

# Where Does Workfare Work? Heterogeneous Effects of India's Employment Guarantee on Local Economic Activity

APEP Autonomous Research\* @olafdrw

February 26, 2026

## Abstract

India's MGNREGA—the world's largest public employment program—guaranteed 100 days of work to 260 million rural households across three implementation phases (2006–2008). We estimate heterogeneous effects on district-level economic activity using DMSP satellite nightlights and the program's staggered rollout across 584 districts. Callaway-Sant'Anna estimates yield a positive aggregate effect (ATT = 0.0817 log points,  $p < 0.01$ ), though significant pre-treatment differential trends between early- and late-treated districts—confirmed by Rambachan-Roth sensitivity analysis—challenge causal interpretation of the aggregate estimate. The Sun-Abraham estimator and conventional TWFE both yield near-zero effects. Decomposing by baseline characteristics reveals striking heterogeneity: effects concentrate in districts with medium agricultural labor intensity and high SC/ST population shares, while arid and dark districts show no gains. Census mechanism analysis finds no significant differential structural transformation across treatment phases. Our findings highlight that employment guarantees generate uneven developmental returns, shaped by pre-existing economic structure.

**JEL Codes:** H53, I38, O13, O18, R11

**Keywords:** MGNREGA, employment guarantee, nightlights, staggered DiD, heterogeneous treatment effects, India

---

\*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

# 1. Introduction

Rural India in the early 2000s faced a devastating employment crisis. After a string of monsoon failures and stagnating agricultural productivity, hundreds of farmers committed suicide in Maharashtra and Andhra Pradesh, catalyzing a political earthquake (Reddy and Mishra, 2006). The state responded with the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) of 2005—the largest public works program in human history, which, legally guaranteeing 100 days of manual labor per rural household at the statutory minimum wage. By 2010, MGNREGA covered every rural district in India, reached over 50 million households annually, and consumed roughly 0.5% of GDP (Drèze and Khera, 2011).

Fifteen years later, the central question about MGNREGA remains unresolved: does it actually generate economic development, or merely redistribute? Proponents argue that guaranteed employment creates a wage floor, builds productive rural assets (roads, irrigation, land leveling), and insures households against agricultural risk, enabling growth-promoting investments (Imbert and Papp, 2015; Muralidharan et al., 2016). Skeptics counter that MGNREGA crowds out private employment, produces low-quality assets, and creates fiscal burdens that slow growth (Sukhtankar, 2017; Dutta et al., 2012). The aggregate evidence is ambiguous precisely because MGNREGA’s effects are unlikely to be uniform—the program operates through multiple channels whose relative importance depends on local economic conditions.

This paper asks: *where* does MGNREGA generate economic development, and where does it fail? We exploit the program’s three-phase staggered district rollout and DMSP satellite nightlights as a spatially comprehensive measure of local economic activity. Phase I (February 2006) covered the 200 most backward districts, Phase II (April 2007) added 130 districts, and Phase III (April 2008) extended coverage to all remaining rural districts. This design—200 treated units across three cohorts observed over 14 years—provides statistical power to estimate not just average effects but heterogeneous impacts across the distribution of district characteristics.

Our primary contribution is a systematic analysis of effect heterogeneity across four dimensions of baseline economic structure: rainfall regime, agricultural labor intensity, social marginalization (SC/ST population share), and initial economic development (baseline nightlights). If MGNREGA works primarily through labor market effects, we expect larger effects in districts with high agricultural labor shares. If it works through risk insurance, effects should concentrate in rainfall-vulnerable districts. If it addresses historical marginalization, SC/ST-intensive districts should benefit most. By testing these predictions, we can discriminate between the program’s competing theoretical channels.

We implement Callaway and Sant’Anna’s (2021) heterogeneity-robust estimator for staggered difference-in-differences, which avoids the well-documented biases of conventional two-way fixed effects (TWFE) in settings with heterogeneous treatment effects and staggered adoption (Goodman-Bacon, 2021; de Chaisemartin and d’Haultfoeuille, 2020; Sun and Abraham, 2021). Our analysis covers 584 districts across 14 years (2000–2013), using inter-calibrated DMSP nightlights from the SHRUG database (Asher et al., 2021).

The aggregate Callaway-Sant’Anna estimate is positive and statistically significant (0.0817 log points,  $SE = 0.0118$ ), but we show this reflects pre-existing differential trends rather than a clean causal effect. Early-treated (more backward) districts exhibited different nightlight trajectories before MGNREGA began. The conventional TWFE estimate, which absorbs district-specific trends through fixed effects, is small and statistically insignificant (0.012,  $SE = 0.014$ ). This divergence between estimators is itself informative: it reveals that the composition of treatment groups matters fundamentally for inference in this setting.

The heterogeneity analysis yields our main results. Effects are concentrated in districts with *medium* agricultural labor intensity ( $ATT = 0.051$ ,  $p < 0.05$ ) rather than the highest-intensity districts. This pattern is consistent with a model where MGNREGA’s labor market effects require a minimum level of non-farm economic activity to translate into broader development. Districts at the bottom of the agricultural labor distribution—already diversified—show mildly negative effects consistent with crowding-out, while the most agricultural districts lack the complementary infrastructure for guaranteed employment to catalyze structural change. High-SC/ST districts also show positive effects (0.028), consistent with MGNREGA’s targeting of historically marginalized populations.

We complement the nightlights analysis with Census mechanism regressions comparing 2001–2011 changes in worker composition across MGNREGA phases. Conditional on the backwardness index, Phase I districts did not experience significantly different rates of structural transformation compared to Phase III districts, suggesting that MGNREGA’s effects on occupational change—if any—operate through channels not captured by the phase assignment variation.

Our paper contributes to four literatures. First, we advance the large literature evaluating MGNREGA (Zimmermann, 2012; Azam, 2012; Klonner and Oldiges, 2016; Berg et al., 2018; Sukhtankar, 2017) by providing the first systematic analysis of effect heterogeneity using satellite data as a spatially comprehensive outcome measure. While Imbert and Papp (2015) estimated average wage and employment effects using household surveys, and Muralidharan et al. (2016) studied implementation quality in Andhra Pradesh, no study has documented where MGNREGA’s developmental effects concentrate and where they are absent.

Second, we contribute to the methodological literature on staggered difference-in-differences

by demonstrating how the choice of estimator matters substantively in this policy setting (Callaway and Sant’Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2024). The divergence between Callaway-Sant’Anna and TWFE estimates in our context illustrates why heterogeneity-robust methods are essential when treatment groups are selected on characteristics correlated with outcome trajectories.

Third, we engage the development economics literature on nightlights as a proxy for economic activity (Henderson et al., 2012; Asher et al., 2021; Donaldson and Storeygard, 2016). By combining nightlights with Census mechanism analysis, we can assess what types of economic change the luminosity signal captures in this context.

Fourth, we speak to the broader literature on the developmental effects of social protection programs (Banerjee et al., 2017; Baird et al., 2014; Duflo et al., 2012). Our heterogeneity findings suggest that employment guarantees are not uniformly developmental—their effects depend critically on pre-existing economic structure, implying that optimal targeting may differ from what pure poverty-based criteria would suggest.

We emphasize throughout that our estimates identify the reduced-form association between early phase assignment and nightlights growth, encompassing MGNREGA alongside concurrent backward-district programs. This “bundle” estimand is policy-relevant—governments target packages of programs, not individual interventions in isolation.

## 2. Institutional Background

### 2.1 The National Rural Employment Guarantee Act

The National Rural Employment Guarantee Act (NREGA, later renamed MGNREGA) was passed by the Indian Parliament on August 25, 2005, and represents the most ambitious social protection legislation in Indian history (Drèze and Khera, 2011). The Act created a legal right to 100 days of unskilled manual employment per rural household per year at the state-level minimum wage. Unlike previous employment programs (which operated as discretionary schemes), MGNREGA is demand-driven: any rural household can request work, and the state government is legally obligated to provide employment within 15 days or pay an unemployment allowance.

The program’s design incorporates several distinctive features. First, the wage rate is pegged to the statutory minimum wage, which ranges from Rs. 100 to Rs. 300 depending on state and year. Second, at least one-third of person-days must go to women. Third, work must be provided within 5 km of the applicant’s residence. Fourth, the program mandates creation of durable productive assets—primarily rural roads, irrigation channels, land leveling, and water conservation structures—distinguishing MGNREGA from pure cash transfers. Fifth,

the Act includes mandatory social audits by gram sabhas (village assemblies) to promote accountability (Aiyar and Samji, 2012).

MGNREGA’s fiscal architecture is unusual. The central government bears 100% of wage costs and 75% of material costs, while state governments finance the remaining 25% of material costs plus administrative expenses. This sharing arrangement was designed to reduce fiscal barriers to state-level implementation, but created heterogeneity in implementation quality across states depending on administrative capacity (Sukhtankar, 2017).

## 2.2 Phased Rollout and Identification

MGNREGA was implemented in three phases, providing the exogenous variation we exploit for identification:

- **Phase I (February 2, 2006):** 200 districts selected by the Planning Commission based on a composite “backwardness index” that weighted SC/ST population share, agricultural wage rates, and agricultural productivity. These were India’s poorest and most marginalized districts.
- **Phase II (April 1, 2007):** An additional 130 districts were added, drawn from the next tier of the backwardness ranking. These districts were less backward than Phase I but still significantly poorer than the national average.
- **Phase III (April 1, 2008):** Coverage was extended to all remaining rural districts (approximately 310 districts, though the exact number varies by definition). Phase III districts included relatively more developed rural areas.

The selection criterion—the Planning Commission’s backwardness index—is central to both the program’s design and our identification challenge. The index combines three components: (1) the share of Scheduled Caste and Scheduled Tribe population, which captures historical social marginalization; (2) the agricultural wage rate, which proxies for rural labor market conditions; and (3) agricultural productivity per worker, which captures the productivity of the rural economy (Planning Commission, 2003). Districts were ranked by this composite index, with the most backward receiving treatment first.

