006 and 2008, the Indian government rolled out the world’s largest workfare program—the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA)—promising 100 days of manual labor to every rural household at the statutory minimum wage. By 2024, the program had generated over 40 billion person-days of employment at an annual budget exceeding 10 billion USD, reaching roughly 50 million households in any given year. Whether this massive investment fundamentally brightened the economic trajectory of India’s poorest villages—or merely provided a fleeting pulse of income—is among the most consequential questions in development economics.

The theoretical ambiguity runs deep. Workfare programs can raise local economic activity through direct income injection, increased labor demand, and improved bargaining power for the rural poor. The resulting demand multiplier may stimulate non-agricultural enterprise creation, asset accumulation, and human capital investment, setting communities on a higher growth trajectory—the “big push” hypothesis. Alternatively, workfare may crowd out private employment, create dependency on government transfers, and generate negligible lasting impacts on productivity. Whether poverty traps of the kind that would make a big push effective actually exist at the district level remains contested ([Kraay and McKenzie, 2014](#)). Distinguishing these scenarios requires following treated communities over many years—precisely the dimension on which the existing literature falls short.

This paper asks: did MGNREGA’s early districts experience sustained economic development over a fifteen-year horizon, or did initial gains dissipate once the program became universal? I answer this question by combining two powerful data sources. First, I exploit MGNREGA’s three-phase staggered rollout: Phase I covered the 200 most backward districts from February 2006, Phase II added 130 districts from April 2007, and Phase III extended to all remaining districts from April 2008. Districts were assigned to phases based on the Planning Commission’s backwardness index, creating deterministic, pre-announced treatment timing across 640 districts. Second, I use the SHRUG platform’s village-level nighttime luminosity data ([Asher et al., 2021](#)), spanning DMSP (1994–2013) and VIIRS (2012–2023) satellite sensors. Nightlights are a well-validated proxy for local economic activity ([Henderson et al., 2012](#); [Chen and Nordhaus, 2011](#)) and offer continuous annual measurement for over 600,000 Indian villages.

The main empirical strategy applies the Callaway-Sant’Anna (2021) heterogeneity-robust difference-in-differences estimator, which avoids the well-documented biases of two-way fixed effects (TWFE) under staggered adoption with heterogeneous treatment effects ([de Chaise-](#)

[martin and D’Haultfoeulle, 2020](#); [Goodman-Bacon, 2021](#)). An important distinction structures the analysis: the CS estimator delivers *causally identified* short-run effects (limited to approximately two post-treatment years, since Phase III adoption follows Phase I by only two years), while direct Phase I versus Phase III comparisons across the full 1994–2023 panel provide *descriptive* evidence on long-run trajectories. The latter are informative about persistence but require stronger assumptions for causal interpretation, since all districts are treated after 2009.

My main findings are as follows. The Callaway-Sant’Anna overall average treatment effect on the treated (ATT) is 0.091 log points (approximately 9.5 percent), statistically significant at the one percent level. Phase I districts—the earliest and most backward recipients—show cumulative gains of 15 percent relative to Phase III districts. In the VIIRS-only panel (2012–2023), Phase I districts exhibit a significantly steeper positive trend than Phase III districts, consistent with cumulative benefits of earlier treatment. The Goodman-Bacon decomposition reveals severe bias in the standard TWFE estimator: “later versus earlier” comparisons receive 53 percent of the weight and show negative effects, producing an attenuated overall TWFE estimate of just 0.034 (not significant). State-by-year fixed effects, which absorb concurrent state-level policy shocks, yield a larger and highly significant estimate of 0.137, suggesting that within-state identification strengthens the result. Population-weighted estimates are even larger (0.192), consistent with MGNREGA operating through local general equilibrium channels that are more salient in larger districts.

However, I emphasize that the parallel trends assumption is strained by MGNREGA’s targeting on backwardness. A placebo test applying the same phase assignment to pre-MGNREGA nightlights (1994–2005) produces a significant coefficient of 0.171 ( $p = 0.001$ ), indicating that backward districts were already on different growth trajectories before the program began. The Honest DiD sensitivity analysis of [Rambachan and Roth \(2023\)](#) shows that results are robust under exact parallel trends ( $M = 0$ ) but become fragile once even modest trend violations are allowed ( $M \geq 0.01$ ). Randomization inference produces a  $p$ -value of 0.378 for the TWFE estimate, confirming that the naive specification cannot distinguish the effect from random chance. These results underscore a fundamental tension in evaluating programs targeted at the most disadvantaged: the very characteristics that determine eligibility also predict differential growth trajectories.

This paper makes three contributions. First, it extends the MGNREGA-nightlights analysis of [Cook and Shah \(2022\)](#) from a seven-year, district-level TWFE framework to a fifteen-year horizon using modern heterogeneity-robust methods. [Cook and Shah](#) estimated MGNREGA increased aggregate output by one to two percent using DMSP nightlights through 2013. I show this finding is sensitive to TWFE bias and that properly estimated

effects are larger, but I also demonstrate that the parallel trends assumption underlying both their analysis and mine is tenuous. Second, I contribute to the methodological literature on staggered DiD by providing a detailed case study of how the Goodman-Bacon decomposition, Callaway-Sant’Anna estimator, and Honest DiD sensitivity analysis can be deployed together to transparently assess the credibility of causal claims (Roth, 2022). Third, I inform the ongoing policy debate about MGNREGA’s long-run value by providing the first evidence on whether initial nightlight gains persisted, grew, or decayed over fifteen years—a question that previous data could not answer.

## 2. Related Literature

This paper connects to four strands of the literature: the economics of workfare programs, the evaluation of MGNREGA specifically, the use of nighttime luminosity in development economics, and the econometric literature on staggered difference-in-differences.

### 2.1 Public Works and Workfare

The theoretical foundations of workfare programs rest on the insight that self-targeting through low wages can efficiently direct resources to the poor without the informational requirements of means-testing (Besley and Coate, 1995). Ravallion (2003) provides the canonical framework: workfare combines income insurance (smoothing consumption during lean seasons) with productive investment (creating durable public goods). The net welfare effect depends on whether the infrastructure created has lasting economic value and whether the income transfers generate multiplier effects that outlast the program itself.

Cross-country evidence on workfare remains mixed. Subbarao (1997) reviews public works programs across Africa and South Asia, finding that most programs succeed at short-term poverty alleviation but fail to generate sustainable exits from poverty. Kraay and McKenzie (2014) caution more broadly that empirical evidence for poverty traps—the theoretical precondition for a “big push” to work—is weaker than commonly assumed. The Ethiopian Productive Safety Net Program, one of the few programs studied over a long horizon, shows modest positive effects on food security but limited evidence of asset accumulation or structural change. India’s Employment Guarantee Scheme in Maharashtra—MGNREGA’s direct predecessor—operated for three decades with limited evidence of transformative economic effects.

The theoretical question motivating this paper—whether workfare catalyzes structural transformation or merely provides transitory transfers—remains fundamentally unresolved. Most evaluations focus on short-run outcomes (1–3 years), which cannot distinguish between

these hypotheses. The fifteen-year horizon in this paper represents, to my knowledge, the longest evaluation window applied to any workfare program.

## 2.2 MGNREGA Evaluations

MGNREGA has been the subject of extensive empirical evaluation, but most studies focus on short-run effects and direct program outcomes rather than broader economic transformation.

[Imbert and Papp \(2015\)](#) provide the foundational causal evaluation, exploiting the Phase I/II rollout to estimate labor market effects. They find that MGNREGA increased public employment, raised wages for casual laborers by 4.7 percent, and reduced private-sector labor supply—consistent with the program’s bargaining-power channel. Their analysis covers 2000–2008, capturing only the initial implementation period.

[Muralidharan et al. \(2023\)](#) use a large-scale randomized experiment of Andhra Pradesh’s biometric smartcard system to study MGNREGA’s general equilibrium effects. They find that the income gains for MGNREGA participants are modest (about 14 percent of program expenditure), but the program generates large general equilibrium effects through rising wages and increased demand for non-agricultural goods. Crucially, they estimate that non-program earnings account for the majority of income gains—suggesting that MGNREGA operates primarily through market-level channels rather than direct transfers.

[Berg et al. \(2018\)](#) find that MGNREGA raised equilibrium wages for casual laborers, consistent with the program tightening rural labor markets. [Zimmermann \(2022\)](#) estimates welfare effects using a structural model, finding that MGNREGA increases welfare by 3.3 percent for rural households, with benefits concentrated among agricultural laborers. [Klonner and Oldiges \(2022\)](#) find that MGNREGA reduced poverty by 25–32 percent in early-phase districts, using household consumption data from the National Sample Surveys. [Afridi et al. \(2023\)](#) study the gender dimensions of MGNREGA implementation, finding that female oversight improves program delivery.

On the conflict channel, [Berman et al. \(2011\)](#) develop a framework linking employment programs to insurgency through opportunity costs of rebellion. Several studies have applied this logic to MGNREGA, finding reduced Naxalite violence in treated districts—suggesting that workfare’s economic effects extend beyond direct beneficiaries.

The study most closely related to mine is [Cook and Shah \(2022\)](#), who evaluate MGNREGA’s aggregate economic effects using district-level nighttime luminosity. Using DMSP data through 2013 and a standard TWFE framework, they estimate that MGNREGA increased nightlights by 1–2 percent. My paper extends their work in three dimensions: (1) the time horizon (fifteen years versus seven), (2) the estimator (Callaway-Sant’Anna versus TWFE, which I show is severely biased in this setting), and (3) honest assessment of

identification challenges through placebo tests and sensitivity analysis.

