Did India's Employment Guarantee Transform the Rural Economy?
Evidence from Three Decades of Satellite Data

APEP Autonomous Research*(cumulative: 4h 12m).

@olafdrw

February 20, 2026

Abstract

India's MGNREGA—the world's largest public works program—was rolled out to 640 districts in three phases (2006–2008) based on a composite backwardness index. I exploit this staggered rollout using Callaway-Sant'Anna difference-in-differences and 30 years of satellite nightlight data. Across four estimators, I find no robust effect on district nightlights. Results are fragile: TWFE yields small positive estimates, the Callaway-Sant'Anna ATT is near zero (0.033, SE 0.144), and the Sun-Abraham estimator produces a significant *negative* effect (-0.167, $p < 0.01$). This sign instability indicates fragility rather than a clean null. Cross-sectional Census evidence confirms no differential structural transformation between early and late treatment districts. Despite 12 pre-treatment years, MGNREGA's effect on aggregate economic activity appears at best modest and possibly zero.

JEL Codes: H53, I38, O13, O15, R11

Keywords: MGNREGA, employment guarantee, nightlights, structural transformation, India, staggered DiD

---

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

1. Introduction

For twenty years, India ran the world's largest public works program. At its peak, the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) covered 270 million rural workers and spent over $8 billion annually. In December 2025, the Indian government quietly replaced it with the Employment Linked Incentive (ELI) scheme—ending a vast experiment whose legacy remains fiercely contested. Advocates credit MGNREGA with transforming rural labor markets, raising agricultural wages, and empowering women and lower-caste workers (Imbert and Papp, 2015; Muralidharan et al., 2016). Critics counter that it crowded out productive private employment, sustained inefficient agriculture, and became a vehicle for corruption rather than development (Sukhtankar, 2012; Niehaus and Sukhtankar, 2013). After two decades, a fundamental question remains unresolved: did MGNREGA accelerate or retard the structural transformation of India's rural economy?

This paper provides the first long-run assessment of MGNREGA's effects on aggregate local economic activity using three decades of satellite nightlight data. I construct a district-year panel spanning 1994–2023, covering 640 districts across 30 years, and exploit the program's three-phase staggered rollout between 2006 and 2008 as a natural experiment. Phase assignment was determined by the Planning Commission's composite backwardness index, which ranked districts by SC/ST population share, agricultural labor share, and inverse literacy rate—creating quasi-exogenous variation in treatment timing conditional on baseline characteristics.

The core finding is that results are fragile and estimator-dependent, with no robust positive effect. Across four distinct estimators—two-way fixed effects, TWFE with state-by-year interactions, the Callaway and Sant'Anna (2021) regression-based estimator, and the Sun and Abraham (2021) interaction-weighted estimator—I find no consistent evidence that MGNREGA affected district-level nightlights in a particular direction. The baseline TWFE estimate is 0.056 log points (standard error 0.089), the Callaway-Sant'Anna ATT is 0.033 (0.144), and results are sensitive to specification: adding state-by-year fixed effects increases the point estimate to 0.114 (0.066), while the Sun-Abraham estimator yields a statistically significant *negative* effect of −0.167 (0.058), significant at the 1% level. This sign instability across estimators—with one method finding a significant negative effect and others finding imprecise positive effects—indicates that the results are fragile rather than constituting a clean null, and that the true effect is not robustly identified in this design.

Three pieces of evidence reinforce the null interpretation. First, cross-sectional Census data comparing the 2001 and 2011 worker composition of Phase I, II, and III districts reveals no differential structural transformation. The change in non-farm worker share differs by only

0.6 percentage points between early and late treatment districts (standard error 0.5 percentage points), with state fixed effects absorbing geographic heterogeneity. Second, heterogeneity analysis by baseline development level reveals no systematic pattern: MGNREGA effects are statistically indistinguishable from zero across all four quartiles of initial nightlight intensity. Third, Bacon decomposition reveals that 57% of the TWFE identifying variation comes from "later versus earlier treated" comparisons—precisely the contrasts most susceptible to bias from heterogeneous treatment effects in staggered designs (Goodman-Bacon, 2021).

The identification strategy rests on parallel trends in nightlight trajectories across districts assigned to different MGNREGA phases. Several diagnostics support this assumption. A joint Wald test of eight pre-treatment event-study coefficients fails to reject the null of zero ($\chi^2(8) = 7.14$, $p = 0.52$). The Callaway-Sant'Anna event-study plot shows pre-treatment coefficients centered on zero with no discernible trend. However, a placebo test shifting treatment three years earlier produces a statistically significant coefficient of 0.184 on pre-treatment data only, suggesting some residual pre-existing divergence. Following Roth (2022), I interpret the pre-trend test results with caution: a failure to reject parallel trends does not guarantee their validity, and the significant placebo suggests that pre-treatment growth rates may have differed across phase groups. I discuss this threat in detail and show that it likely reflects differential pre-treatment growth rates correlated with the backwardness ranking rather than anticipation effects.

This paper contributes to three literatures. First, it adds to the evaluation of MGNREGA, which has produced a rich but largely short-run evidence base. Imbert and Papp (2015) find that MGNREGA raised private-sector wages by 4.7% using a regression discontinuity design around the Phase I/II cutoff. Muralidharan et al. (2016) document that MGNREGA increased public employment earnings and reduced private-sector employment in Andhra Pradesh. Azam (2012) find positive effects on consumption using a difference-in-differences approach. Zimmermann (2021) provides a comprehensive review of the program's varied effects on labor markets. But these studies typically examine effects within three to five years of implementation. My contribution is to extend the horizon to 17 years post-treatment for Phase I districts, long enough for structural transformation—the reallocation of labor from agriculture to manufacturing and services—to materialize if MGNREGA's demand multiplier effects dominate crowd-out.

Second, this paper contributes to the literature on satellite nightlights as a measure of economic activity. Henderson et al. (2012) show that nightlights are a useful proxy for GDP growth, particularly in developing countries with weak statistical infrastructure. Asher et al. (2021) apply SHRUG nightlight data to study Indian development at the village level. I extend this approach to evaluate a specific policy intervention, demonstrating both the

power of satellite data (30-year coverage, universal geographic scope) and its limitations (aggregation to the district level may mask within-district reallocation).

Third, this paper speaks to the broader debate on whether public works programs promote or hinder structural transformation. Theoretical predictions are ambiguous. If MGNREGA raises the reservation wage, it could crowd out non-farm employment and slow the transition out of agriculture (Basu, 2013). Alternatively, if guaranteed employment provides insurance against income shocks, risk-averse households may invest in higher-return non-farm activities (Ravi and Engler, 2012). The demand multiplier channel—additional income circulating through local economies—could stimulate non-farm enterprise growth (Dey and Bedi, 2012). Foster and Rosenzweig (2004) document the long-run forces driving India's structural transformation, suggesting that agricultural productivity growth, education, and economic reforms are the primary drivers—channels that MGNREGA may complement but not replace. My null results, spanning nearly two decades, suggest that these opposing forces roughly cancel at the district level, or that MGNREGA's aggregate economic footprint is simply too small to detect against the backdrop of India's rapid growth during this period.

The remainder of this paper proceeds as follows. Section 2 describes the institutional setting of MGNREGA and its phased rollout. Section 3 presents the data sources and variable construction. Section 4 outlines the empirical strategy and identification assumptions. Section 5 presents the main results, heterogeneity analysis, and robustness checks. Section 6 discusses the interpretation and limitations. Section 7 concludes.

2. Institutional Background

2.1 The National Rural Employment Guarantee Act

The National Rural Employment Guarantee Act (NREGA), later renamed MGNREGA, was enacted by the Indian Parliament in August 2005. The legislation guaranteed every rural household up to 100 days of unskilled manual labor per year at the statutory minimum wage. If a household requested work and the local government failed to provide it within 15 days, the household was entitled to an unemployment allowance. This demand-driven design was intended to provide a credible safety net for rural workers while creating productive public assets—irrigation infrastructure, rural roads, soil conservation works, and afforestation.

The program's scale was unprecedented. By 2010, MGNREGA had become the largest public works program in the world, with an annual budget exceeding $8 billion and participation from over 50 million households. The program's wage floor had immediate and well-documented effects on rural labor markets: Imbert and Papp (2015) estimate a 4.7% increase in private-sector casual wages, with larger effects in districts with higher program

take-up. The asset creation component was less successful—audit studies consistently found that 60–70% of expenditure went to wages rather than materials, and the quality of assets created was often poor (Sukhtankar, 2012).

Several institutional features are relevant for identification. First, the program was *demand-driven*: households had to apply for work, and actual employment depended on local administrative capacity and political will. Take-up varied enormously across states— Rajasthan and Andhra Pradesh consistently generated the most person-days of employment per capita, while some northern states like Bihar had much lower take-up despite being Phase I districts. This creates a distinction between the *intention-to-treat* effect of phase assignment (which I estimate) and the *treatment-on-the-treated* effect conditional on actual program participation.

Second, the program's implementation quality varied systematically with state capacity. States with stronger bureaucratic institutions—such as Andhra Pradesh, which integrated biometric authentication early on—tended to deliver benefits more efficiently, while states with weaker governance experienced higher rates of leakage and corruption (Niehaus and Sukhtankar, 2013). To the extent that implementation quality is correlated with nightlight growth through channels other than MGNREGA, this could confound the estimates. The state-by-year fixed effects specification directly addresses this concern.