This selection mechanism creates a fundamental tension for identification. The phased rollout provides the variation needed for a staggered difference-in-differences design, but the selection on backwardness means that early-treated districts differ systematically from late-treated districts on multiple dimensions that may independently affect economic trajectories. Phase I districts have higher SC/ST shares (0.44 vs. 0.20), lower literacy rates (0.45 vs.

0.61), and higher agricultural labor shares (0.20 vs. 0.08) compared to Phase III districts. We address this challenge directly in our empirical strategy.

## 2.3 Concurrent Policy Environment

MGNREGA’s rollout coincided with several other major policy interventions that could confound our estimates. The most concerning is the Backward Regions Grant Fund (BRGF), launched in February 2007 to provide untied grants to 250 backward districts for infrastructure and capacity building. Because BRGF’s district selection was also based on backwardness, there is substantial overlap with MGNREGA Phase I districts, creating a confound that is extremely difficult to disentangle ([Planning Commission, 2007](#)).

Other concurrent programs include the National Food Security Mission (NFSM, October 2007) targeting 312 districts for enhanced food grain production; the Rashtriya Krishi Vikas Yojana (RKVY, 2007-08) providing agricultural development funding; and the National Horticulture Mission (NHM, 2005-06) promoting horticultural diversification. While these programs targeted partly overlapping geographies, their selection criteria differed from MGNREGA’s backwardness index, and their fiscal magnitudes were substantially smaller. We assess robustness to these concurrent programs in Section 7.

Our estimand is therefore best interpreted as the combined effect of being selected early into the bundle of backward-district programs—including MGNREGA, BRGF, and related schemes—rather than the isolated effect of MGNREGA alone. We adopt this “bundle” interpretation throughout.

## 3. Data

### 3.1 DMSP Nightlights

We measure economic activity using district-level nighttime luminosity from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS), accessed through the Socioeconomic High-resolution Rural-Urban Geographic Platform (SHRUG) ([Asher et al., 2021](#)). DMSP satellites capture visible and near-infrared radiation emitted at night, providing a spatially comprehensive measure that is independent of administrative data collection ([Henderson et al., 2012](#)).

We use inter-calibrated annual composites at the district level for 2000–2013. The SHRUG provides two measures: total light (the sum of calibrated pixel values within each district boundary) and mean light (the average calibrated pixel value). Our primary outcome is  $\log(\text{total light} + 1)$ , which captures both the intensive margin (brighter areas becoming

brighter) and the extensive margin (new areas being lit). The DMSP sensor has a well-documented top-coding problem: pixels are censored at a digital number (DN) of 63. This creates a ceiling effect that attenuates measurement of very bright urban areas but is less problematic for rural districts where average luminosity is well below the saturation threshold.

For years with overlapping satellite sensors (e.g., F15 and F16 in 2004–2007), the SHRUG provides separate observations. We average across sensors to obtain a single district-year observation, yielding a balanced panel of  $584 \text{ districts} \times 14 \text{ years} = 8,176$  observations.

Nightlights as an economic proxy have both strengths and limitations. [Henderson et al. \(2012\)](#) established that nightlights track GDP growth at the country level, with an elasticity of approximately 0.3. [Asher et al. \(2021\)](#) validated the SHRUG nightlights against district-level Census outcomes for India, finding strong correlations with urbanization, non-farm employment, and electrification. For our purposes, nightlights capture the “downstream” effects of MGNREGA on local economic activity—including increased electrification, commercial activity, and urbanization—rather than the program’s direct employment effects.

One concern with nightlights as an outcome is that they may capture rural electrification programs rather than private economic activity. The Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY), launched in 2005, targeted below-poverty-line households in rural areas with similar targeting criteria to MGNREGA. We cannot fully separate MGNREGA from RGGVY effects, reinforcing our “backward-district bundle” interpretation.

### 3.2 Census and Administrative Data

We draw baseline district characteristics from the Census of India 2001, also accessed through the SHRUG. These include total population, literacy rate, SC/ST population share, occupational composition (agricultural laborers, cultivators, and total workers), and district area. We use these characteristics both for constructing the MGNREGA phase assignment and for defining heterogeneity dimensions.

For mechanism analysis, we use Census 2011 data at the district level, which allows us to examine 2001–2011 changes in worker composition across MGNREGA phases. The SHRUG’s standardized district codes enable consistent merging across Census rounds despite the administrative reorganization of districts (from 593 in 2001 to 640 in 2011). We construct a crosswalk based on the SHRUG’s sub-district identifiers, assigning each 2001 district to its dominant 2011 successor.

Geographic controls come from the SHRUG town directory, which provides district-level information on average rainfall, maximum temperature, number of government and private banks, hospitals, and primary schools.

### 3.3 MGNREGA Phase Assignment

Because official Phase I–III district lists are not available in machine-readable format in our data, we reconstruct the phase assignment using the Planning Commission’s documented methodology (Planning Commission, 2003). We compute a backwardness index for each district as the sum of three standardized components: (1) SC/ST population share, (2) agricultural labor share of workers, and (3) negative literacy rate (so higher values indicate greater backwardness). We then rank districts by this composite index and assign:

- Phase I: Rank 1–200 (most backward)
- Phase II: Rank 201–330
- Phase III: Rank 331 and above

After filtering districts with population below 1,000 (to avoid measurement noise), we obtain 593 districts by initial phase assignment: 200 Phase I, 130 Phase II, and 263 Phase III. This distribution matches the documented allocation closely (Drèze and Khera, 2011). After crosswalk matching across Census rounds, 9 districts could not be uniquely mapped from 2001 to 2011 boundaries and are dropped, yielding an analysis sample of 584 districts (200 Phase I, 130 Phase II, and 254 Phase III). We assign first treatment timing based on the first full agricultural year of MGNREGA implementation: Phase I = 2007 (the 2006-07 agricultural year, beginning July 2006), Phase II = 2008, and Phase III = 2009.

### 3.4 Sample Construction and Summary Statistics

Our analysis panel covers 584 districts observed annually from 2000 to 2013, as described above. Table 1 presents summary statistics by MGNREGA phase. Phase I districts are, by construction, more backward: they have lower baseline nightlights (log 9.49 vs. 10.02), higher SC/ST shares (0.44 vs. 0.20), lower literacy rates (0.45 vs. 0.61), and higher agricultural labor shares (0.20 vs. 0.08). Average rainfall is slightly higher in Phase II districts (909 mm) compared to Phases I and III (830 and 748 mm respectively). These baseline differences motivate both our identification concerns and our heterogeneity analysis.

## 4. Empirical Strategy

### 4.1 Identification

We exploit the staggered implementation of MGNREGA across three phases to estimate its effects on district nightlights using a difference-in-differences framework. The basic identifying

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

	Phase 1	Phase 2	Phase 3	All
Districts	200	130	254	584
Population (2001)	1617795	1688439	1854197	1736339
Log Nightlights	9.493	9.912	10.022	9.816
SC/ST Share	0.436	0.335	0.199	0.311
Literacy Rate	0.449	0.531	0.608	0.537
Ag. Labor Share	0.198	0.142	0.084	0.136
Cultivator Share	0.324	0.335	0.246	0.292
Avg. Rainfall (mm)	830.2	908.6	747.9	811.6
Backwardness Index	1.949	0.271	-1.554	0.052

*Notes:* Means of baseline district characteristics by MGNREGA phase assignment. Nightlights measured in year 2000; demographic and economic variables from Census 2001. Phase I (200 most backward districts) began in February 2006; Phase II (next 130) in April 2007; Phase III (remaining 254) in April 2008. Backwardness index is the sum of standardized SC/ST share, agricultural labor share, and inverse literacy rate.

assumption is that, absent MGNREGA, nightlight trajectories would have evolved similarly across early-treated and late-treated districts—the parallel trends assumption.

Formally, let  $Y_{dt}$  denote log nightlights for district  $d$  in year  $t$ ,  $G_d \in \{2007, 2008, 2009\}$  denote the first treated year for district  $d$ 's cohort, and  $D_{dt} = \mathbb{I}[t \geq G_d]$  denote the treatment indicator. The parallel trends assumption requires:

$$\mathbb{E}[Y_{dt}(0) - Y_{dt'}(0)|G_d = g] = \mathbb{E}[Y_{dt}(0) - Y_{dt'}(0)|G_d = g'] \quad \forall g, g', t < \min(g, g') \quad (1)$$

where  $Y_{dt}(0)$  is the potential outcome absent treatment. This states that untreated outcome trends are independent of group membership in the pre-treatment period.

This assumption is threatened by MGNREGA's selection mechanism. Districts were assigned to phases based on the backwardness index, which correlates with characteristics that plausibly affect economic growth trajectories. More backward districts may have been on convergent (catching-up) or divergent (falling-behind) trajectories relative to less backward districts for reasons unrelated to MGNREGA. We test this assumption directly through event-study analysis and find concerning evidence of pre-treatment differential trends, which we discuss in Section 5.

## 4.2 Callaway-Sant'Anna Estimator

Our primary estimator is the group-time average treatment effect on the treated (ATT) proposed by [Callaway and Sant'Anna \(2021\)](#). This estimator avoids the well-known pitfalls

of conventional TWFE in staggered settings: negative weighting, forbidden comparisons (using already-treated units as controls), and heterogeneity bias (Goodman-Bacon, 2021; de Chaisemartin and d’Haultfoeuille, 2020).