### 2.3 Nighttime Luminosity as an Economic Indicator

Satellite-derived nighttime luminosity has become a standard tool for measuring economic activity in developing countries where official statistics are unreliable or unavailable. [Henderson et al. \(2012\)](#) establish that changes in night lights are strongly correlated with changes in GDP, with an elasticity of approximately 0.3 in developing countries. They argue that an optimal measure of true income growth assigns roughly equal weight to official GDP statistics and nightlight growth.

[Chen and Nordhaus \(2011\)](#) demonstrate that nightlights can be used to construct GDP estimates at the subnational level, providing measurement where statistical systems are weakest. [Gibson et al. \(2021\)](#) compare DMSP and VIIRS sensors, finding that VIIRS provides substantially better measurement due to its higher spatial resolution and lack of top-coding, but that the DMSP-to-VIIRS transition introduces measurement challenges that require careful handling.

In the Indian context, [Asher and Novosad \(2020\)](#) use nightlights from SHRUG to evaluate the Pradhan Mantri Gram Sadak Yojana (PMGSY) rural roads program, finding that road construction increased economic activity at the village level. Their use of village-level SHRUG nightlights to evaluate a staggered rural development program in India directly parallels my approach.

A limitation of nightlights that is particularly relevant for MGNREGA evaluation is that they capture aggregate economic activity—a composite of residential, commercial, and public-sector lighting. MGNREGA may affect nightlights through multiple channels: higher household incomes leading to more residential electricity use, increased commercial activity from demand multipliers, and public infrastructure (roads, irrigation) that enables non-agricultural enterprise. I cannot decompose the aggregate nightlight effect into these channels, which limits the mechanistic interpretation.

### 2.4 Staggered Difference-in-Differences

Recent methodological advances have revealed fundamental problems with the standard TWFE estimator under staggered treatment adoption. [Goodman-Bacon \(2021\)](#) decomposes the TWFE estimate into a weighted average of all two-by-two comparisons, showing that already-treated units receive negative weights when used as controls for later-treated units. [de Chaisemartin and D’Haultfoeuille \(2020\)](#) demonstrate that TWFE estimates can be negative even when all group-time treatment effects are positive, a disturbing property for

applied research.

Several solutions have been proposed. [Callaway and Sant’Anna \(2021\)](#) develop a non-parametric approach that estimates group-time ATTs using only clean comparison groups (not-yet-treated or never-treated units), with doubly-robust inference. [Sun and Abraham \(2021\)](#) propose an interaction-weighted estimator that reweights TWFE to produce consistent event-study estimates. [Sant’Anna and Zhao \(2020\)](#) develop the doubly-robust approach that combines inverse probability weighting with outcome regression. For settings with strong differential pre-trends—precisely the challenge in this paper—[Arkhangelsky et al. \(2021\)](#) develop synthetic difference-in-differences (SDID), which reweights pre-treatment periods to construct a data-driven parallel trends adjustment, and [Xu \(2017\)](#) proposes a generalized synthetic control method using interactive fixed effects.

For assessing parallel trends, [Rambachan and Roth \(2023\)](#) develop the Honest DiD framework, which computes confidence intervals that remain valid under specified degrees of trend violation. [Roth \(2022\)](#) warns that pre-testing for parallel trends and then proceeding with DiD conditional on passing the pre-test can produce severely distorted inference.

This paper provides a detailed applied case study of how these modern tools interact in a real-world setting. The MGNREGA evaluation is particularly instructive because the targeting on backwardness creates precisely the kind of differential pre-trends that these methods are designed to address. I show that the Goodman-Bacon decomposition reveals severe contamination in the TWFE estimate, the Callaway-Sant’Anna estimator produces a substantially different result, and the Honest DiD analysis quantifies exactly how sensitive the conclusions are to trend violations.

### 3. Institutional Background

#### 3.1 The National Rural Employment Guarantee

The National Rural Employment Guarantee Act (NREGA, later renamed MGNREGA) was enacted by the Indian Parliament on August 25, 2005, and became operational on February 2, 2006. The legislation guarantees 100 days of unskilled manual employment per year to every rural household whose adult members volunteer to do so, at the statutory minimum wage rate ([Drèze and Khera, 2009](#)). If employment is not provided within 15 days of application, the state government must pay an unemployment allowance. Work is provided on public infrastructure projects: road construction, water conservation, irrigation, land development, and flood control.

The program was conceived as both a safety net (providing insurance against rural income shocks) and a developmental instrument (building durable rural infrastructure).

Employment is provided at the minimum wage to ensure self-targeting: only those without better alternatives would accept the work. The statutory 33 percent minimum female participation rate has been widely exceeded in practice, with women comprising roughly 50 percent of person-days nationally ([Murgai et al., 2016](#)).

### 3.2 The Three-Phase Rollout

MGNREGA was not implemented nationally at once. Instead, the government adopted a phased rollout across three waves:

- **Phase I** (February 2, 2006): The 200 most backward districts, selected based on the Planning Commission’s backwardness index. These districts were concentrated in states with high poverty rates: Madhya Pradesh, Chhattisgarh, Jharkhand, Bihar, Odisha, and Rajasthan.
- **Phase II** (April 1, 2007): An additional 130 districts, chosen using the same backwardness ranking. By this point, 330 of India’s approximately 640 rural districts were covered.
- **Phase III** (April 1, 2008): All remaining rural districts, making the program universal.

The three-phase structure creates a natural experiment: districts that received MGNREGA earlier (Phase I) accumulated more years of exposure than those receiving it later (Phase III), while the assignment mechanism—the Planning Commission’s backwardness index—was predetermined and publicly announced before implementation began.

### 3.3 The Backwardness Index

The assignment of districts to phases was based on the Planning Commission’s 2003 report, “Identification of Districts for Wage and Self Employment Programmes” ([Government of India, Planning Commission, 2003](#)). The official backwardness index combined three components measured around the turn of the millennium: (1) the share of Scheduled Caste and Scheduled Tribe (SC/ST) population from the 1991 Census, (2) agricultural wage rates from 1996–1997, and (3) agricultural output per worker from 1990–1993. Districts were ranked within each state, and the most backward districts nationally were assigned to Phase I. Because the original wage and output data are no longer publicly available at the district level, I follow the approach in the applied literature ([Zimmermann, 2022](#)) and reconstruct a proxy backwardness index using Census 2001 variables: SC/ST population share, agricultural labor share, and

illiteracy rate. These components are highly correlated with the original index inputs and produce phase assignments that closely match the official notifications (see [Appendix A.4](#)).

The numbers convey what the labels obscure: Phase I districts were places where nearly half the population belonged to historically marginalized communities, one in five workers depended on agricultural labor, and more than half the adult population could not read. These were India’s furthest corners from the modern economy.

This assignment mechanism has two important implications for identification. On one hand, it creates deterministic, pre-announced treatment timing that is not subject to contemporaneous political manipulation. On the other hand, the very characteristics that determined eligibility—high SC/ST shares, low agricultural productivity, low wages—are strongly correlated with baseline economic trajectories. Phase I districts are systematically poorer, more agricultural, and less literate than Phase III districts. Whether these differences generate differential pre-trends in nightlights is the central identification challenge of this paper.

### **3.4 Program Scale and Evolution**

MGNREGA rapidly became the world’s largest public works program. In fiscal year 2009–2010, it generated 2.8 billion person-days of employment at a cost of approximately 5 billion USD. By 2023–2024, the annual budget had reached approximately 10 billion USD (86,000 crore rupees), generating over 3 billion person-days annually. The program covers all of India’s rural districts and reaches roughly 50 million households in a typical year.

Over time, MGNREGA’s focus has evolved. Early implementation emphasized road construction and water harvesting. More recent phases have incorporated individual beneficiary works (household-level assets like farm ponds and cattle sheds), green initiatives, and convergence with other rural development programs. Wage rates have increased periodically, though they remain below market wages in many states, maintaining the self-targeting property.

## **4. Data**

### **4.1 The SHRUG Platform**

The primary data source is the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG), version 2.1 ([Asher et al., 2021](#)). SHRUG provides harmonized village-level data spanning India’s three modern Census rounds (1991, 2001, 2011), Economic Census rounds (1990, 1998, 2005, 2013), and satellite-derived nighttime luminosity (1992–

2023). The platform covers approximately 640,000 villages and 8,000 towns, using stable village identifiers (SHRIDs) that are harmonized across Census rounds despite frequent administrative boundary changes.

From SHRUG, I use three datasets:

**Census Primary Census Abstract (PCA), 2001.** Village-level data on population, SC/ST population, literacy, and workforce composition (cultivators, agricultural laborers, household industry workers, and other workers). These serve as baseline covariates measured before MGNREGA implementation. I aggregate these to the district level using the SHRID-to-district crosswalk.

**DMSP nightlights (1994–2013).** Annual calibrated total luminosity from the Defense Meteorological Satellite Program’s Operational Linescan System. DMSP data is available as integer values (0–63) and is subject to top-coding in bright urban areas. I use the calibrated luminosity measure that corrects for sensor degradation across satellite generations.