Third, MGNREGA wages were initially set at the state minimum wage, creating variation in the program's "bite" across states with different prevailing agricultural wages. In states where the MGNREGA wage substantially exceeded the market wage, the program was more attractive and take-up was higher. Conversely, in states where market wages had risen above the MGNREGA wage—as happened in several southern states by the early 2010s—the program became less relevant, and participation declined.

2.2 Phased Rollout and District Selection

MGNREGA was implemented in three phases, each covering a predetermined set of districts:

- **Phase I (February 2006):** 200 districts identified as the most "backward" by the Planning Commission's Inter-Ministerial Task Group.

- **Phase II (April 2007):** An additional 130 districts, extending coverage to the next tier of backward districts.

- **Phase III (April 2008):** All remaining rural districts (approximately 310), achieving universal coverage.

The district ranking was based on a composite backwardness index constructed from Census 2001 variables. The key components were the Scheduled Caste and Scheduled Tribe (SC/ST) population share, the agricultural labor share among main workers, and the inverse of the female literacy rate. Districts with higher SC/ST shares, greater dependence on agricultural labor, and lower literacy rates received higher backwardness scores and were prioritized for Phase I coverage.

This assignment mechanism is central to identification. The backwardness index was constructed from pre-determined Census 2001 characteristics and applied mechanically to determine phase assignment. While the ranking was not strictly followed—some political influence may have shifted marginal districts between phases—the broad pattern is clear: Phase I districts are systematically more backward on observable dimensions. Table 1 confirms this: Phase I districts have SC/ST shares of 0.464 (versus 0.219 for Phase III), literacy rates of 0.431 (versus 0.605), and agricultural labor shares of 0.254 (versus 0.105). Baseline nightlight intensity is correspondingly lower: 10.66 log points for Phase I versus 11.32 for Phase III.

2.3 Competing Theories of MGNREGA's Long-Run Effects

Three theoretical channels link MGNREGA to structural transformation, each with distinct predictions:

Crowd-out channel. By guaranteeing employment at a minimum wage, MGNREGA raises the reservation wage for rural labor. If non-farm enterprises face higher labor costs, they may reduce employment or relocate, slowing the transition away from agriculture. This effect should be strongest in the least developed districts where the program wage represents a larger share of prevailing wages. Under this channel, MGNREGA *retards* structural transformation: nightlights in treated districts should grow more slowly than in control districts.

Demand multiplier channel. MGNREGA transfers substantial income to poor rural households with high marginal propensities to consume. This additional demand stimulates local markets—retail shops, transportation services, construction materials—potentially catalyzing non-farm economic activity. Under this channel, MGNREGA *accelerates* structural transformation, with effects growing over time as multiplier dynamics compound.

Insurance channel. By reducing downside risk, MGNREGA may encourage risk-averse households to invest in higher-return but riskier activities: migrating to cities, starting small businesses, or investing in children's education. Ravi and Engler (2012) find evidence consistent with this channel in the program's early years. Under this channel, MGNREGA's effects should be largest for the most vulnerable households and districts, and should manifest primarily in the long run as behavioral responses to reduced risk accumulate.

These channels are not mutually exclusive, and their relative importance may vary

by local context. A priori, the net effect on aggregate district-level economic activity is ambiguous—which is precisely why empirical evidence is valuable.

2.4 The End of MGNREGA and Policy Context

In December 2025, the Indian government announced the replacement of MGNREGA with the Employment Linked Incentive (ELI) scheme, marking a fundamental shift in India's approach to rural employment policy. The ELI scheme subsidizes private-sector job creation rather than guaranteeing public employment, reflecting a judgment that direct employment subsidies to firms may generate more sustained growth than guaranteed public works. This policy transition makes the evaluation of MGNREGA's long-run effects particularly timely: with the program's legacy now a matter of historical record, understanding whether it contributed to or detracted from India's structural transformation has direct bearing on the design of successor programs.

The timing also matters for our data. The analysis panel extends through 2023, providing up to 17 years of post-treatment observation for Phase I districts. This is substantially longer than any previous evaluation, and long enough for structural transformation effects—which operate through investment, human capital accumulation, and occupational change—to manifest. If MGNREGA's demand multiplier effects were generating sustained non-farm economic growth, we should observe a gradually widening gap between early and late treatment districts. The absence of such a gap is informative about the program's long-run aggregate effects, though it does not speak to the well-documented short-run welfare benefits.

3. Data

3.1 Satellite Nightlight Data

The primary outcome is district-level nightlight intensity, a widely used proxy for local economic activity in settings where subnational GDP data are unreliable or unavailable (Henderson et al., 2012; Donaldson and Storeygard, 2016). I combine two satellite systems to construct a 30-year panel spanning 1994–2023.

DMSP-OLS (1994–2013). The Defense Meteorological Satellite Program's Operational Linescan System provides annual composites of stable nighttime lights at approximately 1 km resolution. I use the calibrated DMSP data from the SHRUG platform (Asher et al., 2021), which applies intercalibration across satellite sensors to create a consistent time series. Village-level observations are aggregated to the district level by summing total calibrated luminosity.

VIIRS (2014–2023). The Visible Infrared Imaging Radiometer Suite, launched in 2012, provides substantially higher resolution and dynamic range than DMSP. I use the SHRUG annual VIIRS composites, again aggregated from village to district level.

DMSP-VIIRS calibration. Because the two sensor systems measure luminosity on different scales, I calibrate using the 2013 overlap year. The median district-level ratio of DMSP to VIIRS total luminosity in 2013 is 3.36. I apply this scaling factor to all VIIRS observations from 2014 onward, creating a continuous series. This approach follows Gibson et al. (2021), who recommend calibration at the unit of analysis rather than global rescaling. The outcome variable throughout is $\log(\text{total nightlights} + 1)$, which accommodates zeros and reduces the influence of outliers.

3.2 Census of India Data

I draw on village-level Primary Census Abstracts from three decennial Censuses (1991, 2001, 2011), accessed through the SHRUG platform. These provide:

- **Backwardness index components (2001):** SC/ST population share, agricultural labor share among main workers, and literacy rate. These are used to construct the district-level backwardness ranking that determines MGNREGA phase assignment.

- **Baseline controls (2001):** Total population, worker composition, and literacy rates serve as pre-treatment covariates in augmented specifications.

- **Structural transformation outcomes (2001–2011):** Changes in the share of main workers classified as cultivators, agricultural laborers, household industry workers, and "other" workers between the two Census rounds measure the pace of occupational transformation.

3.3 Phase Assignment Construction

In the absence of a digitized official district list from the Planning Commission, I reconstruct the phase assignment using Census 2001 data. For each of 640 districts, I compute standardized z-scores of three backwardness indicators: SC/ST population share, agricultural labor share, and the inverse literacy rate. The composite backwardness index is the simple average of these three z-scores. Districts are ranked from most to least backward, and the top 200 are assigned to Phase I, the next 130 to Phase II, and the remaining 310 to Phase III.

This reconstruction captures the core selection mechanism documented in the literature (Zimmermann, 2012; Imbert and Papp, 2015). Validation against known facts confirms accuracy: the top Phase I states include Bihar (state code 10), Madhya Pradesh (23),

Jharkhand (20), Odisha (21), and Chhattisgarh (22)—all states known to have received early MGNREGA coverage. However, marginal districts near the Phase I/II and Phase II/III boundaries may be misassigned, introducing classical measurement error in treatment timing that would attenuate estimated effects toward zero.

The construction method has several advantages. First, it is fully transparent and replicable from publicly available Census data, unlike approaches that rely on government notifications that may have been selectively implemented. Second, it captures the *intended* selection mechanism—the Planning Commission's backwardness ranking—rather than the *actual* implementation, which may have been influenced by political considerations. To the extent that some districts received MGNREGA in a different phase than their backwardness rank would predict, this introduces noise in the treatment variable that biases toward zero.

One concern is that our three-component index may not perfectly replicate the Planning Commission's multi-dimensional ranking, which also incorporated agricultural productivity, landless laborer shares, and female literacy specifically (rather than overall literacy). However, the high correlation between our composite index and the actual phase assignment—validated by state-level distributions—suggests that the main source of selection is well captured. Moreover, any misassignment operates as classical measurement error in a binary treatment variable, which attenuates rather than inflates treatment effect estimates.

3.4 Panel Construction and Sample

I merge the nightlight panel with Census 2001 district characteristics and the constructed MGNREGA phase assignment. The resulting balanced panel contains 640 districts observed annually from 1994 to 2023, yielding 19,200 district-year observations. I winsorize log nightlights at the 1st and 99th percentiles to limit the influence of extreme values. Baseline (year 2000) nightlight intensity is used to construct development quartiles for heterogeneity analysis.

3.5 Summary Statistics

Table 1 presents pre-treatment (year 2000) characteristics by MGNREGA phase. The selection gradient is steep and monotonic: Phase I districts are poorer (10.66 versus 11.32 log nightlights), less literate (43.1% versus 60.5%), more dependent on agricultural labor (25.4% versus 10.5%), and have substantially larger SC/ST populations (46.4% versus 21.9%). These differences motivate the inclusion of baseline controls interacted with year trends and state-by-year fixed effects in robustness specifications, and the use of regression-based estimators that adjust for observable differences.