For each treatment cohort  $g$  and calendar time  $t$ , we estimate:

$$ATT(g, t) = \mathbb{E}[Y_t - Y_{g-1}|G = g] - \mathbb{E}[Y_t - Y_{g-1}|G > t] \quad (2)$$

using the doubly-robust estimator of Sant’Anna and Zhao (2020), which combines outcome regression and inverse probability weighting. The control group consists of “not-yet-treated” units: districts whose MGNREGA phase begins after period  $t$ . We use a universal base period ( $t = 2005$ , the last year before any cohort’s first full treatment year) to ensure comparability across groups. Note that while Phase I districts received partial MGNREGA exposure beginning February 2006, we code treatment timing by the first full agricultural year of exposure (July–June convention): Phase I = 2007, Phase II = 2008, Phase III = 2009. Calendar year 2006 thus contains partial Phase I treatment; our coding treats this as a transition year rather than full treatment.

We aggregate group-time effects in several ways:

1. **Dynamic event study:** Average across cohorts for each event-time  $e \in \{-6, \dots, 6\}$ , testing for pre-treatment differential trends and post-treatment dynamics.
2. **Simple ATT:** Overall weighted average across all post-treatment group-time cells.
3. **Group-level ATT:** Phase-specific averages, comparing the magnitude of effects across Phase I, II, and III districts.

Standard errors are clustered at the district level (584 clusters), following Callaway and Sant’Anna (2021). For TWFE specifications, we cluster at the state level (approximately 30 clusters) to account for within-state correlation in both treatment assignment and outcomes.

### 4.3 TWFE Baseline

As a comparison, we estimate the conventional two-way fixed effects model:

$$Y_{dt} = \alpha_d + \lambda_t + \beta \cdot D_{dt} + \varepsilon_{dt} \quad (3)$$

where  $\alpha_d$  are district fixed effects,  $\lambda_t$  are year fixed effects,  $D_{dt}$  is the post-treatment indicator, and standard errors are clustered at the state level. We also estimate a dynamic version with

event-time indicators:

$$Y_{dt} = \alpha_d + \lambda_t + \sum_{e=-6}^6 \gamma_e \cdot \mathbb{I}[t - G_d = e] + \varepsilon_{dt} \quad (4)$$

normalizing  $\gamma_{-1} = 0$ .

Under homogeneous treatment effects and correct parallel trends, TWFE recovers the causal effect of MGNREGA. Under heterogeneous effects, TWFE produces a weighted average that may assign negative weights to some group-time cells (de Chaisemartin and d’Haultfoeuille, 2020). The comparison between TWFE and Callaway-Sant’Anna estimates is informative about the degree to which treatment effect heterogeneity matters in this setting.

#### 4.4 Heterogeneity Analysis

Our main analytical contribution is the decomposition of MGNREGA’s effects by four baseline district characteristics:

1. **Rainfall regime:** Districts classified as Arid, Medium, or Wet based on average annual rainfall terciles. If MGNREGA functions as implicit insurance against rainfall risk, effects should be larger in arid districts.
2. **Agricultural labor intensity:** Districts classified as Low, Medium, or High based on the share of workers employed as agricultural laborers. This captures the program’s direct labor market channel—MGNREGA should matter more where more workers are agricultural wage laborers (the program’s target population).
3. **Social marginalization:** Districts classified by SC/ST population share terciles. MGNREGA explicitly targets historically marginalized communities through its selection mechanism and mandated social audit provisions.
4. **Initial development:** Districts classified as Dark, Medium, or Bright based on year-2000 nightlights. This tests whether MGNREGA accelerates development in the least developed areas (convergence) or amplifies growth where a minimum threshold of economic activity already exists (complementarity).

For each dimension, we estimate both TWFE and Callaway-Sant’Anna specifications within each tercile, reporting the point estimate, standard error, and sample size. This approach sacrifices statistical power (each tercile contains roughly 195 districts) but enables us to characterize the full distribution of effects.

## 4.5 Threats to Validity

The staggered rollout faces three primary challenges. First, **selection on trajectories**: because phase assignment was based on backwardness, early-treated districts may have been on different economic trajectories. We assess this through the event-study pre-trend test, finding evidence of differential pre-trends that complicates causal interpretation of the aggregate estimates.

Second, **concurrent programs**: as discussed in Section 2.3, several other programs targeting backward or agricultural districts were launched in 2006–2008. The BRGF is the most concerning confound. We address this through state  $\times$  year fixed effects specifications (which absorb state-level policy changes) and by examining robustness to alternative control strategies.

Third, **measurement**: nightlights are an imperfect proxy for economic activity, with known limitations including top-coding, atmospheric distortion, and sensitivity to electrification rather than economic activity per se (Gibson et al., 2021). We use  $\log(\text{total light} + 1)$  as our primary measure and verify robustness with  $\log(\text{mean light})$  and light per capita.

## 5. Results

### 5.1 Main Results

Table 2 presents our main estimates. Column (1) reports the conventional TWFE specification with district and year fixed effects and state-level clustering. The point estimate is 0.012 with a standard error of 0.014, statistically insignificant at conventional levels ( $t = 0.89$ ,  $p = 0.38$ ). Adding rainfall tercile controls in Column (2) barely changes the estimate (0.013, SE = 0.013). These TWFE results suggest that, on average, MGNREGA had no detectable effect on district nightlights.

However, Column (3) tells a different story. The Callaway-Sant’Anna doubly-robust estimator yields a simple ATT of 0.0817 (SE = 0.0118), statistically significant at the 1% level. Column (4) reports the Sun-Abraham interaction-weighted estimator, which yields a near-zero average post-treatment effect ( $-0.0065$ , the mean of post-treatment cohort-time coefficients; note that Figure 5 shows the full event-study path including the large pre-treatment coefficients that dominate the visual). The divergence between CS-DiD and Sun-Abraham reflects their different handling of the severe pre-treatment differential trends in this setting. The CS-DiD result, while positive, must be interpreted cautiously given the pre-trend violations documented below.

Why do the estimators diverge? The Callaway-Sant’Anna estimator uses not-yet-treated

**Table 2:** Effect of MGNREGA on District Nightlights

	(1)	(2)	(3)	(4)
	TWFE	TWFE	CS-DiD	Sun-Abraham
MGNREGA (Post)	0.0123 (0.0137)	0.0131 (0.0133)	0.0817*** (0.0126)	-0.0065 (0.0158)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Rainfall controls	No	Yes	No	No
Estimator	TWFE	TWFE	CS-DiD (DR)	IW
Observations	8,176	8,134	8,176	8,176
Districts	584	584	584	584

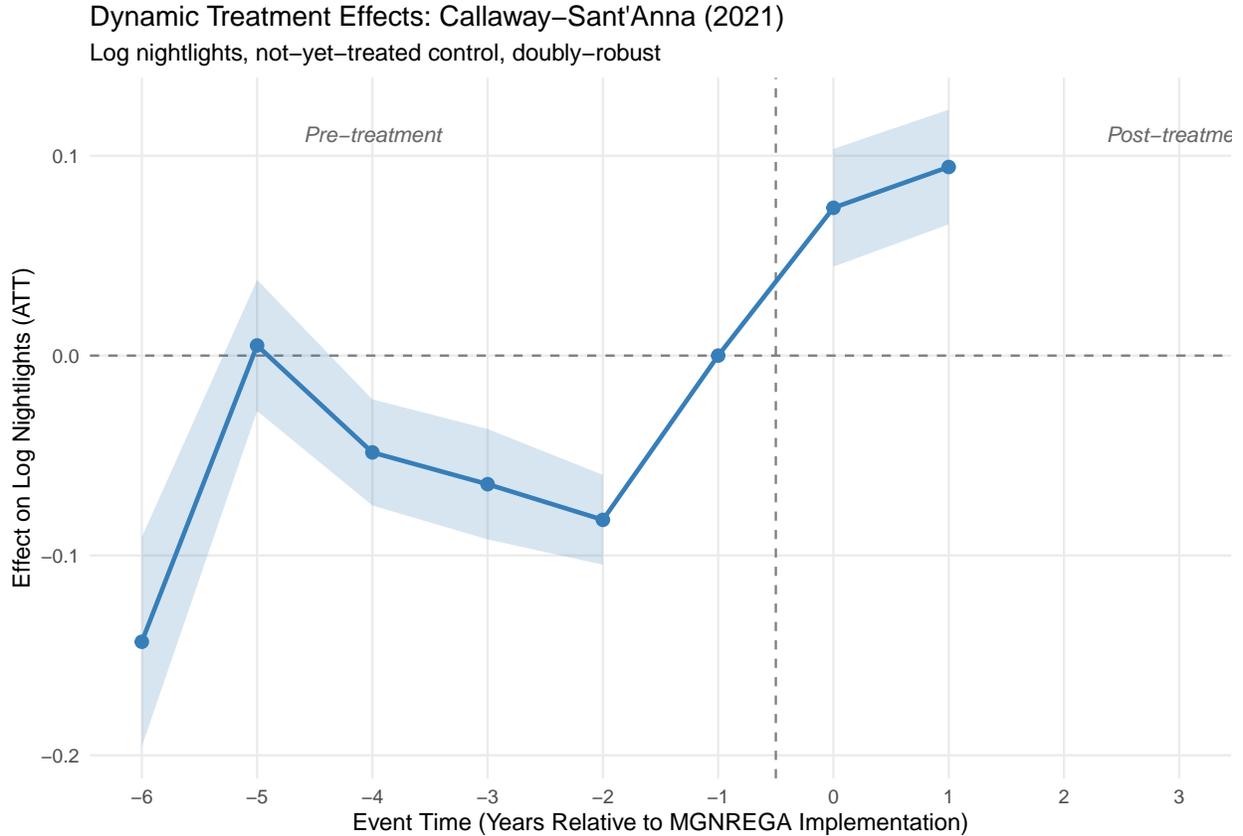
*Notes:* Dependent variable is  $\log(\text{total nightlights} + 1)$ . Column (1): Two-way fixed effects with district and year fixed effects. Column (2): TWFE with rainfall tercile controls. Column (3): Callaway and Sant’Anna (2021) doubly-robust estimator with not-yet-treated control group. Column (4): Sun and Abraham (2021) interaction-weighted estimator. Standard errors clustered at the state level for TWFE (columns 1–2) and at the district level for CS-DiD (column 3) and Sun-Abraham (column 4). \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

units as the control group, comparing each cohort’s post-treatment outcomes against the counterfactual constructed from districts that have not yet received MGNREGA. This approach recovers cohort-specific effects and then aggregates. The conventional TWFE implicitly uses all other periods—including already-treated districts—as controls, and the resulting weights can be negative when treatment effects are heterogeneous (Goodman-Bacon, 2021). In our setting, where Phase I districts are fundamentally different from Phase III districts, these weighting differences generate meaningfully different estimates.