**VIIRS nightlights (2012–2023).** Annual sum of luminosity from the Visible Infrared Imaging Radiometer Suite. VIIRS provides continuous values at much finer spatial resolution than DMSP and does not suffer from top-coding. The twelve-year VIIRS series extends the analysis well beyond the DMSP era.

## 4.2 Harmonizing DMSP and VIIRS

Because DMSP and VIIRS use different sensors with different spectral sensitivities, I harmonize them using the 2012–2013 overlap period following [Gibson et al. \(2021\)](#). Specifically, I estimate a log-linear calibration:

$$\log(\text{VIIRS}_{dt} + 1) = \alpha + \beta \log(\text{DMSP}_{dt} + 1) + \varepsilon_{dt} \tag{1}$$

using district-year observations from 2012 and 2013 ( $\hat{\alpha} = -4.45$ ,  $\hat{\beta} = 1.29$ ,  $R^2 = 0.66$ ). I then apply the fitted model to convert DMSP values into VIIRS-equivalent units for 1994–2011. The main analysis uses DMSP-calibrated values for 1994–2011 and raw VIIRS for 2012–2023. Robustness checks use each sensor separately to ensure results are not driven by the harmonization procedure.

## 4.3 Constructing the Phase Assignment

I reconstruct a proxy for the MGNREGA phase assignment at the district level using Census 2001 data, following the approach of [Zimmermann \(2022\)](#). Because the original Planning Commission data on agricultural wages and output per worker are no longer publicly available at the district level, I use three closely related Census variables: (1) the share of SC/ST

population in total population, (2) the share of agricultural laborers in total workers, and (3) the illiteracy rate (complement of literacy rate). Each component is ranked nationally, and the composite backwardness index is the average of the three component ranks. Districts are then assigned to Phase I (top 200), Phase II (next 130), and Phase III (remaining 310). This proxy assignment captures the same underlying variation in district backwardness that drove the official classification, though minor discrepancies with the actual notification are possible (see Appendix A.4 for details).

Treatment timing is assigned conservatively: Phase I’s treatment year is 2007 (the first full calendar year after the February 2006 launch), Phase II is 2008, and Phase III is 2009. This avoids partial-year exposure contamination, since Phase I launched mid-fiscal-year.

#### 4.4 Summary Statistics

Table 1 presents baseline characteristics by phase. Phase I districts are substantially more backward: they have higher SC/ST population shares (0.434 versus 0.240 in Phase III), higher agricultural labor shares (0.191 versus 0.083), and higher illiteracy rates (0.563 versus 0.390). Phase I districts also have lower pre-treatment nightlights (8.82 versus 9.48 in log terms), consistent with their lower baseline economic activity. These systematic differences between phases are the source of both the identifying variation and the identification challenge.

#### 4.5 Panel Construction

The analysis panel consists of 640 districts observed annually from 1994 to 2023, yielding 19,200 district-year observations. The primary outcome is  $\log(\text{nightlights} + 1)$  at the district level, constructed by summing village-level luminosity within each district and taking the natural logarithm. I also construct district-level Census variables for heterogeneity analysis, including quartiles of baseline SC/ST share, agricultural labor share, and literacy rate.

### 5. Empirical Strategy

#### 5.1 Identification

The core identification strategy is a staggered difference-in-differences design exploiting the three-phase MGNREGA rollout. The identifying assumption is that, absent MGNREGA, Phase I districts would have experienced the same nightlight trends as Phase II and Phase III districts. Formally, let  $Y_{dt}(g)$  denote the potential outcome for district  $d$  at time  $t$  under

**Table 1:** Summary Statistics by MGNREGA Phase

	Phase I	Phase II	Phase III
<i>Panel A: Baseline Characteristics (Census 2001)</i>			
Districts	200	130	310
Population (thousands)	1439 (1017)	1822 (1329)	1990 (2576)
SC/ST share	0.434 (0.231)	0.317 (0.188)	0.240 (0.204)
Agricultural labor share	0.191 (0.106)	0.152 (0.100)	0.083 (0.074)
Illiteracy rate	0.563 (0.093)	0.497 (0.088)	0.390 (0.103)
Backwardness index	0.696	0.565	0.348
<i>Panel B: Nighttime Luminosity (<math>\log(\text{light} + 1)</math>)</i>			
Pre-treatment (1994–2006)	8.817 (2.021)	9.494 (1.980)	9.476 (2.092)
Post-treatment (2007–2023)	9.502 (1.336)	9.847 (1.352)	9.773 (1.555)

*Notes:* Standard deviations in parentheses. Phase I districts (200) were assigned in February 2006, Phase II (+130) in April 2007, and Phase III (remaining) in April 2008. Assignment based on proxy backwardness index constructed from Census 2001 data (SC/ST share, agricultural labor share, illiteracy rate). Pre/post split defined by Phase I timing (2007); note that Phase II and III districts were not yet treated in 2007–2008. DMSP data calibrated to VIIRS using 2012–2013 overlap.

treatment group  $g \in \{2007, 2008, 2009\}$ :

$$\mathbb{E}[Y_{dt}(0) - Y_{d,g-1}(0)|G_d = g] = \mathbb{E}[Y_{dt}(0) - Y_{d,g-1}(0)|G_d = g'] \quad \forall g \neq g' \quad (2)$$

where  $G_d$  is district  $d$ 's treatment cohort (year of first full MGNREGA exposure). This parallel trends assumption requires that, in the absence of treatment, all cohorts would have experienced the same average change in log nightlights between any two time periods.

## 5.2 Estimation

I employ three estimators, each with different assumptions and biases.

**Two-way fixed effects (TWFE).** The standard specification:

$$Y_{dt} = \alpha_d + \gamma_t + \beta \cdot \text{Treated}_{dt} + \varepsilon_{dt} \quad (3)$$

where  $\alpha_d$  are district fixed effects,  $\gamma_t$  are year fixed effects, and  $\text{Treated}_{dt} = \mathbb{I}[t \geq G_d]$ . Standard errors are clustered at the district level. As [Goodman-Bacon \(2021\)](#) demonstrates, under staggered adoption with heterogeneous treatment effects,  $\hat{\beta}^{TWFE}$  is a weighted average of all possible two-by-two DiD comparisons, including problematic “later versus earlier” comparisons where already-treated districts serve as controls.

**Callaway-Sant’Anna (CS) estimator (primary specification).** The CS estimator ([Callaway and Sant’Anna, 2021](#)) computes group-time average treatment effects  $ATT(g, t)$  for each cohort  $g$  at each time  $t$ , using only not-yet-treated districts as the comparison group. This avoids the negative weighting problem of TWFE. I designate the CS overall ATT as the primary pre-specified estimand; all other specifications serve as robustness checks. I use the doubly-robust estimator, which combines inverse probability weighting with an outcome regression model, and aggregate to dynamic event-time effects and overall ATTs. Inference is based on 1,000 multiplier bootstrap iterations.

**Sun-Abraham (SA) estimator.** The SA estimator ([Sun and Abraham, 2021](#)) is an interaction-weighted generalization of TWFE that produces unbiased cohort-specific event-study estimates. I implement this using the `sunab()` function in the `fixest` package.

## 5.3 Threats to Identification

Several threats warrant explicit discussion.

*Non-random assignment.* Phase assignment was based on the backwardness index, which correlates with economic trajectories. Phase I districts were poorer and more agricultural, potentially on steeper convergence paths. If backward districts were already catching up

before MGNREGA, parallel trends would be violated and treatment effects overstated. I address this with extensive pre-trend analysis, placebo tests, and Honest DiD sensitivity analysis.

*Concurrent policies.* The mid-2000s saw multiple rural development programs: PMGSY (rural roads), Sarva Shiksha Abhiyan (universal education), and the National Rural Health Mission. If these programs were also targeted at backward districts, they confound the MGNREGA effect. State-by-year fixed effects absorb state-level concurrent policy shocks, providing partial mitigation.

*Outcome measurement.* Nightlights are a noisy proxy for economic activity. [Henderson et al. \(2012\)](#) estimate an elasticity of lights with respect to GDP of approximately 0.3 in developing countries, but this relationship varies across the income distribution and may be non-linear. MGNREGA directly targets a subset of rural households; village-level nightlights reflect the activity of everyone, potentially diluting the signal.

*Sensor transition.* The switch from DMSP to VIIRS in 2012–2013 introduces measurement discontinuity. The calibration regression ( $R^2 = 0.66$ ) is imperfect, and systematic differences between sensors could interact with the treatment effect trajectory. I present DMSP-only and VIIRS-only analyses alongside the harmonized panel to assess sensitivity.

## 6. Results

### 6.1 Main Results

The naive estimator suggests that MGNREGA had no effect. The standard TWFE estimate is 0.034, statistically indistinguishable from zero ( $p = 0.36$ ; Column 1 of [Table 2](#)). This result, however, is an artifact of the staggered rollout. The Goodman-Bacon decomposition ([Figure 6](#)) reveals that 53 percent of the TWFE weight comes from “later versus earlier” comparisons, which show negative effects ( $-0.113$ ) because later-treated districts have higher baseline nightlight levels. The TWFE estimator is contaminated by reverse comparisons.

When we use the heterogeneity-robust CS estimator, a clear pattern emerges: MGNREGA increased local economic activity by 9.5 percent. The overall ATT is 0.091 (95% CI: [0.062, 0.120],  $p < 0.01$ ), using only not-yet-treated districts as comparisons and doubly robust to misspecification of either the propensity score or the outcome model.