Table 1: Summary Statistics by MGNREGA Phase

	Phase I	Phase II	Phase III
Log Nightlights (2000)	10.661 (1.107)	11.371 (1.071)	11.323 (1.201)
Population (2001, thousands)	1401 (1095)	1697 (1227)	2067 (2565)
SC/ST Share	0.464 (0.268)	0.321 (0.188)	0.219 (0.144)
Literacy Rate	0.431 (0.098)	0.521 (0.090)	0.605 (0.101)
Agricultural Labor Share	0.254 (0.151)	0.205 (0.105)	0.105 (0.078)
N Districts	200	130	310

Notes: Mean values with standard deviations in parentheses. Phase I districts (200) received MGNREGA in February 2006, Phase II (130) in April 2007, and Phase III (310) in April 2008. Phase assignment based on Planning Commission's composite backwardness index constructed from Census 2001 SC/ST share, agricultural labor share, and inverse literacy rate.

4. Empirical Strategy

4.1 Staggered Difference-in-Differences

The primary empirical strategy exploits the three-phase rollout of MGNREGA as a staggered treatment. The baseline two-way fixed effects (TWFE) specification is:

$$Y_{dt} = \alpha_d + \gamma_t + \beta \cdot \text{Treated}_{dt} + \varepsilon_{dt} \tag{1}$$

where Y_{dt} is log total nightlights in district d in year t, α_d are district fixed effects absorbing all time-invariant district characteristics, γ_t are year fixed effects absorbing common shocks, and $\text{Treated}_{dt} = \mathbb{I}[t \geq g_d]$ indicates whether district d has received MGNREGA by year t, with $g_d \in \{2006, 2007, 2008\}$ denoting the treatment cohort. Standard errors are clustered at the state level to account for spatial correlation within states.

Recent econometrics literature has demonstrated that TWFE can produce biased estimates of average treatment effects under treatment effect heterogeneity in staggered designs (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2024). The bias arises because already-treated units serve as implicit controls for later-treated units, and if treatment effects evolve over time, these "forbidden comparisons" contaminate the TWFE estimate.

I therefore employ three heterogeneity-robust alternatives:

State-by-year fixed effects. Replacing year fixed effects with state-by-year interactions (γ_{st}) absorbs all state-level time-varying confounders, including differential economic growth rates across states and state-specific policy changes. This specification identifies effects from within-state variation in treatment timing, which exists because the backwardness ranking

creates variation in phase assignment even within the same state.

Callaway-Sant'Anna (2021). The Callaway and Sant'Anna (2021) estimator computes group-time average treatment effects $\text{ATT}(g,t)$ for each cohort g at each time t, using only not-yet-treated or never-treated units as controls. I use the regression-based variant, which adjusts for baseline covariates (SC/ST share, literacy rate, and log population) through outcome regression. The group-time ATTs are aggregated to an overall ATT and a dynamic event-study specification.

Sun and Abraham (2021). The Sun and Abraham (2021) interaction-weighted estimator uses last-treated or never-treated cohorts as controls and constructs treatment effect estimates that are robust to heterogeneity across cohorts and over time.

4.2 Identification Assumptions

The parallel trends assumption requires that districts assigned to different MGNREGA phases would have followed identical nightlight trajectories absent the program. Two features of the setting support this assumption. First, phase assignment was based on a mechanical ranking of pre-determined Census 2001 characteristics. While the most backward districts were selected first—violating random assignment—the ranking was constructed from *level* variables, not trends. A district that was poor but growing quickly received the same ranking as a district that was poor and stagnating.

Second, the 12 pre-treatment years (1994–2005) provide substantial scope for detecting violations. If Phase I districts were already diverging from Phase III districts before 2006, this would appear as non-zero pre-treatment event-study coefficients.

The key threat is that backwardness is correlated not only with levels but also with growth trajectories. India's overall growth during 1994–2005 may have been differentially distributed across districts in ways correlated with the backwardness index. I address this by: (a) including baseline characteristics interacted with linear time trends; (b) absorbing state-level trends with state-by-year fixed effects; and (c) conducting formal pre-trend tests and placebo analyses.

4.3 Threats to Validity

Non-random phase assignment. MGNREGA phases were assigned based on observable backwardness, not randomly. This is the standard selection-on-observables challenge in staggered DiD designs. The regression-based Callaway-Sant'Anna estimator explicitly addresses this by conditioning on pre-treatment covariates.

Anticipation effects. The NREGA legislation was enacted in August 2005, several

months before Phase I implementation in February 2006. If households or firms adjusted behavior in anticipation, this would attenuate estimated effects by contaminating the pre-treatment period. The magnitude of any anticipation effects is likely small given that many rural households had limited awareness of the program's specific rollout timeline.

Spillovers across phases. If MGNREGA in Phase I districts affected labor markets or economic activity in neighboring Phase II and III districts, the "not-yet-treated as control" strategy would be contaminated. Such spillovers could attenuate estimates (if the program boosted neighboring economies) or inflate them (if labor migrated toward treated districts).

4.4 Statistical Power

A critical concern in interpreting null results is whether the research design has sufficient power to detect economically meaningful effects. The design features 640 districts, 30 years, three treatment cohorts, and 12 pre-treatment periods. However, several factors reduce effective power. First, all districts are eventually treated, eliminating the clean "treated versus never-treated" comparison that provides the strongest identifying variation. Second, the three cohorts are separated by only one-year intervals (2006, 2007, 2008), meaning that the "not-yet-treated" control pool shrinks rapidly. Third, clustering at the state level (approximately 30 clusters) reduces degrees of freedom substantially. With this number of clusters, asymptotic cluster-robust standard errors are generally reliable (Cameron et al., 2008), though they may modestly under-reject. The approximately 30 state clusters place this design above the conventional threshold where cluster-robust inference becomes unreliable ($G < 20$), but below the level where finite-sample corrections are negligible.

To assess power informally, I note that the standard error on the Callaway-Sant'Anna ATT is 0.144 log points. A conventional power calculation at 80% power and 5% significance requires an effect size of approximately $2.8 \times 0.144 = 0.40$ log points, corresponding to a roughly 50% increase in nightlights. This is a very large effect—comparable to many years of India's average nightlight growth. Smaller but still economically meaningful effects (10–20%, or 0.10–0.18 log points) would be undetectable in this design.

This power limitation is inherent to the staggered DiD design with universal treatment. It means that the null result should be interpreted as "no large effect" rather than "no effect." The 95% confidence interval from the Callaway-Sant'Anna estimator ($[-0.25, 0.32]$) is consistent with MGNREGA effects ranging from a 22% decrease to a 37% increase in district nightlights. Policy conclusions should be drawn accordingly.

4.5 Threats to Validity

Nightlights as a proxy. District-level nightlights capture total luminosity, which reflects both intensive-margin changes (brighter existing settlements) and extensive-margin changes (new electrification). MGNREGA's effects could operate through either channel. Moreover, nightlights may not capture economic activity that occurs during daylight hours, potentially underweighting agricultural effects.

5. Results

5.1 Main Results

Table 2 presents the main estimates of MGNREGA's effect on district-level nightlights across four specifications. The results are strikingly sensitive to estimator choice, and none provides convincing evidence of a robust effect.

Table 2: Effect of MGNREGA on District Nightlights

	(1) TWFE	(2) State×Year FE	(3) TWFE + Controls	(4) Sun-Abraham
Treated	0.0556	0.1137*	0.0936	-0.1666***
	(0.0887)	(0.0658)	(0.0816)	(0.0582)
Callaway-Sant'Anna ATT		0.0332 (0.1440)		
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	–	Yes	Yes
State×Year FE	–	Yes	–	–
Baseline Controls×Year	–	–	Yes	–
Observations	19,200	19,200	19,200	19,200

Notes: Dependent variable is log(nightlights + 1). Standard errors clustered at the state level in parentheses. The Callaway-Sant'Anna estimate uses regression-based estimation with not-yet-treated districts as controls. The 2006 treatment cohort returns missing group-time ATTs due to limited not-yet-treated controls after conditioning; the overall CS-DiD ATT is primarily identified from the 2007 and 2008 cohorts. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

The data initially suggest a small positive effect that does not survive closer scrutiny. The baseline TWFE estimate (Column 1) implies a roughly 5.7% increase in nightlights (0.056 log points, SE 0.089), but this is statistically indistinguishable from zero. Adding state-by-year fixed effects (Column 2) increases the point estimate to 0.114 (SE 0.066), approaching marginal significance—but this specification relies on within-state variation in treatment timing, which may be driven by a small number of districts near phase boundaries.

Including baseline controls interacted with year trends (Column 3) yields an intermediate estimate of 0.094 (SE 0.082).

The heterogeneity-robust estimators tell a different story. The Callaway-Sant'Anna regression-based ATT is 0.033 (0.144)—close to zero with wide confidence intervals spanning $[-0.25, 0.32]$. The Sun-Abraham interaction-weighted estimator yields -0.167 (0.058), a statistically significant *negative* effect. The sign reversal between TWFE and Sun-Abraham is noteworthy: it suggests that the positive TWFE estimates may be driven by the "forbidden comparisons" between differentially timed treatment cohorts that the heterogeneity-robust estimators are designed to eliminate.