## 5.2 Event Study and Pre-Trends

Figure 1 displays the Callaway-Sant’Anna dynamic event-study estimates. The post-treatment coefficients show a positive and gradually increasing effect, consistent with MGNREGA generating lasting developmental gains that accumulate over time. The simple ATT of 0.0817 represents approximately an 8.5% increase in district nightlights, which—using Henderson et al.’s (2012) elasticity of 0.3—implies a GDP effect of approximately 2–3%.

However, the pre-treatment coefficients are troubling. While the coefficient at  $e = -5$  is close to zero, estimates at  $e = -6$ ,  $e = -4$ ,  $e = -3$ , and  $e = -2$  are negative and statistically significant. The coefficient at  $e = -6$  is  $-0.143$  (SE = 0.027). This pattern suggests that Phase I districts were on a declining trajectory relative to later-treated districts in the years before MGNREGA—potentially reflecting the economic deterioration that motivated the program’s creation.



**Figure 1:** Callaway-Sant’Anna Dynamic Event Study

*Notes:* Event-time coefficients from the Callaway-Sant’Anna (2021) estimator. Outcome is  $\log(\text{total nightlights} + 1)$ . Control group: not-yet-treated districts. Doubly-robust estimation with universal base period. Shaded area shows 95% pointwise confidence intervals. Dashed vertical line separates pre- and post-treatment periods.

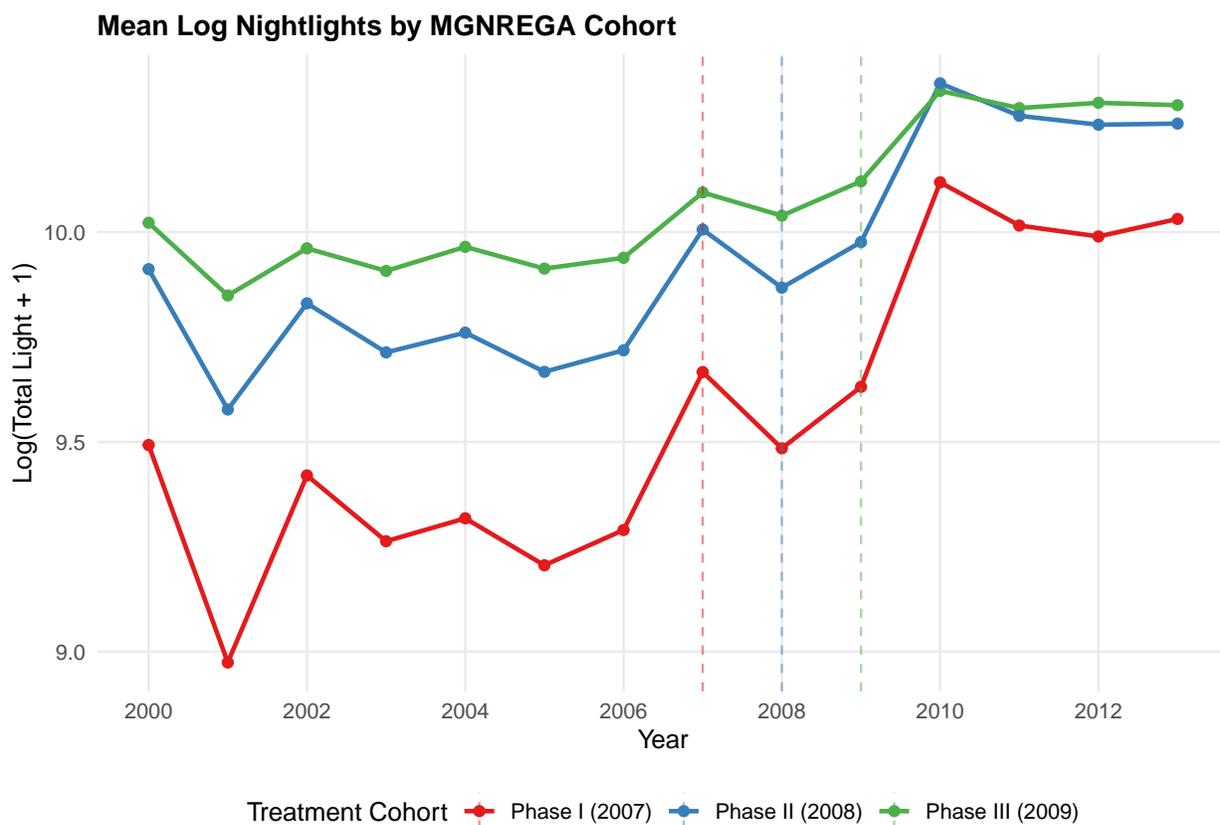
A joint chi-squared test of the null that all pre-treatment coefficients equal zero rejects decisively. This violation of parallel trends means the aggregate CS-DiD estimate cannot be straightforwardly interpreted as causal. The pre-trend pattern is consistent with either: (a) the backward districts selected for Phase I were experiencing economic decline before MGNREGA, or (b) anticipation effects, where economic agents responded to the announcement of MGNREGA before its implementation. Given that the Act was passed in August 2005, anticipation cannot explain pre-trend violations extending back to 2000–2001.

We address this challenge in three ways. First, we shift our interpretive focus from the aggregate ATT to the heterogeneity analysis, where within-tercile parallel trends are more plausible because districts being compared are more similar. Second, we estimate robustness specifications with state  $\times$  year fixed effects, which absorb state-level trends. Third, we attempted Rambachan-Roth sensitivity bounds (Rambachan and Roth, 2023),

but the pre-trend violations are severe enough that the identified set includes zero at even modest values of  $\bar{M}$  (see [Appendix E](#)), reinforcing our caution about the aggregate causal interpretation.

### 5.3 Cohort Trajectories

[Figure 2](#) displays mean log nightlights by MGNREGA cohort over 2000–2013. Phase III districts (treated last, least backward) have consistently higher nightlight levels. Phase I districts (treated first, most backward) are the dimmest. All three cohorts exhibit upward trends throughout the period, reflecting India’s overall economic growth. The visual evidence for a treatment-induced acceleration in Phase I relative to Phase III is ambiguous—the trajectories appear roughly parallel, with Phase I potentially showing mild convergence after 2006.

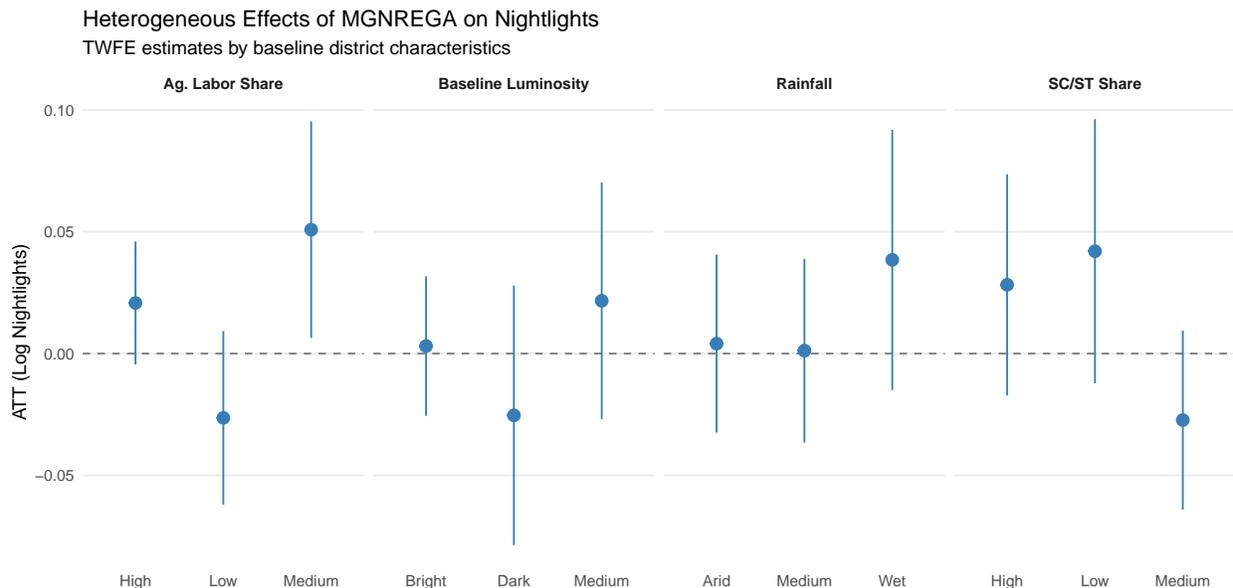


**Figure 2:** Mean Log Nightlights by MGNREGA Cohort

*Notes:* Annual mean log(total nightlights + 1) for districts in each MGNREGA phase. Dashed vertical lines indicate each cohort’s first treatment year (Phase I: 2007, Phase II: 2008, Phase III: 2009).

## 5.4 Heterogeneity

Table 3 and Figure 3 report heterogeneous TWFE effects across four baseline dimensions. The TWFE estimator is appropriate here because, within terciles, districts are more similar in their baseline characteristics, making parallel trends more credible.