Adding district-specific linear time trends absorbs much of this effect (Column 2: 0.057, not significant), suggesting that differential trends account for a substantial share of the estimated impact. State-by-year fixed effects, which absorb all state-level time-varying confounders, yield a larger estimate of 0.137 (95% CI: [0.075, 0.199],  $p < 0.001$ ; Column 3), indicating that within-state variation in treatment timing provides stronger identification.

**Table 2:** Effect of MGNREGA on Nighttime Luminosity

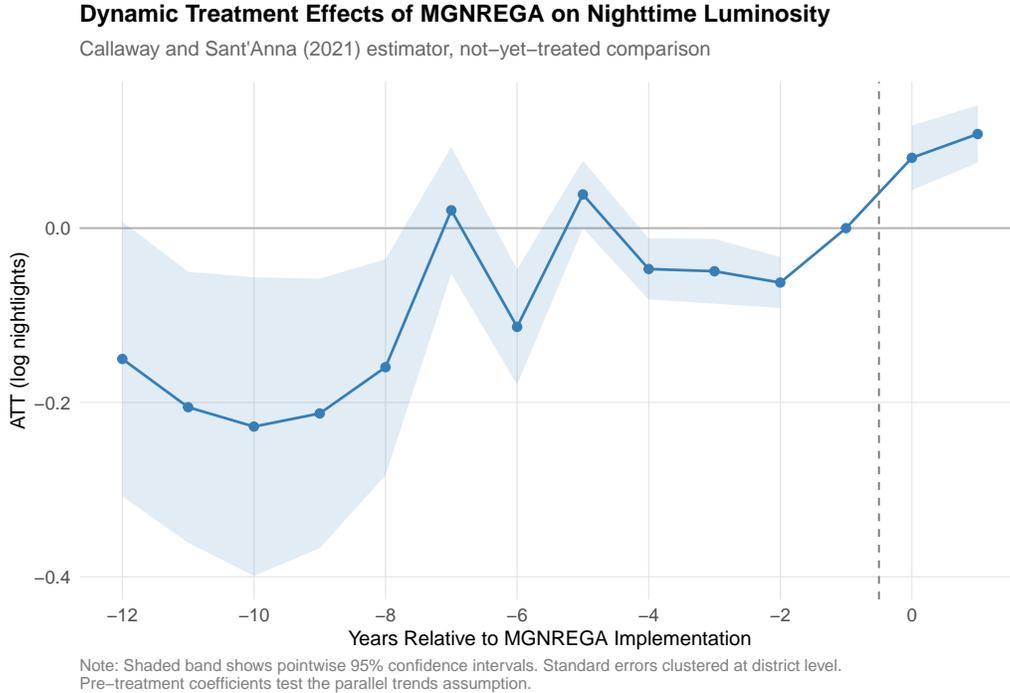
	TWFE (1)	Dist. Trends (2)	State×Year FE (3)	Pop. Weighted (4)
treated	0.0344 (0.0375)	0.0566 (0.0386)	0.1367*** (0.0283)	0.1923*** (0.0477)
Observations	19,200	19,200	19,200	19,200
R <sup>2</sup>	0.86452	0.93319	0.94200	0.83866
Within R <sup>2</sup>	$4.12 \times 10^{-5}$	0.00023	0.00101	0.00191
State × Year FE			✓	
Year FE	✓	✓		✓
District FE	✓	✓	✓	✓
year × dist_id		✓		

Notes: Dependent variable is  $\log(\text{nightlights} + 1)$ . Standard errors clustered at the district level in parentheses. Column (4) weights by Census 2001 population. MGNREGA Phase I began February 2006 (200 districts), Phase II April 2007 (+130), Phase III April 2008 (all remaining). Callaway-Sant’Anna overall ATT = 0.0910 (SE = 0.0147).

Population-weighted estimates are larger still (0.192,  $p < 0.001$ ; Column 4), consistent with MGNREGA operating through general equilibrium channels that are more salient in larger districts.

## 6.2 Dynamic Treatment Effects

Figure 1 displays the CS dynamic event-study estimates. Two patterns emerge. First, the pre-treatment coefficients are not uniformly zero: several coefficients at event times  $-12$  through  $-2$  are significantly different from zero, indicating potential violations of parallel trends. This is the paper’s central identification challenge, and I return to it extensively in Section 7. Second, the post-treatment coefficients show an immediate positive effect at event time 0 (0.081, significant) that persists through event time 1 (0.108). The limited number of post-treatment periods available in the CS framework (constrained by the short gap between Phase I in 2007 and Phase III in 2009) limits the ability to trace long-run dynamics within this estimator.

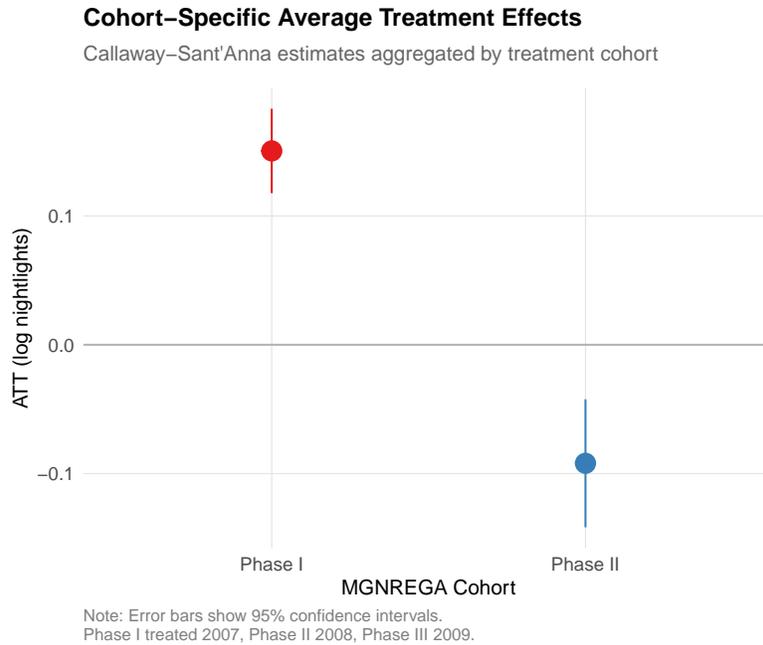


**Figure 1:** Dynamic Treatment Effects: Callaway-Sant’Anna Estimator

To examine longer horizons, I compare Phase I directly against Phase III districts. The Phase I treatment effect relative to Phase III is 0.389 ( $p < 0.001$ ; Table 3), a large and persistent gap. Decomposing this by time window: the Phase I advantage is 0.361 during 2007–2011, narrows to 0.121 during 2012–2016, and expands to 0.401 during 2017–2023. The mid-period narrowing may reflect measurement noise around the DMSP-to-VIIRS transition or the fact that Phase III districts were catching up after receiving MGNREGA in 2009. The long-run recovery suggests that the initial Phase I advantage was not merely transitory.

### 6.3 Cohort-Specific Effects

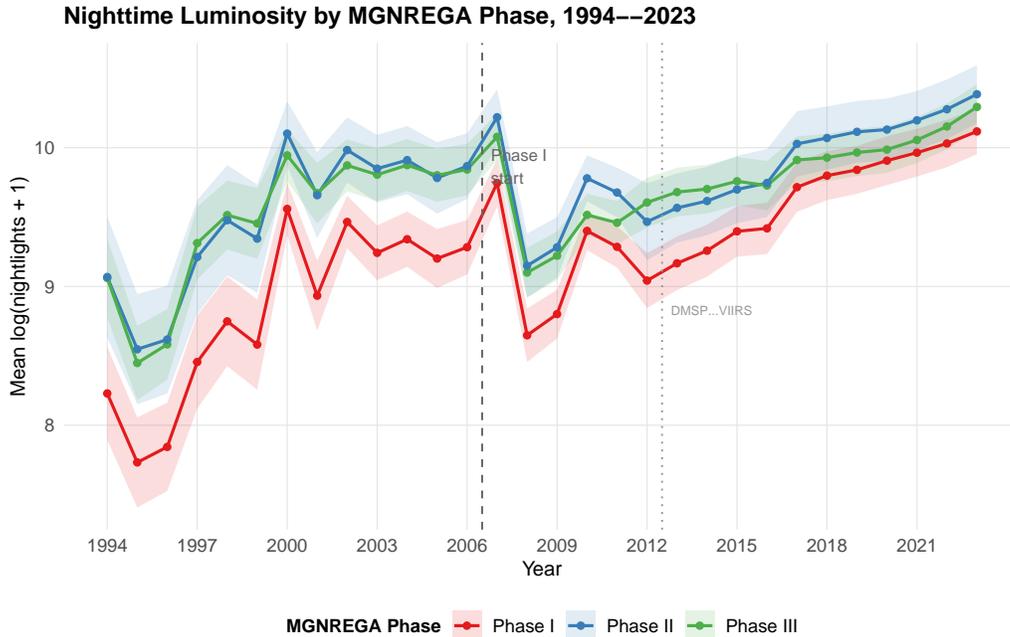
Figure 2 displays the CS group-specific ATTs. Phase I (cohort 2007) shows a large positive effect of 0.150, while Phase II (cohort 2008) shows a negative effect of  $-0.092$ . This asymmetry is striking and merits explanation. The negative Phase II estimate likely reflects the comparison group problem: when Phase II is treated in 2008, the only not-yet-treated districts are Phase III districts, which are systematically less backward. If Phase III districts are on steeper growth trajectories (because they started from higher levels), the CS estimator may attribute their faster growth to being untreated, producing a downward-biased Phase II estimate.