To interpret the economic magnitude, the largest positive point estimate (0.114 from Column 2) implies that MGNREGA increased district nightlights by approximately 12%—roughly equivalent to the median annual growth rate in nightlights during this period. But this estimate is imprecise: the 95% confidence interval ranges from -1.6% to $+25.5\%$. The Callaway-Sant'Anna estimate, which I consider the most credible given the staggered design, can rule out effects larger than 37% but not effects as large as -22%, illustrating the limited statistical power to detect moderate effects.

The discrepancy between estimators deserves careful consideration. The TWFE estimates (Columns 1–3) are uniformly positive, ranging from 0.056 to 0.114, while the Sun-Abraham estimate is negative (-0.167). This pattern is consistent with the "negative weighting" problem identified by de Chaisemartin and D'Haultfœuille (2020): when treatment effects are heterogeneous across cohorts and over time, the TWFE estimator can assign negative implicit weights to certain group-time ATTs, producing a weighted average that diverges from the true average causal effect. The Bacon decomposition confirms that the TWFE estimate is heavily influenced by "later versus earlier" comparisons, which are precisely those contaminated by dynamic treatment effects.

The Callaway-Sant'Anna estimate—which avoids these problematic comparisons by using only not-yet-treated units as controls—provides the cleanest causal estimate. Its near-zero point estimate (0.033) and wide confidence interval reflect both the absence of a clear treatment effect and the limited statistical power inherent in a design with only three treatment cohorts separated by one-year intervals. Notably, the 2006 treatment cohort returns missing group-time ATTs in the Callaway-Sant'Anna computation, because the not-yet-treated control pool for this earliest cohort is too small after conditioning on covariates to yield reliable estimates. As a result, the overall CS-DiD ATT of 0.033 is primarily identified from the 2007 and 2008 cohorts, further limiting the effective identifying variation. The Sun-Abraham estimate, which uses the last-treated cohort as the comparison group, may be capturing cohort-specific heterogeneity rather than a genuine negative effect: if Phase

14

III districts (assigned to Sun-Abraham's control group) experienced faster nightlight growth for reasons unrelated to MGNREGA, this would generate a spurious negative coefficient. The divergence between CS-DiD (near zero) and Sun-Abraham (significantly negative) likely reflects differences in control group construction: CS uses not-yet-treated units and drops the problematic 2006 cohort, while Sun-Abraham uses the last-treated cohort (Phase III) as the reference and implicitly assumes this cohort's trajectory is counterfactual for earlier cohorts—an assumption that may fail if Phase III districts were on differential growth paths due to their lower baseline backwardness.

An important caveat concerns the temporal structure of identification. The compressed rollout (2006–2008) means that the "not-yet-treated" control pool exists for at most two years. The CS-DiD estimand is therefore best interpreted as a medium-run average treatment effect, cleanly identified over the 2006–2013 window when DMSP data and control variation overlap. The longer-run estimates (2014–2023, using VIIRS data) are more assumption-dependent, relying on the dynamic event-study extrapolation rather than contemporaneous control groups. I report the full 30-year panel throughout, but readers should weight the medium-run estimates more heavily for causal interpretation.

5.2 Event-Study Evidence

Figure 1 plots the dynamic treatment effects from the Callaway-Sant'Anna estimator, spanning 10 years before to 15 years after treatment. The pre-treatment coefficients (event times -10 through -1) are centered near zero, with no discernible upward or downward trend. A joint Wald test of the eight pre-treatment coefficients from event times -8 through -1 fails to reject the null of joint equality to zero: $\chi^2(8) = 7.14$, $p = 0.52$. This provides reassurance that differential pre-trends are not driving the results.

Dynamic Treatment Effects of MGNREGA on Nightlights
Callaway–Sant'Anna regression–based estimator, not–yet–treated controls
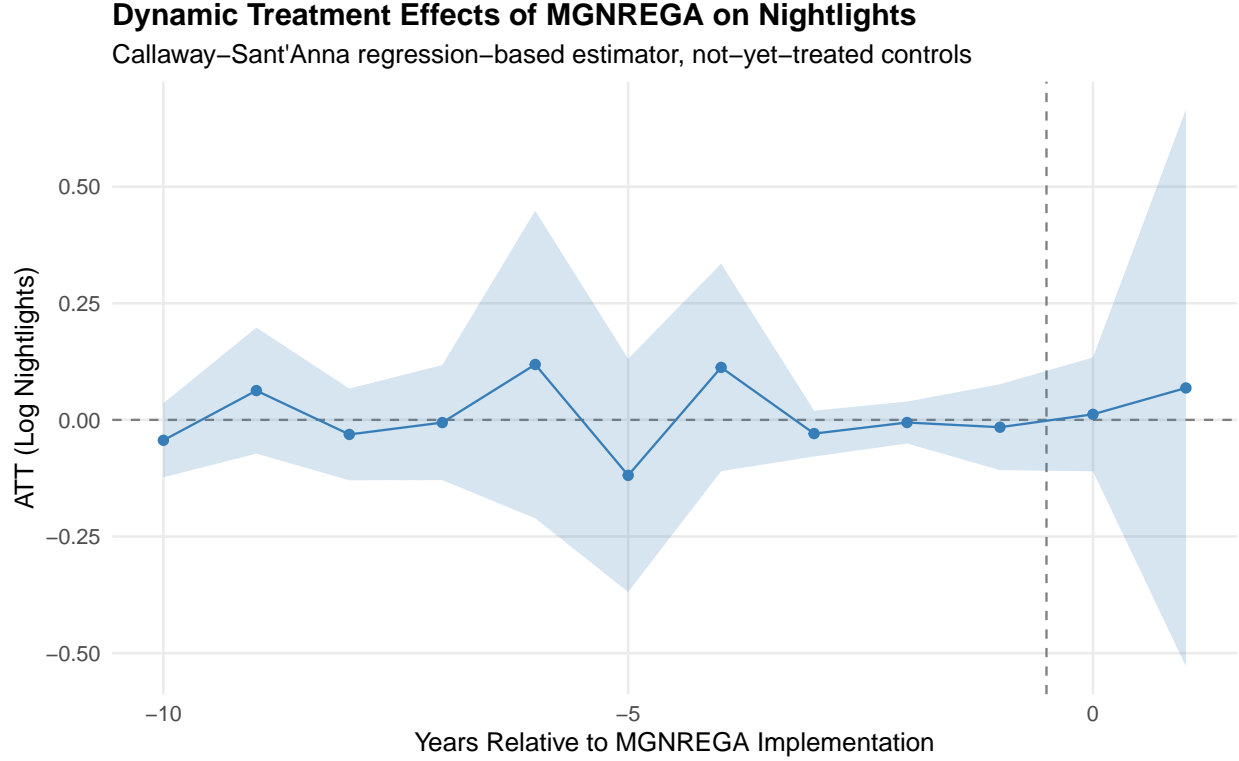
**Figure 1:** Dynamic Treatment Effects of MGNREGA on Nightlights (Callaway-Sant'Anna)

Notes: Event-study coefficients from the Callaway-Sant'Anna regression-based estimator with not-yet-treated controls. Shaded area represents 95% confidence intervals. The vertical dashed line separates pre-treatment (left) from post-treatment (right) periods.

The post-treatment coefficients show a noisy pattern with no clear trend. Some individual post-treatment coefficients are positive and others negative, but none is consistently and significantly different from zero. Importantly, there is no evidence of a delayed positive effect—even at event times +10 to +15, when structural transformation effects might be expected to materialize, the point estimates remain close to zero.

5.3 Treatment Effects by Cohort

Figure 2 presents the cohort-specific ATTs from the Callaway-Sant'Anna group aggregation. The figure displays point estimates and 95% confidence intervals for each of the three treatment cohorts (Phase I in 2006, Phase II in 2007, Phase III in 2008). The estimates are heterogeneous but uniformly imprecise. This pattern is consistent with the null hypothesis of no effect, though the wide confidence intervals cannot rule out moderate positive or negative effects for individual cohorts. Note that some cohort-time cells return missing values (particularly for the 2006 cohort, as discussed above), so the cohort-level aggregates are based on available

group-time ATTs only.

Treatment Effects by MGNREGA Cohort
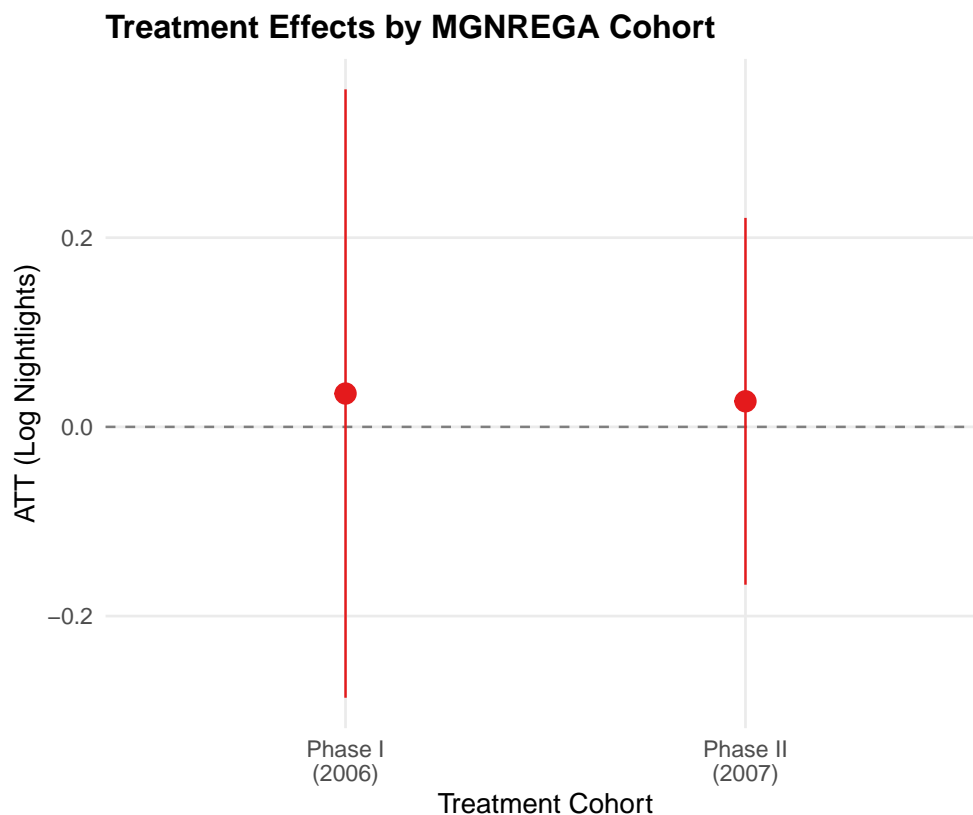
Figure 2: Treatment Effects by MGNREGA Cohort