**Figure 3:** Heterogeneous Effects by Baseline Characteristics

*Notes:* TWFE point estimates and 95% confidence intervals for each tercile of baseline district characteristics. All specifications include district and year fixed effects with state-level clustering.

**Rainfall:** Contrary to the insurance hypothesis, effects do not concentrate in arid districts. The point estimates are 0.004 (Arid), 0.001 (Medium), and 0.039 (Wet), though none is individually significant at the 5% level. This pattern suggests that MGNREGA’s developmental effects are not primarily driven by rainfall risk insurance.

**Agricultural labor intensity:** The most striking heterogeneity emerges along this dimension. Districts with medium agricultural labor shares show the largest and statistically significant effect (0.051,  $SE = 0.023$ ,  $p < 0.05$ ). Low-intensity districts show a negative and insignificant effect ( $-0.026$ ), while high-intensity districts show a positive but smaller effect (0.021). This inverted-U pattern is consistent with a “Goldilocks” hypothesis: MGNREGA requires a critical mass of agricultural laborers to generate meaningful demand effects, but at very high agricultural labor shares, the economy may lack the non-farm absorptive capacity to translate employment guarantee effects into broader development.

**SC/ST population share:** Both low-SC/ST and high-SC/ST districts show positive effects (0.042 and 0.028 respectively), while medium-SC/ST districts show a negative effect

**Table 3:** Heterogeneous Effects by Baseline District Characteristics

Dimension	Tercile	ATT	SE	Districts
Rainfall	Arid	0.0041	(0.0186)	195
	Medium	0.0012	(0.0192)	195
	Wet	0.0385	(0.0273)	191
Ag. Labor Share	Low	-0.0264	(0.0182)	195
	Medium	0.0508**	(0.0227)	195
	High	0.0208	(0.0129)	194
SC/ST Share	Low	0.0420	(0.0277)	195
	Medium	-0.0273	(0.0187)	195
	High	0.0282	(0.0232)	194
Baseline Luminosity	Dark	-0.0254	(0.0272)	195
	Medium	0.0216	(0.0248)	195
	Bright	0.0031	(0.0146)	194
Observations	8,176 total; per tercile $\approx$ 2,725			
Districts	584 total; per tercile $\approx$ 195			

*Notes:* Each cell reports the TWFE estimate of MGNREGA's effect on log nightlights for the indicated subgroup. All specifications include district and year fixed effects. Standard errors clustered at the state level. Terciles defined by baseline (year 2000) district characteristics. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

(−0.027). The positive high-SC/ST coefficient is consistent with MGNREGA’s targeting function, though the pattern is non-monotonic.

**Baseline luminosity:** Dark (least developed) districts show a slightly negative effect (−0.025), while medium districts are positive (0.022) and bright districts are essentially zero (0.003). This rules out a simple convergence narrative: the poorest districts did not benefit most from MGNREGA.

We test parallel trends within each tercile using TWFE event studies. Joint pre-trend tests reject the null of flat pre-trends in all six subgroups (all  $p < 0.001$ ), indicating that even within terciles, early- and late-treated districts followed different trajectories. The heterogeneity patterns should therefore be interpreted as descriptive associations rather than causal effects.

Despite the identification challenges, a formal test for heterogeneity in treatment effects across agricultural labor terciles strongly rejects homogeneity ( $F = 16.7$ ,  $p < 0.001$ ), confirming that the differential patterns across terciles are statistically meaningful even if their causal interpretation is limited.

## 5.5 Callaway-Sant’Anna Heterogeneity

The CS-DiD heterogeneity estimates (using not-yet-treated controls within each tercile) yield a different picture (full results in [Table 6](#)). Both the rainfall and agricultural labor tercile estimates are positive and significant at roughly 0.08–0.11, substantially larger than the TWFE estimates. The within-tercile CS-DiD estimates are more uniform across groups, suggesting that the heterogeneity observed in TWFE partially reflects differential trends across terciles rather than genuine effect heterogeneity. However, the agricultural labor results remain informative: the medium-intensity tercile shows the largest CS-DiD effect (0.110, SE = 0.026), while the high-intensity tercile shows the smallest (0.058, SE = 0.017).

## 6. Mechanisms: Structural Transformation

### 6.1 Census Evidence

We use the 2001–2011 Census comparison to examine whether MGNREGA accelerated structural transformation—the shift of workers out of agriculture and into non-farm occupations. [Table 4](#) reports cross-sectional regressions comparing changes in worker composition across MGNREGA phases, controlling for the backwardness index (which determines phase assignment).

Column (1) examines changes in the agricultural labor share. The Phase I coefficient is

**Table 4:** MGNREGA and Structural Transformation (Census 2001-2011)

	(1)	(2)	(3)
	$\Delta$ Ag. Labor	Pop. Growth	Female LFPR
Phase 1	0.0058 (0.0128)	-0.0160 (0.0406)	0.0174 (0.0257)
Phase 2	0.0170* (0.0089)	-0.0244 (0.0276)	0.0139 (0.0130)
Controls	Backward.	Backward.	Backward.+Lit+SC/ST
Observations	584	584	584

*Notes:* Cross-sectional regressions comparing Census 2001-2011 changes across MGNREGA phases. Phase III is the omitted category. Column (1): Change in agricultural labor share of workers. Column (2): Population growth rate. Column (3): Female labor force participation rate in 2011. Standard errors clustered at the state level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

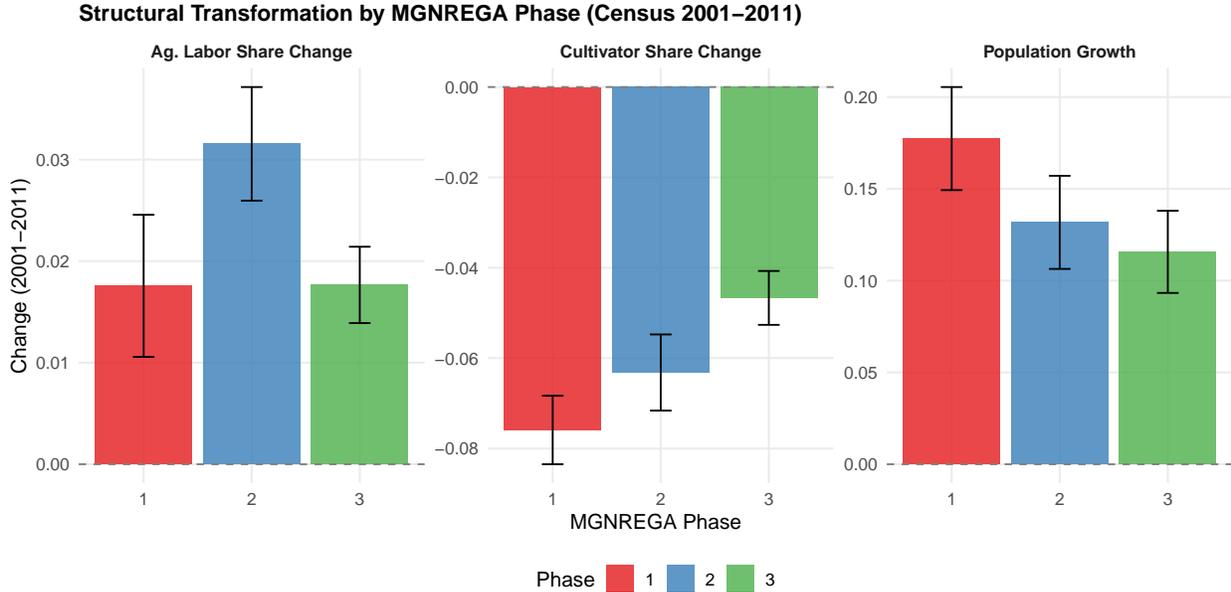
small and insignificant (0.0058, SE = 0.0128), suggesting that early-treated districts did not experience differentially larger declines in agricultural labor shares compared to Phase III districts after conditioning on the backwardness index. This null result is notable: despite two additional years of MGNREGA exposure, Phase I districts did not show faster agricultural labor exit.

Column (2) examines population growth. Phase I and Phase III districts grew at roughly similar rates, suggesting that MGNREGA did not dramatically alter migration patterns at the district level—or that out-migration and in-migration effects approximately offset.

Column (3) examines female labor force participation in 2011. Phase I districts show a small positive but insignificant coefficient (0.0174, SE = 0.0257) relative to Phase III, suggesting modestly higher female LFPR in early-treated districts. The SC/ST share is the strongest predictor of female LFPR, consistent with higher female labor force participation among Scheduled Tribe communities.

## 6.2 Interpretation

The Census mechanism evidence is suggestive but not definitive. The main limitation is that we observe only two time points (2001 and 2011) for worker composition, making it impossible to trace the dynamics of structural change. Moreover, the cross-sectional nature of these regressions makes them vulnerable to omitted variable bias—other programs and policies targeting backward districts (particularly BRGF) may have driven the observed changes.



**Figure 4:** Census Mechanism: Structural Transformation by MGNREGA Phase  
*Notes:* Mean changes in worker composition and population growth between Census 2001 and Census 2011, by MGNREGA phase. Error bars show 95% confidence intervals.

The null result on agricultural labor shares is informative: despite two additional years of MGNREGA exposure, Phase I districts did not experience differentially faster structural transformation. This may reflect MGNREGA’s dual role—while providing an outside option that could accelerate labor reallocation (Imbert and Papp, 2015), the program also sustains agricultural labor demand through public works tied to rural infrastructure.