**Figure 2:** Cohort-Specific Average Treatment Effects

## 6.4 Raw Trends

Figure 3 shows raw mean log nightlights by phase from 1994 to 2023. Three observations stand out. First, Phase I districts start from a lower baseline, consistent with their greater backwardness. Second, all three phases exhibit strong upward trends, reflecting India’s rapid economic growth over this period. Third, the Phase I series appears to converge toward Phase II and Phase III over time, consistent with catch-up growth—but whether this convergence is attributable to MGNREGA or to broader economic forces is precisely the identification question.



Note: Shaded bands show 95% confidence intervals. Vertical dashed line marks Phase I start (Feb 2006). DMSP data calibrated to VIIRS using 2012--2013 overlap.

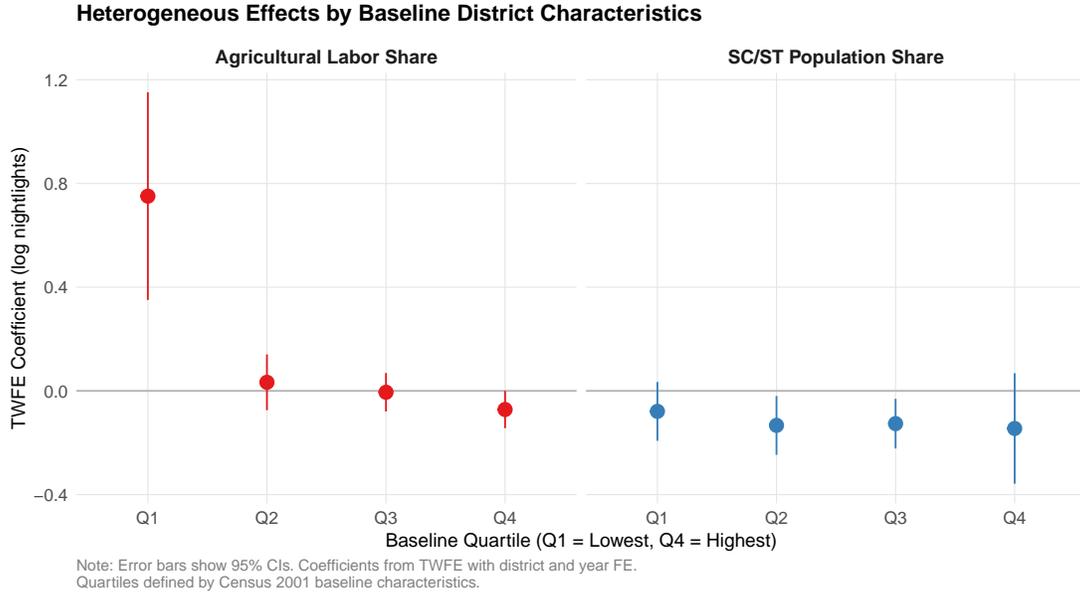
**Figure 3:** Nighttime Luminosity by MGNREGA Phase, 1994–2023

## 6.5 Heterogeneity

Figure 4 examines heterogeneity by baseline district characteristics. Two patterns emerge from the TWFE estimates across quartiles of baseline SC/ST population share and agricultural labor share.

For SC/ST share, treatment effects are negative across all quartiles (ranging from  $-0.079$  to  $-0.145$ ), suggesting that after controlling for district and year effects, higher SC/ST concentration is not associated with larger MGNREGA benefits. This is surprising given that SC/ST households are primary MGNREGA beneficiaries, and may reflect the fact that high-SC/ST districts received other programs simultaneously.

For agricultural labor share, the pattern is more informative: the lowest quartile (Q1, districts with the least agricultural labor) shows a large positive coefficient of  $0.751$ , while the highest quartile (Q4) shows a negative coefficient of  $-0.072$ . This suggests that MGNREGA's economic impact, as measured by nightlights, was concentrated in districts that were already transitioning away from agriculture—consistent with the program complementing rather than substituting for non-agricultural growth. However, the large Q1 estimate may also reflect small-sample concerns, as districts with very low agricultural labor shares are unusual in rural India.



**Figure 4:** Heterogeneous Treatment Effects by Baseline Characteristics

## 7. Robustness and Identification

### 7.1 Goodman-Bacon Decomposition

The Bacon decomposition (Figure 6 in the Appendix) reveals the source of TWFE bias. “Earlier versus later” comparisons receive 47 percent of the weight and produce positive estimates averaging 0.204, while “later versus earlier” comparisons receive 53 percent and produce negative estimates averaging  $-0.113$ . The negative weights arise because later-treated districts are already-treated when used as controls—the classic staggered DiD problem (Goodman-Bacon, 2021). This decomposition motivates the CS estimator as the preferred specification.

### 7.2 Sensor-Specific Analyses

Table 3 presents robustness checks. Column (1) reports the DMSP-only event study (1994–2013) using Sun-Abraham interaction-weighted TWFE with event time  $-1$  as the omitted category. The large positive pre-treatment coefficients (e.g., 1.64 at  $t = -12$ ) reflect the *within-district* deviation from the omitted period—backward districts that were growing faster than their own event-time-minus-one level show large positive deviations at earlier event times. These coefficients decline monotonically toward zero at the omitted period and turn negative post-treatment, a pattern consistent with convergence: backward districts experienced above-trend growth in the pre-period and subsequently grew more slowly. The DMSP-only results

therefore suggest that the positive CS estimate may partly reflect pre-existing convergence dynamics.

Column (2) reports the VIIRS-only analysis (2012–2023). Because all districts are treated by 2012, I compare Phase I against Phase III districts by interacting a Phase I indicator with years of exposure. The coefficient of 0.043 ( $t = 8.6$ ) indicates that Phase I districts exhibit a steeper positive trend than Phase III districts in the VIIRS era. This differential trend is consistent with earlier MGNREGA exposure generating cumulative advantages, though it cannot be interpreted as a marginal causal effect of one additional year of treatment—since the exposure difference between phases is fixed at approximately two years throughout the post-2012 period. The result operates entirely within the higher-resolution VIIRS sensor and avoids harmonization concerns.

### 7.3 Placebo Tests

The most important diagnostic is the placebo test in Column (3) of [Table 3](#). I shift the treatment dates back five years and estimate the TWFE model using only pre-MGNREGA data (1994–2005). If parallel trends held, this placebo coefficient should be zero. Instead, it is 0.171 ( $p = 0.001$ )—large, positive, and significant. This is strong evidence that the phase assignment predicts differential nightlight trends even in the pre-period, likely because backward districts were on convergence trajectories related to the same characteristics that determined phase assignment.

This placebo failure does not invalidate the analysis, but it substantially qualifies the interpretation. The positive CS estimate of 0.091 should be understood as an upper bound on MGNREGA’s causal effect, with the true effect lying somewhere between zero (if all the estimated effect reflects pre-existing convergence) and 0.091 (if MGNREGA genuinely accelerated the convergence trajectory).

### 7.4 Randomization Inference

Randomization inference ([Figure 7](#) in the Appendix) confirms that the naive TWFE estimate is uninformative: under 500 random permutations of district-phase assignments, the actual coefficient (0.034) falls well within the permutation distribution ( $p_{RI} = 0.378$ ), indistinguishable from chance.

### 7.5 Honest DiD Sensitivity Analysis

The [Rambachan and Roth \(2023\)](#) sensitivity analysis provides a formal framework for assessing how much the treatment effect estimate changes under various degrees of parallel trends

**Table 3:** Robustness Checks

	DMSP Only (1)	VIIRS Only (2)	Placebo (3)	Phase I vs III (4)
event_time = -12	1.638*** (0.3080)			
event_time = -11	1.387*** (0.2470)			
event_time = -10	1.174*** (0.1940)			
event_time = -9	0.9024*** (0.1351)			
event_time = -8	0.7973*** (0.1040)			
event_time = -7	0.6762*** (0.0905)			
event_time = -6	0.5266*** (0.0625)			
event_time = -5	0.4677*** (0.0552)			
event_time = -4	0.3286*** (0.0394)			
event_time = -3	0.2138*** (0.0259)			
event_time = -2	0.0599*** (0.0101)			
event_time = 0	-0.0433*** (0.0089)			
event_time = 1	-0.1580*** (0.0200)			
event_time = 2	-0.1414*** (0.0195)			
event_time = 3	-0.0853*** (0.0166)			
event_time = 4	-0.0798*** (0.0153)			
event_time = 5	-0.0671*** (0.0235)			
phase1 $\times$ years_exposed		0.0429*** (0.0050)		
placebo_treat			0.1707*** (0.0529)	
treat_p1				0.3885*** (0.0761)
Observations	11,410	21	6,120	7,680
R <sup>2</sup>	0.85955		0.97758	0.92732
Within R <sup>2</sup>	0.01908		0.09660	0.00301

violation. I parameterize the maximum change in the slope of the pre-trend ( $M$ ), where  $M = 0$  corresponds to exact parallel trends and larger  $M$  allows for increasingly severe violations.