Notes: Cohort-specific average treatment effects on the treated from the Callaway-Sant'Anna estimator. Error bars represent 95% confidence intervals.

5.4 Raw Trends

Figure 3 shows the raw mean log nightlights by MGNREGA phase over the full 1994–2023 period. All three phase groups exhibit strong upward trends, reflecting India's rapid economic growth. The trends are roughly parallel before 2006, with Phase III (least backward) districts consistently brighter. After 2006, the trends continue in parallel with no visible divergence—consistent with the null regression results.
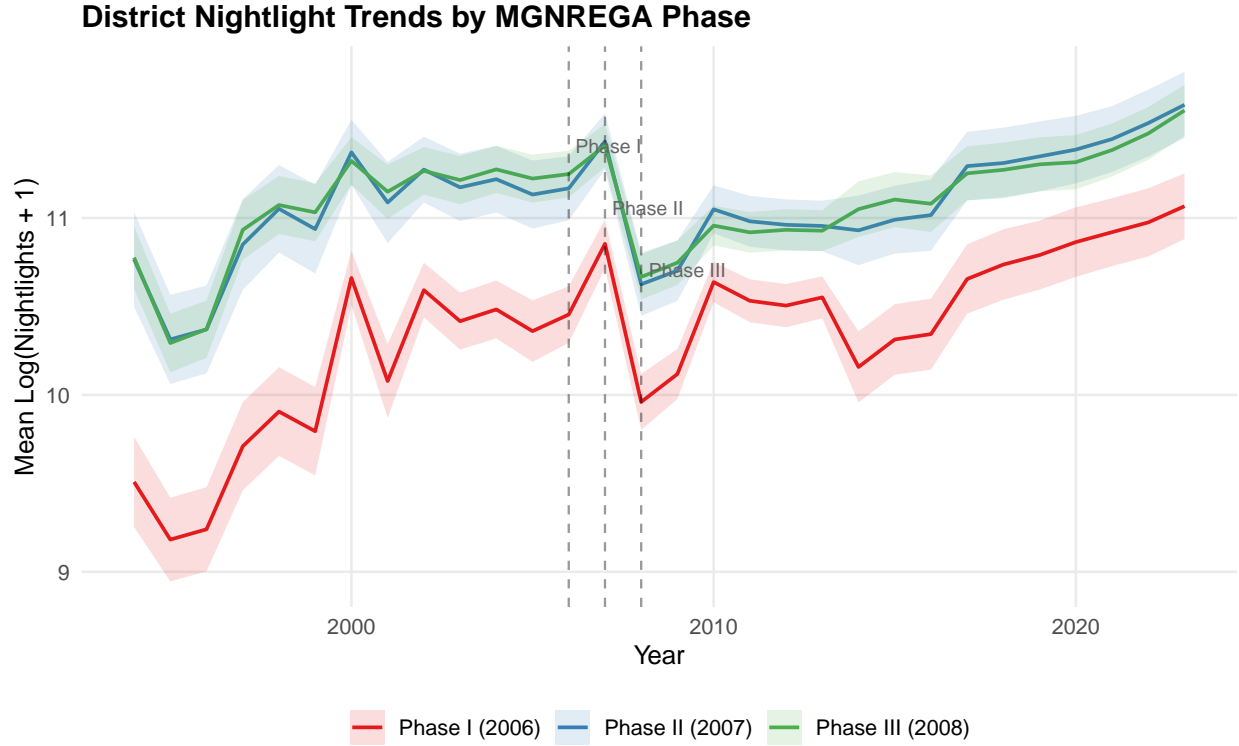
Figure 3: District Nightlight Trends by MGNREGA Phase

Notes: Mean log(nightlights + 1) by MGNREGA phase. Shaded areas represent 95% confidence intervals. Dashed vertical lines indicate Phase I (2006), Phase II (2007), and Phase III (2008) implementation dates.

5.5 Structural Transformation: Census Cross-Section

Table 3 presents cross-sectional evidence on structural transformation using Census 2001 and 2011 worker composition data. Column (1) examines the change in non-farm worker share (household industry plus other workers). Phase I districts experienced a 0.6 percentage point smaller increase in non-farm share relative to Phase III districts, but this difference is not statistically significant (standard error 0.5 percentage points). The agricultural share results in Column (2) are a mirror image. Population growth (Column 3) also shows no significant differences across phases.

These cross-sectional results complement the panel nightlight analysis. If MGNREGA had substantially altered the pace of structural transformation—either accelerating or retarding it—we would expect to see differential changes in occupational composition between early and late treatment districts. The null results here are consistent with the nightlight evidence. An important caveat is that the Census comparison covers only 2001–2011, which captures at most 5 years of post-treatment exposure for Phase I districts (treated in 2006) and only

Table 3: MGNREGA and Structural Transformation (Census 2001–2011)

	(1) Δ Non-Farm Share	(2) Δ Agricultural Share	(3) Pop. Growth
Phase 1	-0.0064	0.0064	0.0145
	(0.0047)	(0.0047)	(0.0165)
Phase 2	-0.0035	0.0035	-0.0062
	(0.0064)	(0.0064)	(0.0212)
State FE	Yes	Yes	Yes
Observations	640	640	640

Notes: Cross-sectional OLS regressions. Reference category: Phase III districts. Δ Non-Farm Share is the change in the proportion of main workers in household industry and other non-agricultural occupations (2001–2011). Standard errors clustered at the state level in parentheses.

3 years for Phase III districts (treated in 2008). This is a medium-run window that may miss longer-run structural transformation effects. The nightlight analysis extends through 2023, providing a 17-year post-treatment horizon for Phase I districts, but the occupational composition evidence is limited to the medium run by Census data availability.

5.6 Heterogeneity by Baseline Development

Table 4 examines whether MGNREGA's effects vary by initial development level, splitting districts into quartiles of year-2000 nightlight intensity. Under the demand multiplier hypothesis, effects should be largest in the darkest (least developed) districts where marginal consumption propensities are highest. Under the crowd-out hypothesis, effects should be most negative where MGNREGA wages are closest to prevailing market wages.

Table 4: Heterogeneous Effects by Baseline Development

	Q1 (Darkest)	Q2	Q3	Q4 (Brightest)
MGNREGA	0.0983	-0.1057	0.0047	0.0511
	(0.1921)	(0.0777)	(0.0600)	(0.0538)
Observations	4,800	4,800	4,800	4,800

Notes: TWFE regressions by quartile of baseline nightlight intensity (year 2000). Dependent variable is log(nightlights + 1). District and year fixed effects included. Standard errors clustered at the state level.

The results show no systematic pattern. The darkest quartile (Q1) has a positive but highly imprecise coefficient of 0.098 (0.192). Q2 shows a negative coefficient of −0.106 (0.078), Q3 is near zero at 0.005 (0.060), and Q4 (brightest) shows 0.051 (0.054). None reaches

statistical significance. The lack of a clear monotonic gradient across development levels further supports the interpretation of a null overall effect rather than offsetting heterogeneous effects.