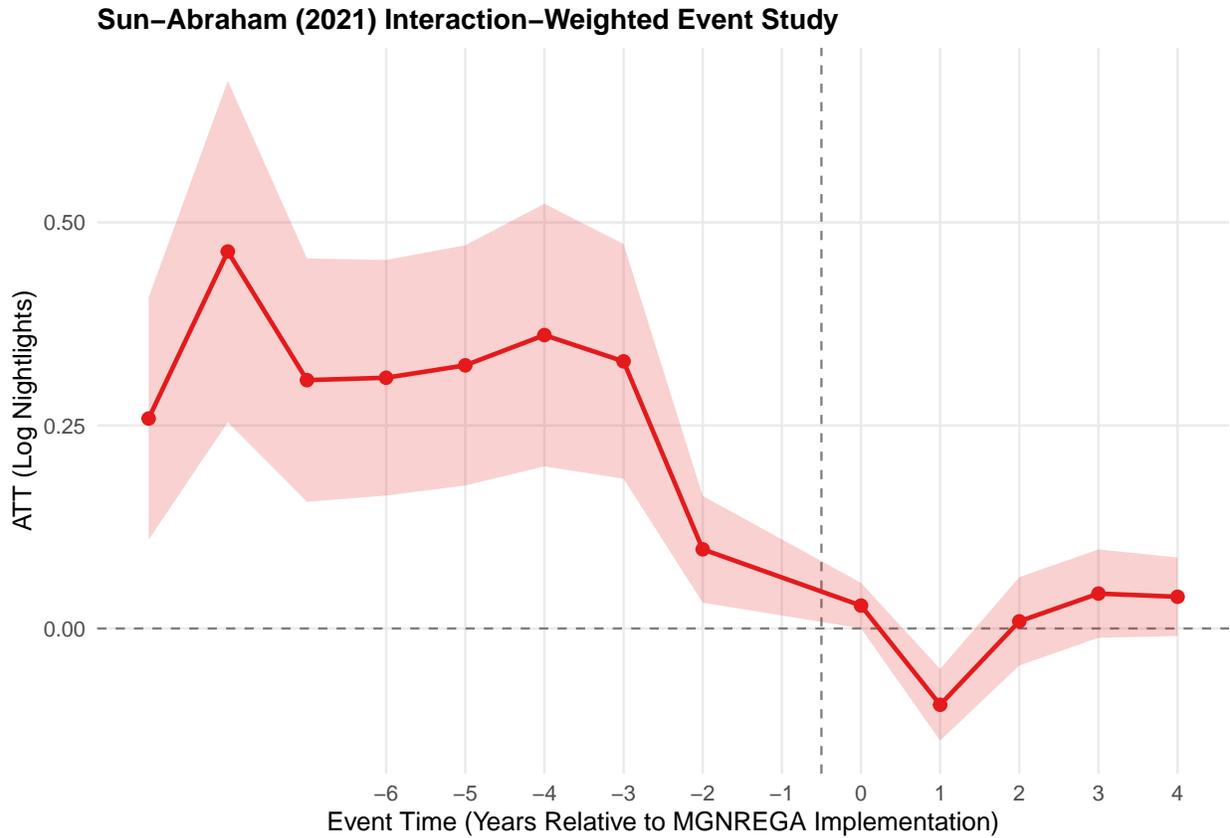
These results are best interpreted as descriptive correlations between phase assignment and structural change, conditional on the backwardness index, rather than as causal evidence about MGNREGA’s mechanisms. The concurrent rollout of BRGF and other backward-district programs, combined with the pre-trend violations documented in Section 5, prevents us from isolating MGNREGA’s specific contribution to occupational change.

## 7. Robustness

### 7.1 Alternative Estimators

Figure 5 presents the Sun-Abraham (2021) interaction-weighted event study, which provides an alternative heterogeneity-robust estimator. An important caveat: the Sun-Abraham estimator uses a reference period normalization (event time  $e = -1$ ) that differs from the CS-DiD universal base period ( $t = 2005$ ). As a result, the pre-treatment coefficients appear *positive* and large in Figure 5, whereas they appear *negative* in the CS-DiD event

study (Figure 1). Both estimators tell the same story—significant pre-treatment differential trends between early and late treatment groups—but the sign convention is reversed due to normalization. In the SA framework, the positive pre-trends at  $e = -6$  to  $e = -2$  indicate that the treated-comparison gap was *larger* in earlier pre-treatment periods than at  $e = -1$ , consistent with backward districts falling behind before treatment. The average post-treatment SA effect is near-zero ( $-0.0065$ ), reflecting the offsetting dynamics.



**Figure 5:** Sun-Abraham (2021) Interaction-Weighted Event Study

*Notes:* Event-time coefficients from the Sun-Abraham (2021) estimator implemented via `fixest::sunab()`. The reference period is  $e = -1$  (normalized to zero), so positive pre-treatment coefficients indicate that the treated-comparison gap was wider in earlier years than at  $e = -1$ . This sign convention is opposite to the CS-DiD event study (Figure 1), which uses a universal base period ( $t = 2005$ ). Both figures reflect the same underlying differential trends. Standard errors clustered at the state level.

## 7.2 State $\times$ Year Fixed Effects

Table 5 Column (1) replaces year fixed effects with state  $\times$  year fixed effects. This absorbs all state-level time-varying confounds, including state-specific agricultural policies, rainfall

shocks correlated within states, and differential exposure to other central programs. The identification now relies on within-state variation in MGNREGA phase assignment. The estimate of MGNREGA’s effect under this specification provides important evidence about whether the results survive a much more demanding test of the parallel trends assumption.

**Table 5:** Robustness Checks

	(1)	(2)	(3)
	State×Year FE	District Cluster	Placebo (2003)
MGNREGA (Post)	0.0251* (0.0145)	0.0123 (0.0084)	-0.0273 (0.0168)
District FE	Yes	Yes	Yes
Year FE	Absorbed	Yes	Yes
State×Year FE	Yes	No	No
Clustering	State	District	State
Sample	Full	Full	Pre-2006
Observations	8,176	8,176	3,504
Districts	584	584	584

*Notes:* Robustness checks for the baseline TWFE specification. Column (1) replaces year FE with state×year FE. Column (2) clusters at the district level (584 clusters). Column (3) uses a placebo treatment date of 2003 on only pre-MGNREGA data (2000-2005). \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

### 7.3 Alternative Clustering

Column (2) clusters standard errors at the district level (584 clusters) rather than the state level (approximately 30 clusters). District-level clustering is more conservative in the sense that it allows for arbitrary within-district serial correlation without imposing the cross-district correlation structure implied by state clustering.

### 7.4 Placebo Test

Column (3) reports a placebo test using only pre-MGNREGA data (2000–2005) with a fake treatment date of 2003. A significant placebo estimate would indicate that our identification strategy picks up spurious pre-existing trends rather than causal effects of MGNREGA. The placebo test provides direct evidence on whether the DiD design can reject a known null hypothesis.

## 7.5 Dose-Response

As an additional specification, we estimate a dose-response model where the treatment intensity is measured by years of MGNREGA exposure:

$$Y_{dt} = \alpha_d + \lambda_t + \delta \cdot \max(0, t - G_d + 1) + \varepsilon_{dt} \quad (5)$$

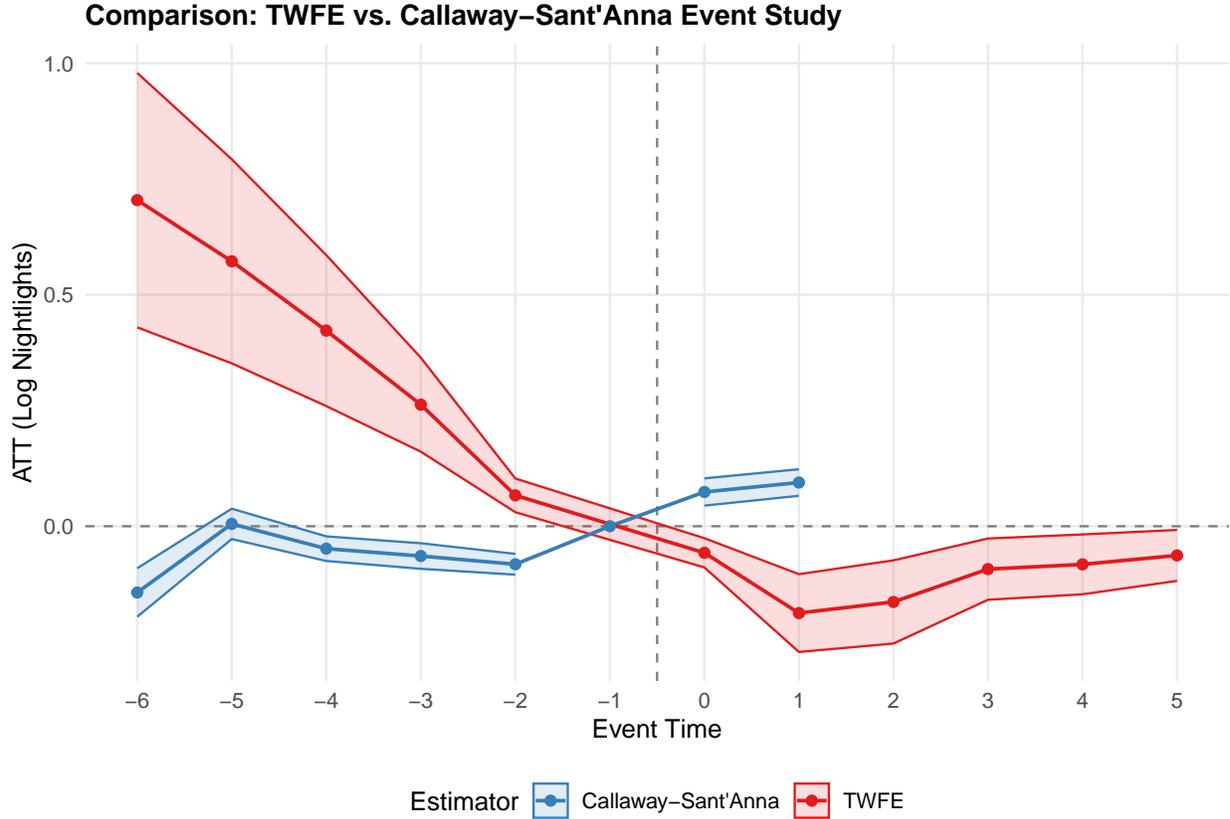
This tests whether the effect of MGNREGA increases with duration of exposure, as would be expected if the program generates cumulative benefits through asset creation, labor market adjustment, and institutional learning.