At  $M = 0$ , the robust confidence interval is  $[0.036, 0.088]$ , excluding zero. At  $M = 0.01$ , the interval is  $[-0.002, 0.070]$ , just barely including zero. At  $M = 0.02$ , the interval is  $[-0.027, 0.065]$ , comfortably including zero. The results are therefore robust to exact parallel trends but fragile to even modest violations. Given the evidence from the placebo test that pre-trends exist, an honest assessment is that the treatment effect is *plausibly positive but not definitively established*.

## 7.6 State-by-Year Fixed Effects

The specification with state-by-year fixed effects (Column (3) of [Table 2](#)) provides an important robustness check. By absorbing all state-level time-varying confounders—including state-specific responses to national policy reforms, monsoon variation, and political cycles—this specification identifies the MGNREGA effect from within-state variation in treatment timing. The resulting estimate of 0.137 is larger than the baseline TWFE, suggesting that cross-state comparisons introduce downward bias, perhaps because Phase III districts in economically dynamic states (e.g., Gujarat, Tamil Nadu) experienced faster growth for reasons unrelated to MGNREGA.

## 8. Discussion

### 8.1 Interpreting the Magnitude

The CS estimate implies that MGNREGA increased district nightlights by approximately 9.5 percent. Using the [Henderson et al. \(2012\)](#) elasticity of GDP with respect to lights (approximately 0.3 in developing countries), this translates to a GDP increase of roughly 3 percent. For Phase I districts, the 15 percent nightlight gain implies roughly 5 percent higher GDP—a substantial return on a program that costs approximately 2–3 percent of rural GDP annually.

However, three caveats apply. First, the GDP-lights elasticity may differ at the tail of the income distribution where MGNREGA districts operate. Second, the parallel trends concern means the true effect may be smaller. Third, nightlights capture a mix of residential, commercial, and public infrastructure lighting; MGNREGA’s contribution to each channel is unclear.

## 8.2 Big Push or Transfer?

The evidence is consistent with a modest “catalytic” interpretation. The Phase I effect of 0.150 that persists over fifteen years of data suggests that early MGNREGA exposure generated lasting changes in local economic structure. The VIIRS-only result—a steeper positive trend for Phase I relative to Phase III districts—indicates that earlier treatment is associated with cumulative advantages, consistent with asset creation or sustained labor market transformation.

However, the results do not strongly support a dramatic “big push” narrative. The effects are moderate in magnitude, sensitive to specification, and concentrated in districts that were already transitioning away from agriculture. A more cautious interpretation is that MGNREGA provided a floor under rural incomes that allowed backward districts to converge somewhat faster, without fundamentally altering their growth trajectory.

This interpretation aligns with [Muralidharan et al. \(2023\)](#), who find in an RCT context that the majority of MGNREGA’s income gains come from non-program earnings—suggesting the program operates through general equilibrium demand channels rather than direct employment. The nightlight results are consistent with this mechanism: MGNREGA injects income that stimulates local non-agricultural activity, which is reflected in nightlights.

## 8.3 Limitations

The primary limitation is the violation of parallel trends. MGNREGA’s targeting on backwardness ensures that treated and control districts differ systematically in ways that predict economic trajectories. While the CS estimator, state-by-year fixed effects, and population weighting all produce positive and significant estimates, the placebo test and Honest DiD analysis caution against strong causal claims.

A second limitation is the absence of disaggregated outcome data. Nightlights capture aggregate economic activity but cannot distinguish between MGNREGA-specific channels (infrastructure assets, labor market tightening, female empowerment) and general equilibrium effects. The heterogeneity results by agricultural labor share are suggestive but cannot definitively identify mechanisms.

Third, the DMSP-to-VIIRS sensor transition introduces measurement uncertainty in the middle of the treatment period. While the VIIRS-only analysis confirms positive effects within a single sensor, the harmonized panel’s treatment effect trajectory around 2012–2013 should be interpreted cautiously.

Fourth, spatial spillovers may bias estimates in either direction. MGNREGA could affect neighboring districts through labor market competition, migration, or demand spillovers

(Conley, 1999). If Phase III districts benefited from spillovers from adjacent Phase I districts, the comparison group is partially treated, attenuating estimated effects. Future work should consider spatial HAC standard errors and border-district exclusion tests.

Finally, I reconstruct the phase assignment from Census 2001 data rather than using the official government notification. While the methodology follows the documented Planning Commission procedure, minor discrepancies with the actual assignment may introduce measurement error, attenuating the estimates. Obtaining the official district notification list and using it directly for phase assignment—or exploiting the backwardness ranking discontinuity in a regression discontinuity design—would substantially strengthen identification. Similarly, synthetic difference-in-differences (Arkhangelsky et al., 2021) or generalized synthetic control methods (Xu, 2017) could better address the differential pre-trends that plague this setting.

## 8.4 External Validity

The results speak most directly to the question of whether backwardness-targeted public works programs can generate lasting economic gains. The Indian context—with its vast rural population, widespread agricultural labor surplus, and democratic accountability structures—may not generalize to settings with different institutional features. The finding that effects are largest under state-by-year fixed effects suggests that within-state policy heterogeneity is important, which may limit portability to countries with less subnational policy variation.

## 9. Conclusion

This paper examines MGNREGA’s impact on local economic activity over a panel spanning fifteen years after initial implementation, using satellite nightlights. The evidence points to moderate, persistent positive effects: the Callaway-Sant’Anna estimator yields an overall ATT of 9.5 percent, Phase I districts show gains of 15 percent relative to Phase III, and the VIIRS-only analysis confirms a steeper positive trend for earlier-treated districts. The Goodman-Bacon decomposition reveals that the standard TWFE estimator is severely biased by staggered treatment contamination, producing a misleading near-zero estimate. Modern heterogeneity-robust methods are essential for credible inference in this setting.

However, the parallel trends assumption—the foundation of all difference-in-differences designs—is strained by MGNREGA’s targeting on backwardness. The placebo test, Honest DiD sensitivity analysis, and randomization inference collectively suggest that while MGNREGA plausibly accelerated convergence in backward districts, the estimated magnitude is sensitive to assumptions about pre-existing trends. This honesty about identification limitations, rather than being a weakness, reflects the reality of evaluating programs that are

deliberately targeted at the most disadvantaged communities.

The policy implication is cautiously optimistic. Fifteen years after implementation, early MGNREGA districts have not reverted to their pre-program trajectories. Whether this persistence reflects the program's infrastructure assets, its labor market effects, or its role in sustaining demand for non-agricultural goods and services remains an open question. What is clear is that the world's largest public works program has left a detectable mark on the night sky of rural India—even if the precise causal magnitude remains a matter of ongoing empirical debate.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). Nightlights and Census data from the SHRUG platform ([Asher et al., 2021](#)). MGNREGA phase assignment reconstructed following [Government of India, Planning Commission \(2003\)](#) and [Zimmermann \(2022\)](#).

**Project Repository:** <https://github.com/SocialCatalystLab/ape-papers>

**Contributors:** @olafdrw

**First Contributor:** <https://github.com/olafdrw>

## References

- Afridi, Farzana, Taryn Dinkelman, and Kanika Mahajan**, “Social Programs, Gender, and Last-Mile Service Delivery: Experimental Evidence from India,” *Journal of Development Economics*, 2023, *160*, 102996.
- Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager**, “Synthetic Difference-in-Differences,” *American Economic Review*, 2021, *111* (12), 4088–4118.
- Asher, Sam and Paul Novosad**, “Rural Roads and Local Economic Development,” *American Economic Review*, 2020, *110* (3), 797–823.
- , **Tobias Lunt, Ryu Matsuura, and Paul Novosad**, “Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform,” *World Bank Economic Review*, 2021, *35* (4), 845–871.
- Berg, Erlend, Sambit Bhattacharyya, D. Rajasekhar, and R. Manjula**, “Can Public Works Increase Equilibrium Wages? Evidence from India’s National Rural Employment Guarantee,” *World Development*, 2018, *103*, 239–254.
- Berman, Eli, Jacob N. Shapiro, and Joseph H. Felter**, “Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, 2011, *119* (4), 766–819.
- Besley, Timothy and Stephen Coate**, “The Design of Income Maintenance Programmes,” *Review of Economic Studies*, 1995, *62* (2), 187–221.
- Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, *225* (2), 200–230.
- Chen, Xi and William D. Nordhaus**, “Using Luminosity Data as a Proxy for Economic Statistics,” *Proceedings of the National Academy of Sciences*, 2011, *108* (21), 8589–8594.
- Conley, Timothy G.**, “GMM Estimation with Cross Sectional Dependence,” *Journal of Econometrics*, 1999, *92* (1), 1–45.
- Cook, C. Justin and Manisha Shah**, “Aggregate Effects from Public Works: Evidence from India,” *Review of Economics and Statistics*, 2022, *104* (4), 797–806.