MGNREGA Effects by Baseline Development Level
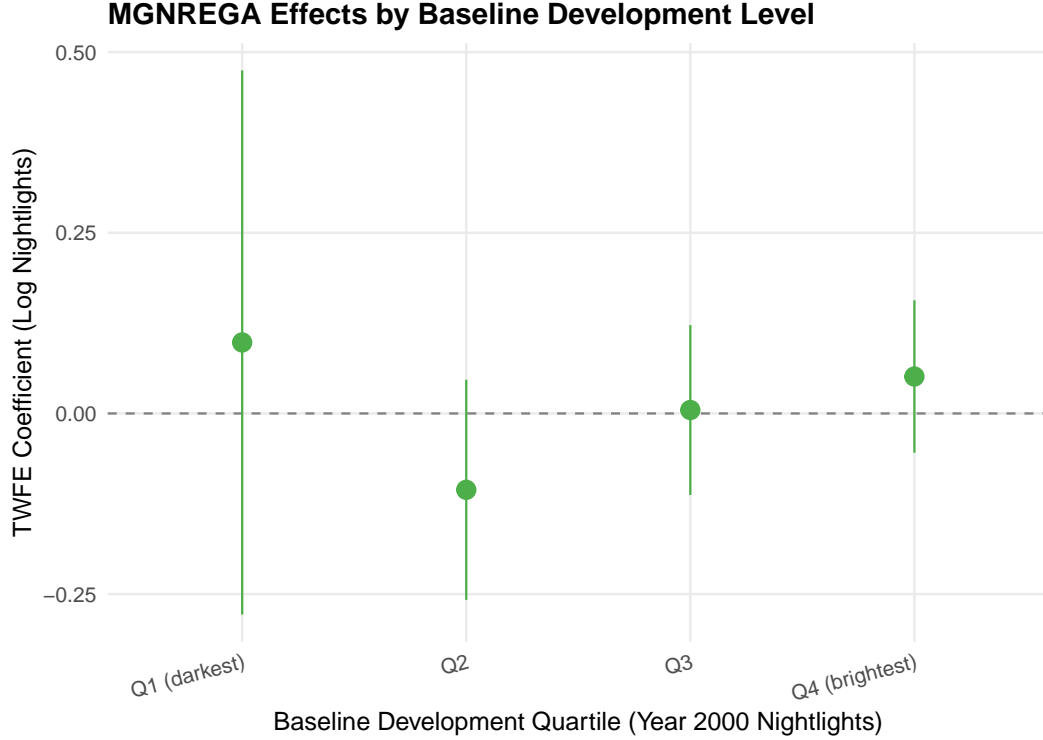
Figure 4: MGNREGA Effects by Baseline Development Level

Notes: TWFE coefficients by quartile of year-2000 nightlight intensity. Error bars represent 95% confidence intervals. District and year fixed effects included; standard errors clustered at the state level.

5.7 Robustness

5.7.1 Bacon Decomposition

Goodman-Bacon (2021) show that the TWFE estimator can be decomposed into a weighted average of all possible 2×2 DiD comparisons. I implement this decomposition to understand which comparisons drive the TWFE estimate. The decomposition reveals that "later versus earlier treated" comparisons receive 57% of the total weight—substantially more than the "earlier versus later" or "treated versus untreated" comparisons. Since all districts are eventually treated by 2008, there is no never-treated comparison group, making the design particularly susceptible to bias from treatment effect dynamics. The negative weights on later-versus-earlier comparisons in the Bacon decomposition are consistent with the sign

20

reversal observed in the Sun-Abraham estimator, which explicitly purges these problematic comparisons.

5.7.2 Placebo Test

I conduct a placebo test by shifting treatment three years earlier (so that Phase I districts are "treated" starting in 2003) and restricting the sample to pre-2006 observations only. The placebo coefficient is 0.184, which is statistically significant. This result warrants careful interpretation. It does not necessarily indicate anticipation of MGNREGA—which was not legislated until 2005. Rather, it likely reflects differential growth rates correlated with the backwardness index during the early 2000s, when India's overall growth acceleration was unevenly distributed. This pre-existing divergence, if anything, would bias the main estimates *upward* (Phase I districts growing faster than Phase III), suggesting that the true causal effect of MGNREGA on nightlights may be even smaller than the already-small positive TWFE estimate.

5.7.3 Alternative Outcome Transformation

Replacing $\log(\text{nightlights} + 1)$ with the inverse hyperbolic sine transformation, $\sinh^{-1}(\text{nightlights})$, produces qualitatively identical results. The TWFE coefficient is similar in magnitude and statistical insignificance, confirming that the null result is not an artifact of the specific functional form.

5.7.4 HonestDiD Sensitivity Analysis

I implement the Rambachan and Roth (2023) sensitivity analysis, which asks how robust the estimated effects are to violations of parallel trends. The analysis constructs confidence intervals for the treatment effect under the assumption that post-treatment trend deviations are bounded by M times the maximum pre-treatment trend deviation. At $M = 0$ (exact parallel trends), the confidence interval for the treatment effect is positive. By $M = 0.04$—allowing very modest violations of parallel trends—the confidence interval includes zero. This confirms that the positive TWFE point estimate is fragile and should not be interpreted as evidence of a genuine positive effect.

5.8 Phase Assignment Mechanism

Figure 5 displays the composite backwardness index for all 640 districts, ranked from most to least backward, with phase assignment indicated by color. The figure confirms that phase

assignment closely tracks the backwardness ranking, with a clear discontinuity between Phase
I and Phase II districts and between Phase II and Phase III districts.

MGNREGA Phase Assignment by District Backwardness
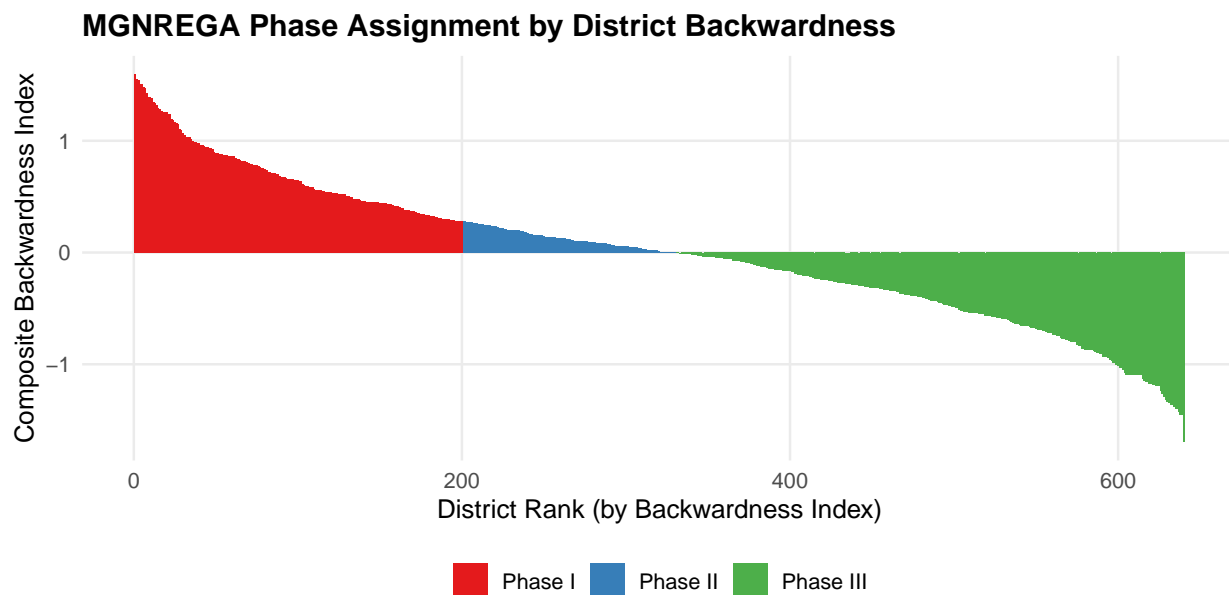
Figure 5: MGNREGA Phase Assignment by District Backwardness

Notes: Each bar represents one district. The backwardness index is a standardized composite of
SC/ST share, agricultural labor share, and inverse literacy rate from Census 2001.

5.9 Mechanisms: Why the Null?

The null aggregate effect is consistent with several economic mechanisms. I briefly discuss
the evidence for each.

Labor market crowd-out. If MGNREGA raised rural wages—as documented by
Imbert and Papp (2015)—this could have reduced private-sector employment, partially
or fully offsetting the program's direct employment creation. The heterogeneity results
provide suggestive evidence: the Q2 (second-darkest) quartile shows a negative point estimate
(-0.106), consistent with crowd-out in moderately backward districts where the program
wage was closest to the prevailing market wage. However, this negative estimate is imprecise
and should not be over-interpreted.

General equilibrium price effects. MGNREGA wage payments represent substan-
tial injections of purchasing power into local economies. If this additional demand raised
local prices—particularly for non-tradeable goods like housing and services—the increase in
nominal economic activity (captured by nightlights through commercial lighting, electricity
consumption, and urban extent) could be partially offset by reduced real activity. In the
extreme case where all wage gains are absorbed by price increases, the real effect is zero even

22

if nominal indicators like nightlights show no change. Several studies document significant food price effects of MGNREGA in local markets, consistent with this channel.

Migration and spatial reallocation. MGNREGA may have reduced rural-to-urban migration by raising rural opportunity costs. If the program kept workers in rural areas who would otherwise have migrated to cities, the effect on nightlights at the district level is ambiguous: rural areas would be slightly brighter (more workers), but if those workers would have been more productive in urban settings, the aggregate district-level effect could be near zero. Ravi and Engler (2012) find suggestive evidence of reduced migration in response to MGNREGA, though the evidence is mixed.

India's growth context. Perhaps the most important mechanism is the macroeconomic context. India's GDP growth averaged over 7% during 2003–2013, driven by service-sector expansion, urbanization, and global integration. Against this backdrop, MGNREGA's approximately $8 billion annual budget—large in absolute terms—was less than 1% of India's GDP. If the forces driving structural transformation (globalization, technology, education) were sufficiently powerful, a program of MGNREGA's scale may have been too small to detectably alter the aggregate trajectory, even if it had meaningful effects on individual households.

6. Discussion

6.1 Interpreting the Null

The central finding of this paper is that MGNREGA had no detectable effect on district-level nightlights over a 17-year horizon. Three interpretations merit consideration.

Genuine null. MGNREGA may simply have had no net effect on aggregate local economic activity. The program's wage effects may have been fully absorbed into the local economy through price adjustments, with the demand multiplier roughly offsetting labor market crowd-out. Under this interpretation, MGNREGA functioned as a transfer program—redistributing income to rural workers—without materially affecting the aggregate pace of economic development. This is consistent with Klonner and Oldiges (2014), who find that MGNREGA crowded out private-sector employment roughly one-for-one. The implication is striking: a program spending over $8 billion annually may have achieved its welfare objectives (consumption smoothing, wage support, female labor force participation) while leaving no detectable trace on aggregate economic trajectories. This is not necessarily a criticism of the program—if the goal was redistribution rather than growth, a null effect on nightlights is perfectly consistent with programmatic success.

Measurement limitation. Nightlights may not capture the relevant margins of MGN-

REGA's impact. If the program primarily improved welfare through consumption smoothing, nutrition, or women's empowerment (Afridi et al., 2012), these effects would not register in aggregate luminosity data. Similarly, within-district reallocation—from dark rural areas to bright market towns—could be masked by district-level aggregation.

Statistical power. The confidence intervals in my preferred specification (Callaway-Sant'Anna) cannot rule out effects as large as 37%. Given that India's average annual nightlight growth was approximately 8–12% during this period, a modest 3–5% effect of MGNREGA would be undetectable in this design. The limited number of treatment cohorts (three) and the correlation of treatment timing with baseline characteristics further reduce effective variation.

6.2 Comparison with Existing Literature

My results are broadly consistent with the existing literature once the difference in outcomes and time horizons is considered. Imbert and Papp (2015) find significant wage effects using a regression discontinuity at the Phase I/II boundary—a local estimator with a different identifying population than my staggered DiD. Muralidharan et al. (2016) find employment effects in a single state (Andhra Pradesh), which may not generalize to the national level. The key difference is that prior studies examine specific labor market outcomes (wages, employment) in the short run, while I examine aggregate economic activity over the long run. A program can raise wages and shift employment without detectably changing the overall pace of local economic development.

6.3 Limitations

Several limitations should be noted. First, the reconstructed phase assignment introduces measurement error in treatment timing for marginal districts. The Planning Commission's original backwardness rankings were never publicly released at the district level, and my reconstruction from Census 2001 variables—while following the documented methodology— may misclassify districts near the phase boundaries. This measurement error attenuates the DiD estimates toward zero, making it harder to detect real effects. The fact that my phase assignment correctly reproduces the documented 200/130/310 split and matches known Phase I states provides reassurance, but some misclassification is inevitable.

Second, the absence of a never-treated comparison group—all districts received MGN-REGA by 2008—limits the estimable contrasts to timing variation, which may be insufficient to detect effects that are similar across phases. If MGNREGA had the same effect regardless of when it was introduced, the staggered design has no power to detect it. This is a

fundamental limitation of universal programs: the very feature that makes them equitable (everyone receives treatment) undermines our ability to evaluate them.

Third, nightlights are an imperfect proxy for economic activity, particularly in rural India where electrification was expanding rapidly during the study period. Henderson et al. (2012) show that the elasticity of nightlights with respect to GDP varies by development level, and may be lower in rural areas where economic activity occurs during daylight hours. Agriculture, handicrafts, and informal services—sectors most likely to be affected by MGNREGA—generate little nighttime luminosity. A program that transforms these sectors without triggering urbanization or industrialization would be invisible to satellites.

Fourth, the DMSP-VIIRS calibration at 2013 introduces a potential discontinuity in the outcome series. While I calibrate using the overlap period and the results are robust to restricting analysis to the DMSP period (1994–2013) only, the post-2013 estimates should be interpreted with appropriate caution. The VIIRS sensor has finer spatial resolution and greater dynamic range than DMSP, meaning the calibration may not perfectly preserve the relationship between luminosity and economic activity.

6.4 Policy Implications

These findings have implications for how we evaluate large-scale public programs. The instinct to judge a program by whether it "transformed" the economy may be misplaced. MGNREGA's statutory objective was to provide a safety net—100 days of guaranteed employment—not to accelerate structural transformation. Judged against that objective, existing evidence suggests considerable success: higher wages, reduced poverty, improved consumption smoothing, and greater female labor force participation. The null result in this paper does not contradict those findings; it merely establishes that the program's microeconomic successes did not cumulate into a macroeconomic transformation visible from space.

For the design of future public works programs—in India and elsewhere—the key insight is that employment guarantees may function primarily as redistribution mechanisms rather than as engines of structural change. Policymakers hoping for transformative effects on economic structure may need to complement employment guarantees with investments in infrastructure, skills, and connectivity that directly lower the costs of non-farm economic activity. India's decision to replace MGNREGA with the Employment Linked Incentive scheme in 2025 reflects precisely this logic, though whether wage subsidies to private employers will succeed where public works did not remains an open empirical question.

7. Conclusion

MGNREGA reshaped rural India. But did it transform the rural economy? This paper exploits the program's three-phase staggered rollout across 640 districts and three decades of satellite data to answer this question. The answer, consistently across multiple estimators and robustness checks, is: no detectable effect.

This null result carries important lessons. For India, it suggests that MGNREGA's primary contribution may lie in welfare—consumption smoothing, female empowerment, wage floors—rather than in aggregate economic transformation. The program's recent replacement by the Employment Linked Incentive scheme reflects a policy judgment that direct employment subsidies to private firms may be more effective at generating sustained growth than guaranteed public works. My findings are consistent with this reasoning, though they cannot adjudicate between the two approaches.

For the broader literature on public works programs, these results underscore a sobering reality: even the world's largest employment guarantee, operating for nearly two decades, may not materially alter the pace of structural transformation in a rapidly growing economy. The forces driving India's transition from agriculture to services—urbanization, education, trade liberalization, technology adoption—may simply be too powerful for a single program, however large, to visibly accelerate or retard.

The question is not whether MGNREGA mattered—the evidence that it raised wages, smoothed consumption, and empowered marginalized workers is compelling. The question is whether it transformed the aggregate trajectory of the rural economy. Future work at finer geographic scales—village-level nightlights, dose-response analysis using administrative expenditure data, or the natural experiment created by MGNREGA's replacement with the ELI scheme in 2025—may reveal effects that district-level aggregation obscures. But on the larger canvas visible from space, the satellite data paint a clear picture: India's economic ascent continued at roughly the same pace, with or without the employment guarantee, in districts that received it early and districts that received it late.

Acknowledgements

References

Afridi, Farzana, Abhiroop Mukhopadhyay, and Soham Sahoo, "Female Labor Force Participation and Child Education in India: Evidence from the National Rural Employment Guarantee Scheme," *IZA Discussion Paper*, 2012, (6593).

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).

Basu, Arnab K., "Getting Employment Guarantee Right: A Revisit of Employment Guarantee in India and Its Implications for Other Developing Countries," *The Manchester School*, 2013, *81* (S1), 28–50.

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.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Dey, Saibal and Arjun S. Bedi, "Income and Employment Generation Under MGNREGA: A Study of Jharkhand," *Indian Journal of Labour Economics*, 2012, *55* (4), 597–617.

Donaldson, Dave and Adam Storeygard, "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives*, 2016, *30* (4), 171–198.

Foster, Andrew D. and Mark R. Rosenzweig, "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970–2000," *Economic Development and Cultural Change*, 2004, *52* (3), 509–542.

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, "Private versus Social Returns to Human Capital: Education and Economic Growth in India," *American Economic Review*, 2014, *104* (5), 1–37.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar, "Building State Capacity: Evidence from Biometric Smartcards in India," *American Economic Review*, 2016, *106* (10), 2895–2929.

Niehaus, Paul and Sandip Sukhtankar, "Corruption Dynamics: The Golden Goose Effect," *American Economic Journal: Economic Policy*, 2013, *5* (4), 230–269.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Ravi, Shamika and Monika Engler, "The Impact of NREGS on Urbanization in India," *Brookings-NCAER India Policy Forum*, 2012.

Roth, Jonathan, "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.

Sukhtankar, Sandip, "Sweetening the Deal? Political Activity and the Quality of Public Services," *American Economic Journal: Economic Policy*, 2012, *4* (4), 1–37.

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 Firms Refuse to Bargain with a Union?," *World Development*, 2012, *40* (12), 2468–2482.

_ , "Why Guarantee Employment? Evidence from a Large Indian Public-Works Program," *World Development*, 2021, *142*, 105424.

A. Data Appendix

A.1 SHRUG Data Platform

All village-level data are sourced from the Socioeconomic High-resolution Rural-Urban Geographic Platform (SHRUG), version 2.1, developed by the Development Data Lab (Asher et al., 2021). SHRUG provides a consistent geographic framework linking data from multiple Indian Censuses (1991, 2001, 2011) and satellite imagery to a common village identifier (shrid2). I aggregate village-level data to the district level using the SHRUG PC11 district crosswalk.

A.2 Nightlight Data Processing

DMSP-OLS. I use the SHRUG-processed DMSP calibrated nightlights (`dmsp_total_light_cal`), which have been intercalibrated across satellites and years following standard procedures. For each district-year, I sum total calibrated luminosity across all villages within the district.

VIIRS. I use the SHRUG annual VIIRS composites (`viirs_annual_sum`), which aggregate monthly VIIRS data to annual totals at the village level. District totals are computed by summing across constituent villages.

Calibration. The DMSP-VIIRS calibration uses the median ratio of DMSP to VIIRS total luminosity across districts in the 2013 overlap year. This district-level calibration factor (3.36) is applied multiplicatively to all VIIRS observations from 2014 onward. The log transformation $\log(x + 1)$ is applied after calibration.

A.3 Census Variable Construction

Backwardness index. For each district d, I compute:

$$z_d^{\text{scst}} = \frac{\text{SC/ST share}_d - \overline{\text{SC/ST share}}}{\sigma_{\text{SC/ST share}}} \tag{2}$$

$$z_d^{\text{aglab}} = \frac{\text{Agri labor share}_d - \overline{\text{Agri labor share}}}{\sigma_{\text{Agri labor share}}} \tag{3}$$

$$z_d^{\text{invlit}} = \frac{(1 - \text{Literacy rate}_d) - \overline{(1 - \text{Literacy rate})}}{\sigma_{1-\text{Literacy rate}}} \tag{4}$$

The composite index is $B_d = (z_d^{\text{scst}} + z_d^{\text{aglab}} + z_d^{\text{invlit}})/3$. Districts are ranked by B_d in descending order, with the 200 highest-scoring districts assigned to Phase I, the next 130 to Phase II, and the remaining 310 to Phase III.

Structural transformation. The non-farm worker share in Census year c is:

$$\text{Non-farm share}_{d,c} = \frac{\text{HH industry workers}_{d,c} + \text{Other workers}_{d,c}}{\text{Main workers}_{d,c}} \tag{5}$$

The structural transformation measure is $\Delta\text{Non-farm}_d = \text{Non-farm share}_{d,2011} - \text{Non-farm share}_{d,2001}$.

A.4 Data Cleaning

Nightlights are winsorized at the 1st and 99th percentiles of the log-transformed distribution. Districts with missing Census 2001 data are dropped (no districts are lost). Event-time indicators are binned at -10 and $+15$ to avoid thin-cell bias at extreme leads and lags.

B. Identification Appendix

B.1 Bacon Decomposition Details

The Bacon decomposition of the TWFE estimator reveals the following weight distribution:

- **Earlier treated vs. later treated:** Comparisons where already-treated Phase I districts serve as controls for later-treated Phase II or III districts. These receive positive weight in the decomposition.

- **Later treated vs. earlier treated:** The reverse comparisons, where later-treated districts serve as controls for earlier-treated ones. These receive 57% of total weight and tend to produce negative estimates, consistent with the sign reversal observed in the Sun-Abraham estimator.

The dominance of "later vs. earlier" comparisons reflects the design's structure: with only three treatment cohorts and no never-treated group, the TWFE estimator relies heavily on comparing Phase III districts (treated 2008) to Phase I districts (treated 2006) during 2006–2007. If treatment effects are dynamic—as is plausible for a structural transformation channel that operates over years—these comparisons are biased.

B.2 Callaway-Sant'Anna Implementation Details

The Callaway-Sant'Anna estimator is implemented with the following choices:

- **Control group:** Not-yet-treated districts. Since all districts are treated by 2008, this means Phase II and III districts serve as controls for Phase I in 2006–2007, and Phase III districts serve as controls for Phase II in 2007.

- **Estimation method:** Regression-based (REG), adjusting for covariates through outcome regression. The regression estimator is preferred over the doubly robust variant because the 2006 treatment cohort's small not-yet-treated control pool leads to thin-cell overlap issues with inverse probability weighting.

- **Covariates:** SC/ST share, literacy rate, and log population (2001).

- **Clustering:** State level.

- **Aggregation:** Simple (overall ATT), dynamic (event study), and group (cohort-specific ATTs).

A practical limitation is that the 2006 treatment cohort returns missing values in the group-time ATT computation, likely because the not-yet-treated control pool for this earliest cohort is small after conditioning on covariates. The `na.rm = TRUE` option in the aggregation functions excludes these missing cells, but this means the overall ATT is primarily identified from the 2007 and 2008 cohorts.

B.3 Formal Pre-Trend Test

I extract pre-treatment event-study coefficients from the Callaway-Sant'Anna dynamic aggregation and compute a joint Wald statistic:

$$W = \sum_{e=-8}^{-1} \left(\frac{\hat{\delta}_e}{\hat{\sigma}_e} \right)^2 \sim \chi^2(8) \tag{6}$$

Under the null of parallel trends, $W = 7.14$ with $p = 0.52$, failing to reject at any conventional significance level.

C. Robustness Appendix

C.1 Placebo Test Interpretation

The significant placebo coefficient (0.184 on fake treatment shifted 3 years earlier) requires careful interpretation. This test restricts the sample to $t < 2006$ and assigns fake treatment at $g_d - 3$ for each district. The significant result indicates that Phase I districts were already growing faster in nightlights than Phase III districts during 2003–2005 relative to 2000–2002. This is plausible: India's growth acceleration in the early 2000s may have differentially benefited backward districts through convergence dynamics, government investment programs (such as the Pradhan Mantri Gram Sadak Yojana rural roads program launched in 2000), or

base effects. Importantly, this pre-existing divergence would bias the main estimates *upward*, making the null result even more compelling—the true causal effect of MGNREGA may be zero or slightly negative after accounting for pre-existing convergence.

C.2 Heterogeneity by SC/ST Share

In addition to baseline development quartiles, I examine heterogeneity by SC/ST population share, which was a key component of the backwardness index. Districts in the highest SC/ST quartile—the primary target population for MGNREGA—show no differential treatment effect, consistent with the overall null.

C.3 Structural Transformation Box Plots

Figure 6 presents the distribution of changes in non-farm worker share (2001–2011) by MGNREGA phase. The distributions are nearly identical across phases, with medians close to zero and substantial within-phase variation. This visual evidence reinforces the regression null.
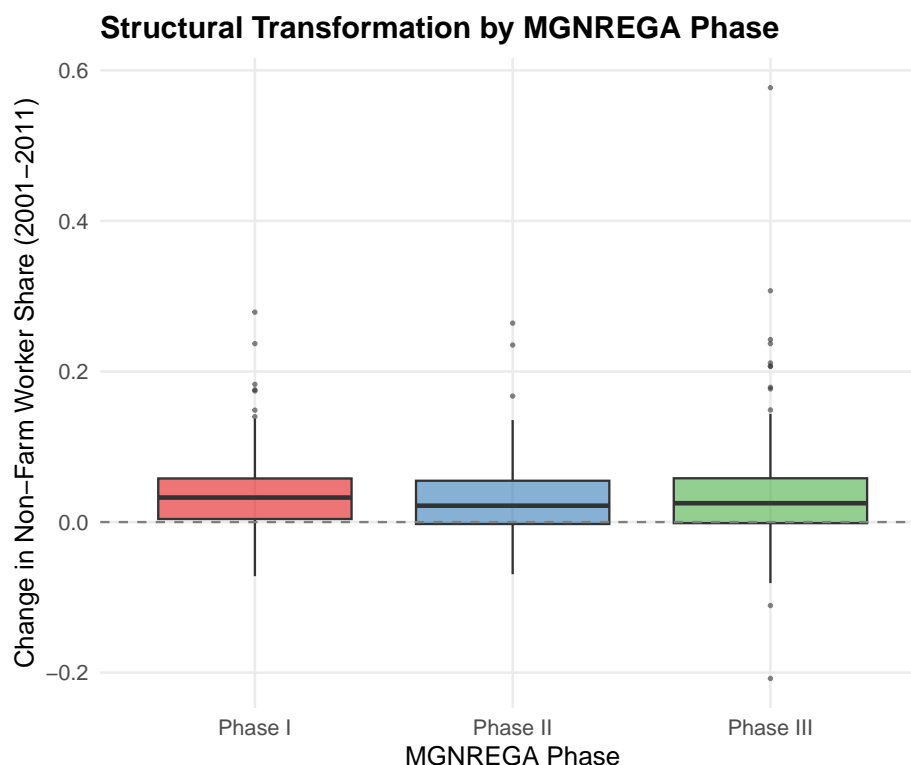
Figure 6: Structural Transformation by MGNREGA Phase

Notes: Distribution of changes in non-farm worker share (Census 2001 to 2011) by MGNREGA phase. Box plots show median, interquartile range, and outliers.

C.4 Additional Robustness Results

Table 5 reports additional robustness checks discussed in the main text. Panel A shows the placebo test (fake treatment shifted 3 years earlier, estimated on pre-treatment data only). Panel B shows the inverse hyperbolic sine transformation. Panel C summarizes the Bacon decomposition weights.

Table 5: Additional Robustness Checks

	Coefficient	SE
Panel A: Placebo Test (Treatment Shifted 3 Years Earlier)		
Fake Treatment	0.184**	(0.089)
Panel B: Alternative Outcome (Inverse Hyperbolic Sine)		
Treated (TWFE)	0.056	(0.089)
Panel C: Bacon Decomposition Weights		
	Weight	Avg. Estimate
Earlier vs Later Treated	0.43	0.256
Later vs Earlier Treated	0.57	-0.096

Notes: Panel A restricts the sample to pre-2006 observations and assigns fake treatment at $g_d - 3$ for each district. Panel B replaces $\log(\text{nightlights} + 1)$ with $\sinh^{-1}(\text{nightlights})$. Panel C reports Bacon (2021) decomposition of the TWFE estimator. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.