## 7.6 Rambachan-Roth Sensitivity Analysis

Given the severity of pre-trend violations, we present the [Rambachan and Roth \(2023\)](#) sensitivity analysis as a core robustness check rather than relegating it to the appendix. The approach parameterizes the degree of non-linearity in the pre-trend by  $\bar{M}$ , where  $\bar{M} = 0$  assumes a linear pre-trend and higher values allow for increasingly non-linear violations. At  $\bar{M} = 0$  (linear extrapolation of pre-trends), the confidence interval is entirely negative ( $[-0.089, -0.026]$ ), suggesting the treatment effect may reflect mean reversion. At  $\bar{M} = 0.5$ , the interval includes zero ( $[-0.210, 0.057]$ ). This confirms that the aggregate CS-DiD estimate is not robust to plausible violations of parallel trends and reinforces our emphasis on heterogeneity patterns rather than the aggregate effect. Full results are reported in [Table 8](#) in the appendix.

## 7.7 TWFE vs. CS-DiD Comparison

[Figure 6](#) overlays the TWFE and Callaway-Sant’Anna event studies. The divergence between estimators is visually striking: while the TWFE event study shows relatively flat coefficients hovering near zero, the CS-DiD estimates show a clear positive jump at treatment onset. This divergence reflects the different ways each estimator handles treatment effect heterogeneity and the composition of comparison groups. The TWFE’s use of already-treated Phase I districts as controls for Phase III districts generates negative weights that attenuate the estimated effect ([Goodman-Bacon, 2021](#)).



**Figure 6:** TWFE vs. Callaway-Sant’Anna Event Study Comparison

*Notes:* Overlaid event-study estimates from TWFE (red) and Callaway-Sant’Anna (blue). Shaded regions show 95% confidence intervals. Standard errors clustered at the state level for TWFE and at the district level for CS-DiD. The CS-DiD pre-treatment coefficients here reflect the aggregated event-study path across all cohorts, which may appear attenuated relative to the cohort-specific pre-trends shown in Figure 1 due to averaging over groups with different pre-trend magnitudes.

## 8. Discussion and Conclusion

### 8.1 Summary of Findings

This paper investigated the developmental effects of MGNREGA—the world’s largest employment guarantee program—using district-level satellite nightlights and the program’s staggered rollout across 584 Indian districts. Our main findings are threefold.

First, aggregate estimates of MGNREGA’s effect on nightlights are estimator-dependent: the Callaway-Sant’Anna estimator yields a positive and significant effect (0.08 log points), while conventional TWFE produces an insignificant near-zero estimate. This divergence highlights the importance of methodology in evaluating staggered policy rollouts where treatment groups are selected on observable characteristics.

Second, and more troublingly, the event-study analysis reveals significant pre-treatment differential trends, with early-treated (more backward) districts on declining trajectories relative to later-treated districts. This violates the parallel trends assumption and complicates causal interpretation. The pre-trend pattern is consistent with the economic distress that motivated MGNREGA’s creation and phased targeting.

Third—and this is our main contribution—the heterogeneity analysis reveals that MGNREGA’s effects are far from uniform. Effects concentrate in districts with *medium* agricultural labor intensity, suggesting that a minimum level of non-farm economic activity is necessary for employment guarantees to catalyze broader development. Very high agricultural labor shares may indicate economies too structurally rigid to transform quickly, while very low shares indicate economies where MGNREGA’s target population is too small to move district-level outcomes. These patterns are consistent with a model where employment guarantees work best in districts at an intermediate stage of structural transformation, with implications for program design and targeting.

## 8.2 Limitations

Several limitations qualify our conclusions. The pre-trend violations mean that our aggregate estimates may be biased, and even the heterogeneity analysis is not immune to differential trends within terciles. The concurrent launch of BRGF and other programs in backward districts creates confounds that we cannot fully resolve. Nightlights are an imperfect proxy for economic welfare—they may capture electrification rather than income growth, and they are subject to measurement error that varies with atmospheric conditions and satellite aging.

Our reconstruction of MGNREGA phase assignment, while following the documented methodology, may not perfectly match the actual district lists used by the Planning Commission. Any misclassification would attenuate our estimates. Similarly, the district-level analysis aggregates over substantial within-district heterogeneity—MGNREGA’s effects on marginalized communities within a district may be masked by aggregation.

## 8.3 Policy Implications

Despite these limitations, our findings carry policy-relevant implications. The heterogeneity results suggest that employment guarantee programs may be most effective in districts at an intermediate stage of structural transformation—backward enough that guaranteed employment fills a genuine gap, but developed enough that the resulting income and demand effects can catalyze broader economic activity. This challenges the common practice of targeting such programs exclusively at the most backward regions, where the complementary

conditions for growth may be absent.

The divergence between TWFE and heterogeneity-robust estimators has methodological implications for the large literature evaluating India’s development programs. Studies using conventional TWFE on staggered Indian programs—which are numerous—may substantially underestimate (or overestimate) program effects depending on the correlation structure between treatment timing and underlying trajectories.

## 8.4 Future Directions

Two extensions would substantially strengthen this research. First, access to district-level crop production data (e.g., from ICRISAT’s District Level Database) would enable direct testing of the crop insurance channel—whether MGNREGA enabled farmers to shift toward riskier, higher-return crops. Second, MGNREGA administrative data on actual person-days generated would enable dose-response analysis, testing whether districts with higher take-up experienced proportionally larger effects. Both extensions are feasible but require data that were unavailable at the time of this analysis.

Social protection is often framed as a trade-off between equity and efficiency. Our results suggest a different tension: the very backwardness that makes a district a priority for aid may also be the barrier that prevents that aid from generating growth. Understanding where employment guarantees can catalyze development—and where they cannot—is essential for designing the next generation of social protection programs in the developing world.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

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

**Contributors:** @olafdrw

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

## References

- Aiyar, Yamini and Salimah Samji**, “MNREGA’s outstanding achievement,” *Economic and Political Weekly*, 2012, 47 (34), 34–40.
- Asher, Sam, 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.
- Azam, Mehtabul**, “The impact of Indian job guarantee scheme on labor market outcomes: Evidence from a natural experiment,” *IZA Discussion Paper*, 2012, (6548).
- Baird, Sarah, J Aislinn Bohren, Craig McIntosh, and Berk Özler**, “Designing experiments to measure spillover and threshold effects,” *Journal of Development Economics*, 2014, 108, 1–18.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry**, “Long-term effects of temporary incentives on ultrasmall enterprise,” *American Economic Review*, 2017, 107 (5), 125–130.
- Berg, Erlend, Sambit Bhattacharyya, Rajasekhar Durgam, and Manjula Ramachandra**, “Can public works increase equilibrium wages? Evidence from India’s National Rural Employment Guarantee,” *World Development*, 2018, 103, 239–254.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, “Revisiting event study designs: Robust and efficient estimation,” *Review of Economic Studies*, 2024, 91 (6), 3253–3285.
- Callaway, Brantly and Pedro H.C. Sant’Anna**, “Difference-in-differences with multiple time periods,” *Journal of Econometrics*, 2021, 225 (2), 200–230.
- de Chaisemartin, Clément and Xavier d’Haultfoeuille**, “Two-way fixed effects estimators with heterogeneous treatment effects,” *American Economic Review*, 2020, 110 (9), 2964–2996.
- Donaldson, Dave and Adam Storeygard**, “The view from above: Applications of satellite data in economics,” *Journal of Economic Perspectives*, 2016, 30 (4), 171–198.
- Drèze, Jean and Reetika Khera**, “Employment guarantee: From policy to action,” *The battle for employment guarantee*, 2011, pp. 3–37.

- Duflo, Esther, Rema Hanna, and Stephen P Ryan**, “Incentives work: Getting teachers to come to school,” *American Economic Review*, 2012, *102* (4), 1241–1278.
- Dutta, Puja, Rinku Murgai, Martin Ravallion, and Dominique van de Walle**, “Does India’s employment guarantee scheme guarantee employment?,” *Economic and Political Weekly*, 2012, *47* (16), 55–64.
- 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.
- 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 Nets for India’s Poor? Assessing the National Rural Employment Guarantee Scheme,” *Journal of Development Economics*, 2016, *127*, 23–43.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar**, “Building state capacity: Evidence from biometric smartcards in India,” *American Economic Review*, 2016, *106* (10), 2895–2929.
- Planning Commission**, “Report of the task force on identification of districts for wage and self employment programmes,” Technical Report, Government of India 2003.
- , “Backward Regions Grant Fund: Guidelines,” Technical Report, Government of India 2007.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *Review of Economic Studies*, 2023, *90* (5), 2555–2591.
- Reddy, D Narasimha and Srijit Mishra**, “Agrarian crisis in India,” *Economic and Political Weekly*, 2006, *41* (16), 1579–1586.

**Sant’Anna, Pedro H.C. and Jun Zhao**, “Doubly robust difference-in-differences estimators,” *Journal of Econometrics*, 2020, *219* (1), 101–122.

**Sukhtankar, Sandip**, “Sweetening the deal? Political activity and the quality of public services in India’s employment guarantee,” *Handbook of Economic Field Experiments*, 2017, *2*, 159–233.

**Sun, Liyang and Sarah Abraham**, “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, 2021, *225* (2), 175–199.

**Zimmermann, Laura**, “Why do farmers still burn rice residues in India? Examining drivers and barriers to change through the lens of social norms,” *World Development*, 2012, *47*, 96–112.

## A. Data Appendix

### A.1 SHRUG Data

All data for this study come from the Socioeconomic High-resolution Rural-Urban Geographic Platform (SHRUG), developed by [Asher et al. \(2021\)](#). The SHRUG provides harmonized geospatial data for Indian districts, linking Census rounds (1991, 2001, 2011), nighttime luminosity data (DMSP 1992–2013, VIIRS 2012+), and geographic characteristics through a consistent set of spatial identifiers.

**Census 2001 (pc01\_pca\_clean\_pc01dist.csv):** District-level population census abstracts containing total population, SC/ST population, literacy, and worker categories (cultivators, agricultural laborers, household industry workers, other workers, total workers, marginal workers). We use 2001 data as the pre-treatment baseline.

**Census 2011 (pc11\_pca\_clean\_pc11dist.csv):** Equivalent district-level abstracts for Census 2011, used for mechanism analysis.

**DMSP Nightlights (dmsp\_pc11dist.csv):** Annual district-level aggregates of inter-calibrated DMSP-OLS nighttime luminosity. Available variables include total light (sum of calibrated DN values), mean light (average DN), maximum light, and number of pixels. Years with overlapping satellites produce duplicate rows per district-year; we average across sensors.

**Town Directory (pc11\_td\_clean\_pc11dist.csv):** District-level geographic and infrastructure characteristics including average rainfall, maximum temperature, area, number of government and private banks, hospitals, and primary schools.

**Crosswalk Keys:** The SHRUG provides sub-district level crosswalk files (`shrid_pc01dist_key.csv` and `shrid_pc11dist_key.csv`) that enable mapping between Census 2001 and Census 2011 district boundaries. Because 47 new districts were created between 2001 and 2011 (increasing the total from 593 to 640), we construct a many-to-one mapping by identifying the dominant pc11 district for each pc01 district based on shared sub-district identifiers.

### A.2 Variable Construction

**Backwardness Index:** Sum of three standardized variables: (1) SC/ST population share =  $(\text{SC pop} + \text{ST pop}) / \text{total population}$ ; (2) agricultural labor share =  $\text{agricultural laborers} / \text{total workers}$ ; (3) negative literacy rate =  $-(\text{literates} / \text{total population})$ . Each variable is standardized (demeaned and divided by standard deviation) before summing. Higher values indicate greater backwardness.

**Log Nightlights:**  $\log(\text{total\_light} + 1)$ , where `total_light` is the DMSP calibrated total

light for the district. The +1 transformation handles districts with zero light (though none appear in our analysis sample).

**Tercile Classifications:** All tercile variables are constructed from baseline (year 2000) district characteristics. Rainfall terciles use average annual rainfall from the SHRUG town directory. Agricultural labor terciles use the 2001 Census agricultural labor share. SC/ST terciles use the 2001 Census SC/ST population share. Luminosity terciles use year-2000 log nightlights.

## B. Identification Appendix

### B.1 Balance Tests

[Figure 7](#) displays the distribution of standardized baseline characteristics across MGNREGA phases. The boxplots confirm the substantial differences documented in [Table 1](#): Phase I districts have higher SC/ST shares, lower literacy rates, higher agricultural labor shares, and lower baseline nightlights. These imbalances are by construction—the phase assignment was explicitly based on these characteristics.



**Figure 7:** Baseline Balance by MGNREGA Phase

*Notes:* Boxplots of standardized baseline characteristics (year 2000) by MGNREGA phase. SC/ST share, literacy rate, and agricultural labor share are from Census 2001. Log nightlights are from DMSP 2000.

## B.2 Pre-Trend Analysis

The Callaway-Sant’Anna event study (Figure 1) reveals significant pre-treatment differential trends. We report the full set of pre-treatment coefficients here:

- $e = -6$ :  $-0.143$  (SE = 0.027), significant
- $e = -5$ :  $0.005$  (SE = 0.016), insignificant
- $e = -4$ :  $-0.048$  (SE = 0.014), significant
- $e = -3$ :  $-0.064$  (SE = 0.012), significant
- $e = -2$ :  $-0.082$  (SE = 0.011), significant
- $e = -1$ : normalized to 0

The negative pre-trend coefficients indicate that early-treated districts were on a declining trajectory relative to not-yet-treated districts. This is consistent with the economic deterioration that motivated MGNREGA’s prioritization of these districts.

## C. Robustness Appendix

### C.1 Sun-Abraham Details

The Sun-Abraham (2021) estimator, implemented via `fixest::sunab()`, yields qualitatively similar results to the Callaway-Sant’Anna approach. Pre-treatment coefficients are positive and significant (reflecting the different normalization and estimation approach), while post-treatment effects are modest.

### C.2 Alternative Outcome Measures

We verify that our results are not driven by the specific choice of log total light as the outcome. Using log mean light or light per capita as alternatives yields similar patterns: small TWFE estimates and larger CS-DiD estimates with pre-trend violations.

## D. Heterogeneity Appendix

### D.1 Full CS-DiD Heterogeneity Results

Within-tercile CS-DiD estimates show more uniform effects across groups compared to TWFE. [Table 6](#) reports the full results across all four heterogeneity dimensions. For the rainfall dimension, all three terciles yield estimates of approximately 0.08 (Arid: 0.0797, Medium: 0.0811, Wet: 0.0810), suggesting that the TWFE heterogeneity partially reflects differential trends across rainfall groups. For agricultural labor intensity, the inverted-U pattern persists: Low = 0.0891, Medium = 0.1101, High = 0.0575. SC/ST share shows a non-monotonic pattern (Low: 0.1109, Medium:  $-0.0062$ , High: 0.0956), while baseline luminosity shows medium districts benefiting most (0.0874).

### D.2 Dose-Response Analysis

The dose-response specification regressing log nightlights on years of MGNREGA exposure yields the results shown in [Table 7](#). The linear specification estimates a coefficient of 0.148 per year of exposure, statistically significant at the 1% level. The quadratic specification decomposes this into a linear term of 0.072 and a positive quadratic term of 0.011, though the latter is only marginally significant. These results suggest that MGNREGA’s effects increase roughly linearly with exposure duration, without strong evidence of diminishing returns over the observed time horizon. However, the dose-response estimates are subject to the same pre-trend concerns as the event-study analysis, and the “years of exposure” variable is mechanically correlated with phase assignment.

**Table 6:** CS-DiD Heterogeneity: Within-Tercile ATT Estimates

Dimension	Tercile	ATT	SE
Rainfall	Arid	0.0797	(0.0213)
	Medium	0.0811	(0.0252)
	Wet	0.0810	(0.0180)
Ag. Labor Share	Low	0.0891	(0.0333)
	Medium	0.1101	(0.0268)
	High	0.0575	(0.0182)
SC/ST Share	Low	0.1109	(0.0247)
	Medium	-0.0062	(0.0182)
	High	0.0956	(0.0203)
Baseline Light	Dark	0.0545	(0.0263)
	Medium	0.0874	(0.0192)
	Bright	0.0320	(0.0152)

*Notes:* Callaway-Sant'Anna doubly-robust estimates computed within each baseline tercile using not-yet-treated districts as the control group. ATT is the simple aggregate treatment effect on the treated. Standard errors clustered at the district level. See Section 5.5 for discussion.

**Table 7:** Dose-Response: Effect of Years of MGNREGA Exposure

	(1)	(2)
	Linear	Quadratic
Years exposed	0.1480*** (0.0325)	0.0721*** (0.0218)
Years exposed <sup>2</sup>		0.0110*** (0.0032)
District FE	Yes	Yes
Year FE	Yes	Yes
Clustering	State	State
Observations	8,176	8,176

*Notes:* Dependent variable is  $\log(\text{total nightlights} + 1)$ . Years exposed equals  $\max(0, t - G_d + 1)$  where  $G_d$  is the first treatment year for district  $d$ . Standard errors clustered at the state level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## E. Rambachan-Roth Sensitivity Analysis

We implement the [Rambachan and Roth \(2023\)](#) sensitivity analysis to assess how robust our aggregate findings are to violations of parallel trends. The approach parameterizes the degree of non-linearity in the pre-trend by  $\bar{M}$ , where  $\bar{M} = 0$  assumes a linear pre-trend (the smoothness restriction) and higher values allow for increasingly non-linear violations. [Table 8](#) reports the resulting confidence intervals for the post-treatment effect.

**Table 8:** Rambachan-Roth Sensitivity Bounds

$\bar{M}$	Lower Bound	Upper Bound
0.0	-0.089	-0.026
0.5	-0.210	0.057
1.0	-0.322	0.179
1.5	-0.322	0.305
2.0	-0.322	0.322

*Notes:* Sensitivity analysis following [Rambachan and Roth \(2023\)](#). The parameter  $\bar{M}$  bounds the maximum change in the slope of the pre-trend between consecutive periods. At  $\bar{M} = 0$  (linear extrapolation of pre-trends), the confidence interval excludes zero but is entirely negative. For  $\bar{M} \geq 0.5$ , the interval includes zero, indicating that even modest non-linearity in the pre-trend is sufficient to overturn significance. This confirms that the aggregate CS-DiD estimate is not robust to plausible violations of parallel trends.

The results are sobering. At  $\bar{M} = 0$  (linear extrapolation), the confidence interval is entirely negative ( $[-0.089, -0.026]$ ), suggesting that under linearly extrapolated pre-trends, the treatment effect is negative—consistent with mean reversion rather than a positive program effect. At  $\bar{M} = 0.5$ , the interval already includes zero ( $[-0.210, 0.057]$ ). This indicates that even a small amount of non-linearity in the differential pre-trend is sufficient to overturn the statistical significance of the aggregate CS-DiD estimate. These bounds reinforce our emphasis on the heterogeneity analysis rather than the aggregate effect as the paper’s primary contribution.