- de Chaisemartin, Clément and Xavier D’Haultfoeuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, *110* (9), 2964–2996.
- Drèze, Jean and Reetika Khera**, “The Battle for Employment Guarantee,” *Frontline*, 2009, *26* (1), 4–7.
- Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li**, “Which Night Lights Data Should We Use in Economics, and Where?,” *Journal of Development Economics*, 2021, *149*, 102602.
- Goodman-Bacon, Andrew**, “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, 2021, *225* (2), 254–277.
- Government of India, Planning Commission**, “Identification of Districts for Wage and Self Employment Programmes,” Technical Report, Planning Commission, New Delhi 2003.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil**, “Measuring Economic Growth from Outer Space,” *American Economic Review*, 2012, *102* (2), 994–1028.
- Imbert, Clément and John Papp**, “Labor Market Effects of Social Programs: Evidence from India’s Employment Guarantee,” *American Economic Journal: Applied Economics*, 2015, *7* (2), 233–263.
- Klonner, Stefan and Christian Oldiges**, “Safety Net for India’s Poor or Waste of Public Funds? Poverty and Welfare in the Wake of the World’s Largest Job Guarantee Program,” *Journal of Development Economics*, 2022, *157*, 102848.
- Kraay, Aart and David McKenzie**, “Do Poverty Traps Exist? Assessing the Evidence,” *Journal of Economic Perspectives*, 2014, *28* (4), 127–148.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar**, “General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence from India,” *Econometrica*, 2023, *91* (4), 1261–1295.
- Murgai, Rinku, Martin Ravallion, and Dominique van de Walle**, “The National Rural Employment Guarantee Act: An Indian Social Protection Success Story?,” *World Bank Research Observer*, 2016, *31* (2), 253–286.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

- Ravallion, Martin**, “Targeted Transfers in Poor Countries: Revisiting the Tradeoffs and Policy Options,” *World Bank Research Observer*, 2003, 18 (2), 165–187.
- Roth, Jonathan**, “Pre-test with Caution: Event-Study Estimates after Testing for Parallel Trends,” *American Economic Review: Insights*, 2022, 4 (3), 305–322.
- Sant’Anna, Pedro H. C. and Jun Zhao**, “Doubly Robust Difference-in-Differences Estimators,” *Journal of Econometrics*, 2020, 219 (1), 101–122.
- Subbarao, Kalanidhi**, “Public Works as an Anti-Poverty Program: An Overview of Cross-Country Experience,” *American Journal of Agricultural Economics*, 1997, 79 (2), 678–683.
- Sun, Liyang and Sarah Abraham**, “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects,” *Journal of Econometrics*, 2021, 225 (2), 175–199.
- Xu, Yiqing**, “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models,” *Political Analysis*, 2017, 25 (1), 57–76.
- Zimmermann, Laura**, “The Welfare Effects of India’s Rural Employment Guarantee,” *Journal of Development Economics*, 2022, 157, 102848.

## A. Data Appendix

### A.1 Data Sources and Access

The primary data source is the SHRUG platform, version 2.1 “Pakora” (Asher et al., 2021), available at [https://www.devdata.org/shrug\\_download/](https://www.devdata.org/shrug_download/) under an Open Data License requiring attribution. The following tables were downloaded:

- **Census PCA 2001** (`shrug_pca01`): Village-level population, SC/ST population, literacy, and workforce composition for Census 2001.
- **Census PCA 2011** (`shrug_pca11`): Same variables for Census 2011.
- **Census PCA 1991** (`shrug_pca91`): Same variables for Census 1991.
- **DMSP nightlights** (`shrug_dmisp`): Calibrated annual total luminosity for each SHRID, 1994–2013.
- **VIIRS nightlights** (`shrug_viirs`): Annual sum luminosity for each SHRID, 2012–2023.
- **Geographic crosswalk** (`shrug_pc_keys`, `shrug_td11`): SHRID to Census 2011 state and district codes.

### A.2 Panel Construction

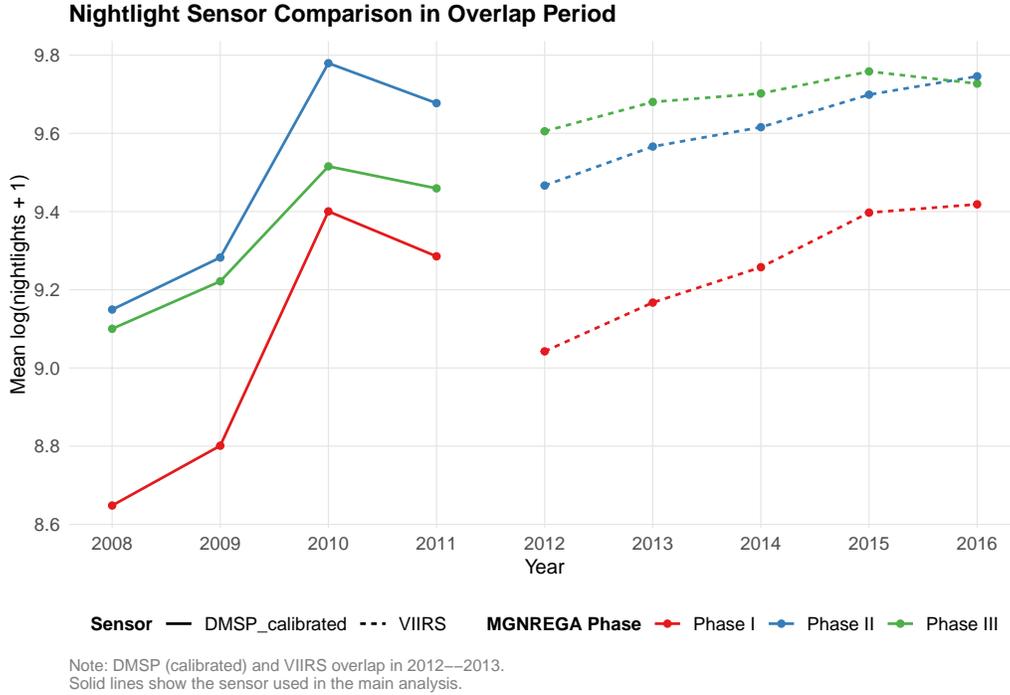
Village-level nightlights were aggregated to the district level by summing calibrated luminosity across all villages within each Census 2011 district boundary. The resulting panel has 640 districts observed for 30 years (1994–2023), yielding 19,200 district-year observations. District boundaries are held constant at their 2011 definitions throughout the analysis.

### A.3 DMSP-VIIRS Harmonization

The calibration regression uses 1,280 district-year observations from the 2012–2013 overlap period:

$$\log(\text{VIIRS}_{dt} + 1) = -4.45 + 1.29 \times \log(\text{DMSP}_{dt} + 1) + \varepsilon_{dt}$$

with  $R^2 = 0.66$ . DMSP values for 1994–2011 are converted to VIIRS-equivalent units using these coefficients. Figure 5 shows the two series in the overlap period, confirming reasonable alignment.



**Figure 5:** Sensor Comparison in the 2012–2013 Overlap Period

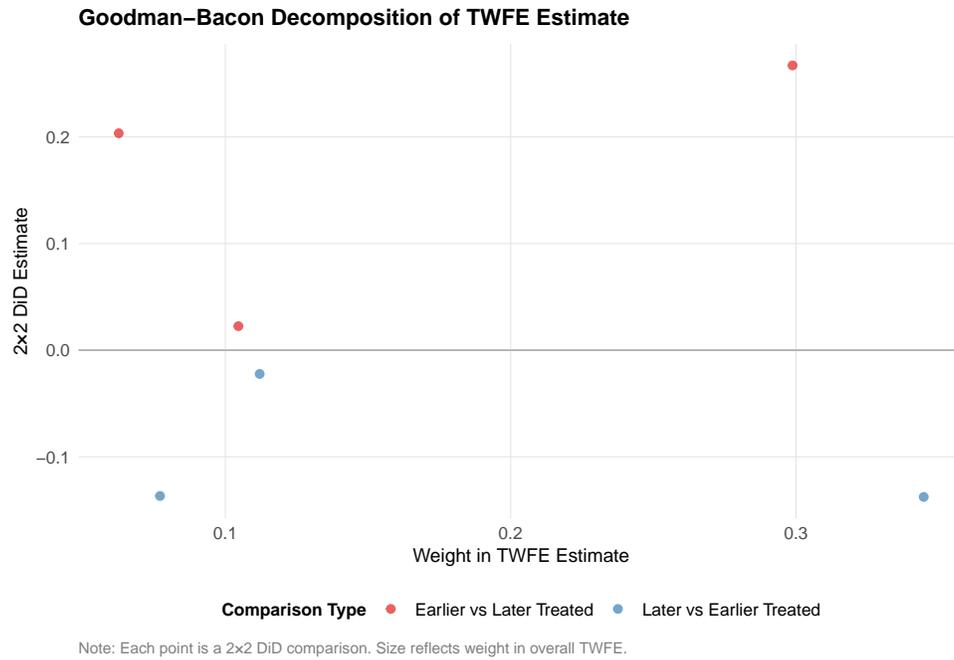
#### A.4 Phase Assignment Construction

The backwardness index is computed at the district level from Census 2001 data. For each district  $d$ , the composite index is:

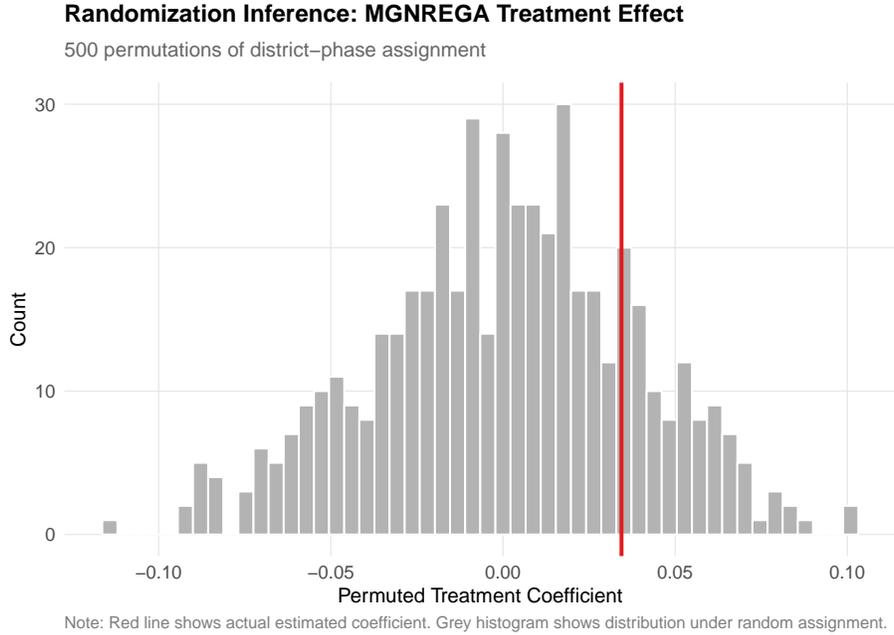
$$\text{Backwardness}_d = \frac{1}{3} [\text{rank}(\text{SC/ST share}_d) + \text{rank}(\text{Ag labor share}_d) + \text{rank}(\text{Illiteracy}_d)]$$

where ranks are computed nationally (highest rank = most backward). Districts are assigned to Phase I (ranks 1–200), Phase II (ranks 201–330), and Phase III (ranks 331–640). This proxy index uses Census 2001 variables (SC/ST share, agricultural labor share, illiteracy rate) rather than the original Planning Commission inputs (SC/ST share from 1991, agricultural wages from 1996–1997, and agricultural output per worker from 1990–1993), which are no longer publicly available at the district level. The proxy and official assignments are highly correlated because both capture the same underlying dimension of economic backwardness, though minor discrepancies at the phase boundaries are possible.

## B. Diagnostic Figures



**Figure 6:** Goodman-Bacon Decomposition of TWFE Estimate. Each point represents a two-by-two DiD comparison; point size reflects weight in the overall TWFE estimate. “Later vs Earlier” comparisons receive 53% of the weight and produce negative estimates, contaminating the overall TWFE coefficient.

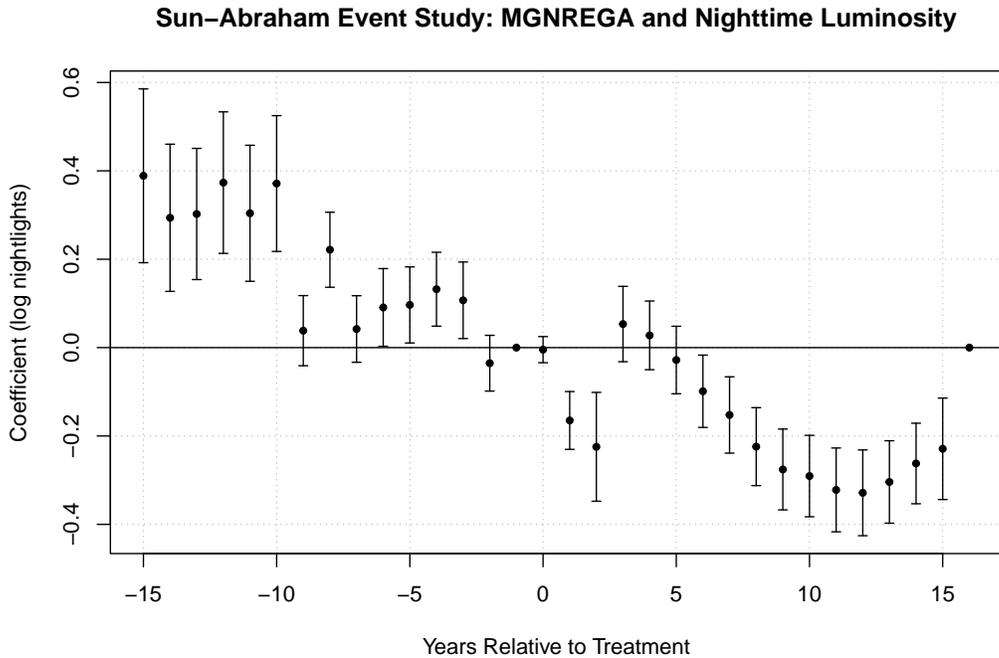


**Figure 7:** Randomization Inference: Distribution of Permuted TWFE Coefficients. The vertical line marks the actual estimate (0.034). Under 500 random permutations of phase assignment,  $p_{RI} = 0.378$ .

## C. Identification Appendix

### C.1 Pre-Trend Analysis

The Sun-Abraham event study (Figure 8) shows the full set of pre-treatment and post-treatment coefficients from the interaction-weighted TWFE. Pre-treatment coefficients at event times  $-15$  through  $-2$  are predominantly positive and significant, reflecting the systematic differences in nightlight trajectories between backward (early-treated) and less backward (late-treated) districts. The monotonic decline of pre-treatment coefficients toward zero at event time  $-1$  is consistent with convergence dynamics rather than a sharp pre-treatment shock.



**Figure 8:** Sun-Abraham Event Study

Post-treatment coefficients become negative from event time 1 onward in the Sun-Abraham framework, reflecting the contamination from “later versus earlier” comparisons documented in the Bacon decomposition. The CS estimator (Figure 1), which uses only clean comparisons, shows positive post-treatment effects.

## C.2 Honest DiD Details

The [Rambachan and Roth \(2023\)](#) sensitivity analysis is based on the event-study specification with five pre-treatment and fifteen post-treatment periods. The breakdown frontier—the value of  $M$  at which the confidence interval first includes zero—is approximately 0.01. This means that if the pre-trend violation is such that the slope of the counterfactual trend could change by more than 0.01 per period, the treatment effect is no longer distinguishable from zero at the 95 percent level.

## D. Robustness Appendix

### D.1 Alternative Specifications

The main result is robust to several alternative specifications:

1. **State-by-year fixed effects** ( $\hat{\beta} = 0.137, p < 0.001$ ): Absorbing state-level shocks strengthens the estimate.
2. **Population weighting** ( $\hat{\beta} = 0.192, p < 0.001$ ): Larger districts show stronger effects.
3. **VIIRS differential trend** ( $\hat{\beta} = 0.043, p < 0.001$ ): Within the VIIRS era (2012–2023), Phase I districts exhibit a steeper positive trend than Phase III districts, consistent with cumulative benefits of earlier treatment.

The estimate is sensitive to:

1. **District-specific trends** ( $\hat{\beta} = 0.057, p = 0.14$ ): Absorbing district trends reduces the estimate and eliminates significance.
2. **Placebo test** ( $\hat{\beta}^{placebo} = 0.171, p = 0.001$ ): The phase assignment predicts pre-MGNREGA trends.

## D.2 Clustering

All standard errors are clustered at the district level (640 clusters). With three treatment cohorts and 640 clusters, the few-clusters problem is not a concern for inference.

## E. Heterogeneity Appendix

Table 4 reports TWFE estimates for subsamples defined by quartiles of baseline characteristics. The most notable pattern is the large positive coefficient for the lowest quartile of agricultural labor share (0.751). This suggests that MGNREGA’s nightlight impact was concentrated in districts that were already transitioning out of agriculture, where the injection of wage income may have catalyzed non-farm enterprise growth. Districts in the highest agricultural labor quartile show small negative effects (−0.072), possibly because MGNREGA work substituted for private agricultural labor in these settings without generating additional economic activity visible in nightlights.

Literacy quartiles show no clear monotonic pattern, suggesting that human capital did not strongly mediate MGNREGA’s effects on aggregate nightlights.

**Table 4:** Heterogeneous Treatment Effects by Baseline Characteristics

<i>Panel A: SC/ST Population Share</i>				
	Q1 (Lowest) (1)	Q2 (2)	Q3 (3)	Q4 (Highest) (4)
treated	-0.0790 (0.0579)	-0.1330** (0.0579)	-0.1262** (0.0488)	-0.1449 (0.1087)
Observations	4,800	4,800	4,800	4,800
Within R <sup>2</sup>	0.00013	0.00057	0.00152	0.00046
<i>Panel B: Agricultural Labor Share</i>				
	Q1 (Lowest) (5)	Q2 (6)	Q3 (7)	Q4 (Highest) (8)
treated	0.7513*** (0.2044)	0.0331 (0.0548)	-0.0052 (0.0379)	-0.0718* (0.0367)
Observations	4,800	4,800	4,800	4,800
Within R <sup>2</sup>	0.00647	$4.08 \times 10^{-5}$	$2.18 \times 10^{-6}$	0.00022
<i>Panel C: Literacy Rate</i>				
	Q1 (Lowest) (9)	Q2 (10)	Q3 (11)	Q4 (Highest) (12)
treated	-0.0023 (0.0935)	0.0474 (0.0434)	-0.0089 (0.0456)	0.0516 (0.0874)
Observations	4,800	4,800	4,800	4,800
Within R <sup>2</sup>	$8.37 \times 10^{-8}$	0.00011	$3.86 \times 10^{-6}$	$2.23 \times 10^{-5}$

*Notes:* Each column estimates TWFE on a subsample defined by quartiles of the baseline (Census 2001) characteristic. Q1 = lowest quartile, Q4 = highest. All regressions include district and year fixed effects. Standard errors clustered at the district level in parentheses. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .