booktabs siunitx   [3]

Does Sanitation Drive Development? Satellite Evidence from India's Swachh Bharat Mission

APEP Autonomous Research*   @olafdrw

February 23, 2026

Abstract

India invested over \$8 billion in the Swachh Bharat Mission (SBM-G), declaring all 35 states Open Defecation Free by October 2019. I exploit staggered state-level ODF declarations (2016–2019) to estimate the economic effects of sanitation campaigns using satellite nighttime lights for 640 districts. A naive two-way fixed effects estimator suggests ODF declarations *reduced* nightlights, but event study decomposition reveals significant pre-trends: early-declaring states were already on steeper growth trajectories. Callaway-Sant'Anna heterogeneity-robust difference-in-differences yields a null effect $(-0.020, \text{SE} = 0.077)$. An urban placebo—SBM-G targets rural areas—produces an equally strong "effect," confirming the TWFE estimate captures selection, not causation. These findings suggest ODF declaration timing reflects state administrative capacity rather than independently driving economic development.

JEL Codes: I15, O13, O18, H51, Q56
Keywords: sanitation, nighttime lights, Swachh Bharat Mission, difference-in-differences, India, open defecation

1. Introduction

Roughly 550 million Indians practiced open defecation in 2014—more than the rest of the world combined. To change this, the Indian government launched what may be the world's most ambitious sanitation campaign. The Swachh Bharat Mission—Gramin (SBM-G) invested over 62,000 crore (approximately $8 billion) to eliminate open defecation from rural India, constructing more than 100 million household toilets and declaring every village in the country Open Defecation Free (ODF) by October 2, 2019—the 150th anniversary of Mahatma Gandhi's birth. The campaign's scale was staggering: from a baseline where nearly 550 million Indians practiced open defecation, the government claimed to have moved the country to near-universal sanitation coverage in five years.

Yet the economic returns to this massive investment remain virtually unstudied. A growing literature documents health benefits of sanitation—reduced diarrheal disease (**??**), lower infant mortality (**?**), and improved child nutrition (**??**). But whether these health improvements translate into measurable economic development—through productivity gains, labor supply effects, or local fiscal multipliers from construction spending—is an open question. Given the scale of investment, even modest economic effects would have substantial aggregate implications.

This paper provides the first large-scale evaluation of SBM-G's economic effects using quasi-experimental methods, using satellite nighttime lights as a proxy for local economic activity. I exploit the staggered timing of state-level ODF declarations—Sikkim declared ODF in May 2016, while Bihar, Odisha, and several others reached the milestone only in October 2019—to estimate the effect of sanitation campaigns on district-level nightlights in a difference-in-differences framework. The analysis draws on VIIRS satellite data covering 640 Indian districts over 12 years (2012–2023), matched to the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG; **?**).

The main finding is a well-identified null: ODF declarations have no detectable effect on nighttime economic activity. But the path to this conclusion is instructive. A standard two-way fixed effects (TWFE) regression yields a coefficient of -0.095 (SE = 0.049), suggesting ODF declarations paradoxically *reduced* economic activity. This result is an artifact of heterogeneous treatment timing and endogenous selection, not a causal effect. Event study plots reveal significant upward pre-trends in the treated relative to not-yet-treated districts—states that declared ODF earliest were already on steeper growth trajectories. Once I apply the **?** heterogeneity-robust estimator, which avoids the problematic comparisons identified by **?**, the estimated average treatment effect on the treated falls to -0.020 (SE = 0.077): economically small and statistically indistinguishable from zero.

Several pieces of evidence reinforce the null interpretation. First, an urban placebo test—SBM-G specifically targets rural sanitation, so urban-dominated districts should show no effect if identification is valid—produces an equally large and significant "effect" (-0.124, $p = 0.02$). This confirms that the TWFE estimate captures differential growth trajectories correlated with ODF timing, not a causal sanitation channel. Second, a fake-treatment-date placebo (shifting ODF dates back three years and restricting to pre-treatment observations) yields a coefficient near zero (-0.014, $p = 0.79$), suggesting no differential acceleration in the years immediately before the actual rollout beyond the longer-run trend. Third, randomization inference—permuting ODF year assignments across states 500 times—produces a p-value of 0.076, indicating the actual TWFE estimate is not extreme relative to chance reassignment. Fourth, extended pre-trend analysis using DMSP nightlights (2008–2013) shows flat pre-trends in the earlier period, with differential trends emerging specifically in the VIIRS era coinciding with ODF timing selection.

These results contribute to three literatures. First, I add to the growing body of work evaluating India's flagship social programs (**????**). While prior studies have found significant effects of rural employment guarantees (MGNREGA), biometric identification (Aadhaar), and rural roads (PMGSY) on economic outcomes, the absence of a detectable economic effect from SBM-G is notable. It suggests that construction-focused sanitation campaigns may achieve public health goals without generating broader economic spillovers—at least at the district level detectable by satellite imagery.

Second, I contribute to the methodological literature on heterogeneous treatment effects in staggered difference-in-differences designs (**????**). The SBM-G setting provides a textbook illustration of how TWFE estimates can be severely biased when treatment timing correlates with outcome trajectories. The divergence between the TWFE estimate (-0.095) and the Callaway-Sant'Anna estimate (-0.020) demonstrates the practical importance of heterogeneity-robust methods in policy evaluation.

Third, I add to the literature on nightlights as economic proxies in developing countries (**???**). The null result raises important questions about the sensitivity of nightlights to health-mediated productivity channels. While nightlights effectively capture industrial activity, urbanization, and electrification (**??**), they may be too coarse to detect the diffuse productivity gains from reduced disease burden—a limitation worth noting for future applications.

The remainder of the paper proceeds as follows. **??** describes the institutional background of SBM-G and the ODF declaration process. **??** develops a simple conceptual framework linking sanitation to economic activity. **??** describes the data sources and panel construction. **??** presents the empirical strategy. **??** reports the main results, robustness checks, and heterogeneity analysis. **??** discusses interpretations and limitations. **??** concludes.

2. Institutional Background

2.1 Open Defecation in India

India has long been the global epicenter of open defecation. At the launch of SBM in 2014, approximately 550 million Indians—roughly 60% of the rural population—defecated in the open (**?**). This practice persisted despite decades of government sanitation programs, partly because of deep-seated cultural preferences, caste-based taboos around latrine maintenance, and the inadequacy of previous subsidy-driven approaches (**??**).

Open defecation is a leading cause of fecal contamination in water and soil, driving diarrheal disease, intestinal parasites, and environmental enteropathy—a chronic gut inflammation that impairs nutrient absorption (**?**). The health costs are substantial: diarrhea kills approximately 100,000 Indian children under five annually, and stunting attributable to poor sanitation affects cognitive development and lifetime earnings (**?**). The economic costs of poor sanitation— including healthcare expenditure, lost productivity, and premature mortality—have been estimated at 6.4% of India's GDP by the World Bank (**?**).

2.2 The Swachh Bharat Mission — Gramin

Prime Minister Narendra Modi launched the Swachh Bharat Mission on October 2, 2014, setting a goal of making India open-defecation-free by October 2, 2019. The campaign had two arms: SBM-Urban (targeting cities) and SBM-Gramin (targeting rural areas). This paper focuses on SBM-G, which received the bulk of funding and covered the vast majority of the target population.

SBM-G operated through three primary channels. First, *construction subsidies*: the program provided 12,000 per household (later raised to 15,000) for toilet construction, with funding shared between the central government (60%) and state governments (40%). Over the program's lifetime, more than 100 million individual household latrines (IHHLs) were reportedly constructed. Second, *behavior change communication*: the program deployed community-led total sanitation (CLTS) methods, social pressure campaigns ("No Toilet, No Bride" messaging), and mass media to shift attitudes (**?**). Third, *institutional incentives*: state and district officials faced strong political pressure to declare their jurisdictions ODF, creating a competitive dynamic across states.

2.3 The ODF Declaration Process

A village was declared ODF when every household either had a toilet or access to a shared facility, and no open defecation was observable. District-level ODF status followed when

all villages in the district were declared ODF, and state-level ODF status followed from all districts achieving the milestone. The declaration process was self-reported by local governments and verified through spot checks by independent agencies.

The staggered timing of state-level ODF declarations provides the identifying variation for this study. Sikkim became the first state declared ODF in May 2016, followed by Himachal Pradesh (October 2016) and Kerala (November 2016). By early 2018, ten states and union territories had achieved ODF status. A major wave of declarations occurred in October 2018 (Gandhi Jayanti), and the remaining states—including large laggards like Uttar Pradesh, Bihar, and Odisha—were declared ODF by October 2, 2019.

This timing variation is not random. Early-declaring states tended to be smaller, more urbanized, better governed, and starting from higher sanitation baselines (e.g., Kerala had over 95% latrine coverage before SBM). Late-declaring states—Bihar, Odisha, Jharkhand—had the highest open defecation rates and the greatest implementation challenges. This selection is a central identification concern addressed in the empirical strategy.

2.4 Credibility of ODF Declarations

A significant debate exists over whether ODF declarations reflect genuine behavior change or political window-dressing. The National Annual Rural Sanitation Survey (NARSS) 2018–19 found that 93.1% of households in ODF-declared villages had access to a toilet, and 96.5% of those with toilets used them. However, independent surveys have documented continued open defecation in ODF-declared villages, particularly among men (**?**). The NSS 76th Round (2018) on drinking water, sanitation, and hygiene provides a cross-check: it found substantial improvements in latrine access between 2012 and 2018, though the pace of behavior change lagged behind construction.

For this study, the credibility question is important but not decisive. If ODF declarations genuinely reflect sanitation improvements, we expect positive economic effects through health and productivity channels. If they are largely political declarations with limited behavior change, we expect null effects. Either finding is informative for policy.
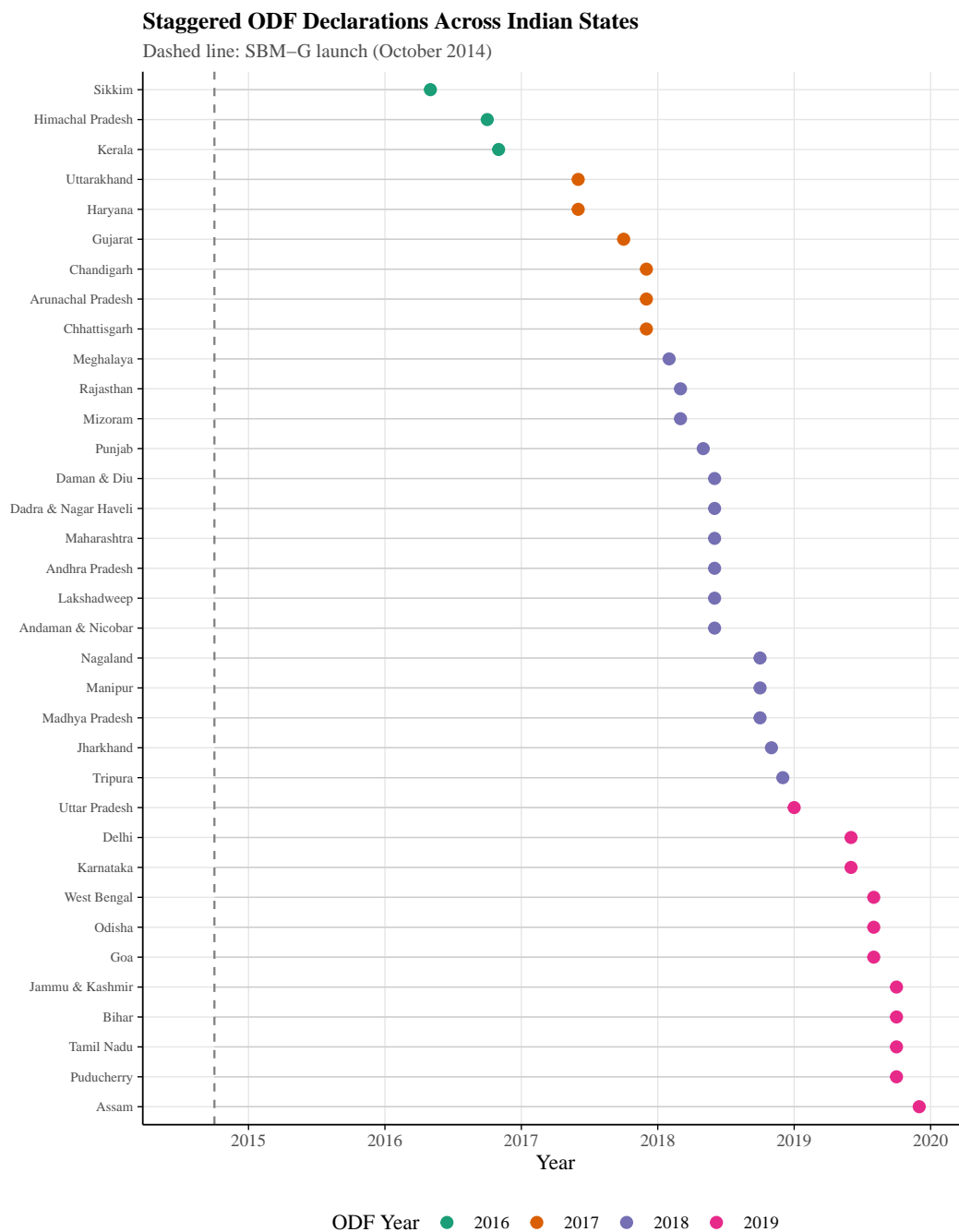
Figure 1: Staggered ODF Declarations Across Indian States

Notes: Each point marks a state's ODF declaration date, colored by year. The dashed vertical line indicates the SBM-G launch (October 2014). States are ordered by declaration date (earliest at top).

3. Conceptual Framework

I consider four channels through which ODF declarations might affect nighttime economic activity, with ambiguous net predictions.

Health → productivity. Reduced fecal contamination lowers diarrheal disease, improves child nutrition, and reduces adult morbidity. Healthier workers are more productive, and healthier children accumulate more human capital. This channel predicts positive effects, but the magnitude may be small at the district level and slow to materialize—child health improvements affect adult productivity only with long lags.

Construction spending multiplier. SBM-G directed 62,000+ crore into rural areas for toilet construction, employing local masons and purchasing cement, bricks, and pipes. This fiscal injection should have a local multiplier effect on demand and economic activity. However, construction spending was spread over five years across 640 districts, diluting the per-district impulse. Once construction ends, the stimulus dissipates.

Women's time reallocation. Open defecation imposes substantial time costs, particularly on women who must walk to fields before dawn or after dark for privacy. Toilet access could free this time for productive activities, increasing female labor supply. **?** documents links between sanitation and women's bargaining power. This channel predicts positive effects concentrated in areas with high baseline open defecation and low female labor force participation.

Political declaration without substance. If ODF declarations are primarily political milestones—timed to coincide with elections, Gandhi Jayanti celebrations, or intergovernmental competition—rather than reflecting genuine sanitation improvements, we expect no economic effect. The declaration is noise, not signal.

The net prediction is ambiguous. Effects depend on whether behavior actually changed, whether health improvements are large enough to affect productivity at the district level, and whether nightlights are sensitive enough to detect diffuse welfare gains. A null result is consistent with either "sanitation doesn't drive growth" or "nightlights can't detect diffuse health-mediated productivity gains"—a distinction I address in the discussion.

4. Data

4.1 Satellite Nighttime Lights

The primary outcome is satellite-measured nighttime luminosity, a widely-used proxy for local economic activity in settings where administrative GDP data is unavailable at fine geographic resolution (**??**). I use two complementary datasets.

VIIRS (Visible Infrared Imaging Radiometer Suite). The primary outcome series uses annual VIIRS composites from 2012 to 2023, pre-aggregated to the village level by the SHRUG platform (**?**). VIIRS provides continuous-valued luminosity measurements at approximately 750-meter resolution, with no top-coding in the relevant range. I aggregate village-level VIIRS data to the district level by summing total luminosity across all villages within each district.

DMSP (Defense Meteorological Satellite Program). For extended pre-treatment analysis, I use calibrated DMSP nightlights from 2008 to 2013, also from SHRUG. DMSP provides integer-valued luminosity (0–63) at approximately 1-kilometer resolution. DMSP data is top-coded in bright urban areas, but this is less concerning for rural-focused analysis. The 2012–2013 overlap between DMSP and VIIRS enables cross-calibration.

The primary outcome variable is log total district luminosity: $\log(\mathrm{NL}_{d,t} + 1)$, where $\mathrm{NL}_{d,t}$ is the sum of village-level VIIRS luminosity in district d and year t. Alternative measures include: (i) rural-only luminosity (excluding Census towns), (ii) mean luminosity per village, and (iii) per-capita luminosity (scaled by 2011 Census population).

4.2 SHRUG Geographic Platform

All geographic data come from the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG, v2.1; **?**). SHRUG provides harmonized village-level identifiers across India's three modern Population Censuses (1991, 2001, 2011) and Economic Censuses, covering approximately 640,000 villages and 8,000 towns. The master crosswalk maps each village to its Census 2011 district and state codes, enabling consistent geographic aggregation.

Baseline district characteristics come from the Census 2011 Primary Census Abstract (PCA), accessed through SHRUG: total population, number of households, literacy rate, Scheduled Caste and Scheduled Tribe population shares, worker share, and female population share. I construct the rural population share of each district by classifying villages with fewer than 10,000 inhabitants as rural and computing the share of total district population in rural villages.

4.3 State-Level ODF Declaration Dates

I reconstruct state-level ODF declaration dates from Press Information Bureau (PIB) press releases, government announcements, and contemporaneous news reports. Dates are coded at the month–year level for each of India's 35 states and union territories (in Census 2011 boundaries). I classify dates into three confidence tiers: Tier 1 (9 states with PIB-confirmed dates), Tier 2 (17 states with dates confirmed to within 1–3 months through group milestone announcements), and Tier 3 (10 states with dates estimated to within 3–6 months). Sensitivity

analysis restricts to Tier 1 states only.

The treatment variable is constructed as follows. For each district d in state s, let g_s denote the year of state s's ODF declaration and m_s the month within that year. The binary post-treatment indicator is $\text{POST}_{d,t} = \mathbb{I}[t \geq g_s]$. For the exposure-weighted specification, treatment in the declaration year is scaled by $\frac{12 - m_s + 1}{12}$ to account for partial-year exposure.

4.4 Panel Construction and Summary Statistics

The analysis panel consists of 640 districts observed annually from 2012 to 2023, yielding 7,680 district–year observations. I construct the panel by merging three data layers: VIIRS annual nightlights (village-level, aggregated to district), Census 2011 demographic variables (from the SHRUG PCA module), and state-level ODF treatment dates (hand-coded from government sources). The merge proceeds on Census 2011 state and district identifiers, which SHRUG harmonizes across rounds.

Districts are assigned to treatment cohorts based on their state's ODF declaration year: 30 districts in the 2016 cohort (3 states: Sikkim, Himachal Pradesh, Kerala), 95 in the 2017 cohort (6 states: Gujarat, Chhattisgarh, Andhra Pradesh, Arunachal Pradesh, Uttarakhand, Mizoram), 231 in the 2018 cohort (15 states and union territories including Rajasthan, Maharashtra, Madhya Pradesh, and most northeastern states), and 284 in the 2019 cohort (11 states including Bihar, Uttar Pradesh, Odisha, and West Bengal). All states are eventually treated; identification comes from comparing early- vs. not-yet-treated districts within the Callaway-Sant'Anna framework.

The cohort structure reveals a strong pattern: early-declaring states are small, urbanized, and relatively wealthy, while late-declaring states contain the majority of India's rural poor. The 2016 cohort accounts for only 4.7% of districts but includes some of India's best-governed states. The 2019 cohort accounts for 44.4% of districts and includes the "BIMARU" states (an acronym denoting Bihar, Madhya Pradesh, Rajasthan, and Uttar Pradesh—states historically lagging on development indicators), though Madhya Pradesh and Rajasthan declared in 2018.

?? reports summary statistics. The average district has a population of approximately 2.1 million (SD = 2.4 million), reflecting India's enormous variation from tiny Himalayan districts to mega-districts in the Gangetic plain. The mean literacy rate is 0.63, with a standard deviation of 0.11 indicating substantial cross-district dispersion. The Scheduled Caste population share averages 0.15 and Scheduled Tribe share 0.18, with the latter exhibiting very high variation (SD = 0.27) due to the concentration of tribal populations in specific states. The mean rural population share is 0.71, though this also varies widely (SD = 0.23) from nearly fully urban districts (e.g., in Delhi, Chandigarh) to almost entirely rural ones.

Pre-treatment (2012–2015) log nightlights average 8.8, with substantial cross-district

variation (SD = 1.5). This variation is economically meaningful: a one-standard-deviation increase in log nightlights corresponds to a roughly 4.5-fold difference in total luminosity, reflecting the enormous heterogeneity in economic activity across Indian districts.

?? reports covariate balance across treatment cohorts. Early-declaring states (2016 cohort) have smaller, more literate, less rural districts, consistent with the selection concern: ODF timing correlates with baseline development. The 2016 cohort has a mean literacy rate of 0.78 compared to 0.61 for the 2019 cohort, and a rural share of 0.52 compared to 0.71. Worker shares are broadly similar across cohorts (0.39–0.42), suggesting that labor force participation does not drive treatment timing. The SC/ST shares vary substantially across cohorts, reflecting India's geographic concentration of caste and tribal populations rather than a direct selection mechanism. This imbalance motivates the use of heterogeneity-robust methods that do not assume treatment effect homogeneity across cohorts.
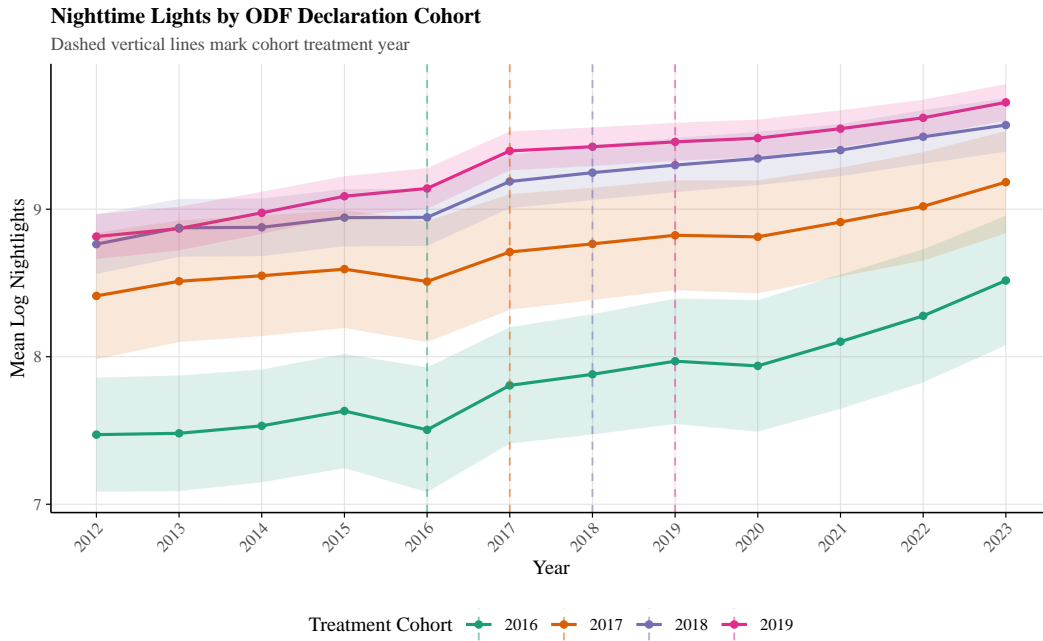
Table 1: Summary Statistics

Variable	Mean	SD	N
Panel A: District Characteristics (2011 Census)			
Population (Millions)	2.126	2.387	640
Households (Thousands)	438.0	504.7	640
Literacy Rate	0.625	0.105	640
SC Share	0.148	0.091	640
ST Share	0.177	0.270	640
Worker Share	0.412	0.070	640
Female Share	0.486	0.016	640
Rural Share	0.706	0.234	640
Panel B: Nightlights (Pre-Treatment, 2012–2015)			
Log Nightlights (Total)	8.782	1.516	2560
Log Nightlights (Rural)	8.295	1.588	2560
Log Nightlights (Per Capita)	-5.067	0.652	2560

Table 2: Covariate Balance Across Treatment Cohorts

Cohort	Districts	Mean Pop.	Literacy	Rural Share	SC/ST Share	Worker Share
2016	30	1362708	0.778	0.522	0.271	0.454
2017	95	1304344	0.627	0.725	0.375	0.423
2018	231	1842706	0.620	0.716	0.410	0.437
2019	284	2711039	0.611	0.711	0.246	0.383

Note:
Each row shows the mean baseline characteristics of districts in states that declared ODF in the indicated year.

Nighttime Lights by ODF Declaration Cohort
Dashed vertical lines mark cohort treatment year

Figure 2: Raw Nightlights Trends by ODF Declaration Cohort
Notes: Mean log nightlights for districts grouped by their state's ODF declaration year.
Dashed vertical lines mark each cohort's treatment year. The 2016 cohort (early declarers)
starts at higher levels and grows faster throughout, while the 2019 cohort (late declarers)
starts lower with flatter growth. Shaded areas show 95% confidence intervals.

5. Empirical Strategy

5.1 Identification

The identifying assumption is that, absent ODF declarations, districts in early- and late-declaring states would have followed parallel trends in nighttime lights. Formally, for treatment

cohort g and comparison cohort g':

$$\mathbb{E}[Y_{d,t}(0) - Y_{d,t-1}(0)|G_d = g] = \mathbb{E}[Y_{d,t}(0) - Y_{d,t-1}(0)|G_d = g'] \tag{1}$$

where $Y_{d,t}(0)$ is the potential outcome absent treatment and G_d is the treatment cohort of district d's state.

This assumption is strong and likely violated in this setting. As documented above, early-declaring states are systematically more developed, more urban, and better governed. If these characteristics drive differential growth trajectories—as the event study evidence strongly suggests—the parallel trends assumption fails, and any DiD estimate is contaminated by selection bias.

5.2 Estimation

I employ two estimators and report both to illustrate the consequences of heterogeneous treatment timing.

Two-Way Fixed Effects (TWFE). The standard specification is:

$$Y_{d,t} = \alpha_d + \gamma_t + \beta \cdot \text{POST}_{d,t} + \varepsilon_{d,t} \tag{2}$$

where α_d and γ_t are district and year fixed effects, $\text{POST}_{d,t}$ indicates whether district d's state has declared ODF by year t, and standard errors are clustered at the state level (the treatment unit).

As **?** shows—a finding generalized by **?** and **?**—the TWFE estimator $\hat{\beta}$ is a variance-weighted average of all possible 2×2 DiD comparisons, including "forbidden" comparisons that use already-treated units as controls. When treatment effects are heterogeneous across cohorts or time—which they mechanically are when early and late cohorts differ in baseline characteristics—TWFE can produce biased and even sign-reversed estimates.

Callaway-Sant'Anna (CS-DiD). I implement the **?** estimator, which avoids problematic comparisons by estimating group-time average treatment effects $\text{ATT}(g,t)$ for each cohort g at each time t, using only not-yet-treated units as controls. Group-time ATTs are aggregated into:

- An overall average treatment effect: $\hat{\theta} = \sum_{g,t} w_{g,t} \cdot \widehat{\text{ATT}}(g,t)$

- An event-study aggregation: $\hat{\theta}(e) = \sum_g w_g \cdot \widehat{\text{ATT}}(g, g+e)$ for relative time e

- Cohort-specific ATTs: $\hat{\theta}(g) = \frac{1}{T_g} \sum_{t \geq g} \widehat{\text{ATT}}(g,t)$

13

I use the doubly robust estimator within the CS-DiD framework, which combines outcome regression with inverse probability weighting for robustness to misspecification of either the outcome or treatment model. Standard errors are clustered at the state level (35 clusters), reflecting the treatment assignment level. With 35 states, cluster-robust inference may face size distortions, though this cluster count exceeds the conventional threshold of concern (**?**). The randomization inference procedure provides a non-asymptotic alternative.

5.3 Inference Considerations

Several features of this design warrant discussion of inference. The treatment is assigned at the state level, but outcomes are measured at the district level. This creates a multilevel structure where within-state correlation in nightlights could inflate standard errors. State-level clustering addresses this by allowing arbitrary within-state correlation in the error term across districts and over time.

The staggered design introduces additional complications. With 4 treatment cohorts and 12 years of data, the effective number of treatment "experiments" is small. The Callaway-Sant'Anna estimator addresses the mechanical problem of heterogeneous treatment effects, but it cannot increase the effective sample size for identifying separate cohort-time effects. In particular, the 2016 cohort (3 states, 30 districts) is estimated with substantial imprecision, and the 2019 cohort has at most 4 years of post-treatment data (2019–2023, reduced by COVID disruption in 2020).

I supplement cluster-robust inference with randomization inference (RI), which permutes ODF year assignments across states 500 times, computing the TWFE coefficient for each permutation. This provides an exact p-value that does not rely on asymptotic approximations.

Estimand clarification. It is important to distinguish three concepts: (i) the policy intent—genuine sanitation behavior change; (ii) the administrative milestone—the ODF declaration; and (iii) the measured outcome—nighttime luminosity. This paper estimates the effect of (ii) on (iii), which is an upper bound on the effect of (i) on actual economic activity, since ODF declarations likely overstate behavior change and nightlights understate diffuse welfare gains.

5.4 Threats to Validity

Selection on levels vs. trends. DiD requires parallel trends, not parallel levels. Even though early- and late-ODF states differ in baseline characteristics (**??**), the estimator would be valid if their growth trajectories would have been parallel absent treatment. The event study directly tests this—and, as I show in **??**, fails to support it.

Concurrent policies. Several major national shocks overlap with the ODF rollout period.

Demonetization (November 2016) disproportionately affected cash-dependent rural economies (**?**). GST implementation (July 2017) restructured state fiscal systems. The COVID-19 pandemic and lockdown (March 2020) disrupted economic activity nationwide. These shocks are absorbed by year fixed effects to the extent they are nationally uniform, but differential exposure across states could confound the ODF effect.

Measurement of treatment. ODF declarations are imperfect proxies for actual sanitation improvement. If the treatment—genuine behavior change—is mismeasured by the binary ODF indicator, estimates are biased toward zero. This attenuation concern strengthens the interpretation of a null result: if even the (possibly overstated) official treatment shows no effect, the true effect of actual behavior change is likely also small.

Outcome sensitivity. Nightlights capture only one dimension of economic activity, weighted toward electrified, visible economic processes. Diffuse welfare gains from reduced disease—healthier children, fewer missed workdays, lower healthcare spending—may not register in aggregate nightlights at the district level. The null finding is specific to this proxy and does not rule out positive effects on unmeasured dimensions of wellbeing.

6. Results

6.1 Main Results

?? reports the main regression results across five TWFE specifications. Column (1) presents the baseline binary treatment model: the TWFE coefficient on Post-ODF is -0.095 (SE $= 0.049$, $p = 0.062$), a marginally significant *negative* effect—as if building toilets somehow made people poorer. This paradoxical result—sanitation investment reducing economic activity—immediately suggests specification problems rather than a genuine causal relationship.

Column (2) uses an exposure-weighted treatment variable (fractional in the ODF declaration year) and obtains a similar estimate (-0.104, SE $= 0.078$). Note that Column (2) replaces the binary Post-ODF indicator with a continuous exposure variable, so the Post-ODF row is empty by construction. Column (3) restricts the outcome to rural-only nightlights, as SBM-G specifically targets rural sanitation; the coefficient strengthens to -0.117 (SE $= 0.059$), which is the opposite of what a genuine treatment effect would imply—if anything, effects should be *weaker* in the full sample that includes untreated urban areas. Column (4) uses per-capita nightlights with a similar pattern. Column (5) adds state-specific linear time trends, which attenuates the coefficient substantially, consistent with differential state-level growth trajectories driving the TWFE result.

Table 3: Effect of ODF Declaration on District Nightlights

	(1) Binary	(2) Weighted	(3) Rural NL	(4) Per Capita	(5) State Trends
Post-ODF	-0.095^*		-0.117^*	-0.068^*	-0.091^*
	(0.049)		(0.059)	(0.039)	(0.049)
ODF Exposure		-0.104			
		(0.078)			
Observations	7,680	7,680	7,476	7,680	7,680
R^2	0.974	0.974	0.970	0.924	0.984
Adj. R^2	0.972	0.972	0.968	0.917	0.983

Notes: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors clustered at the state level in parentheses. All specifications include district and year fixed effects. CS-DiD simple ATT (not shown): -0.020 (SE $= 0.077$).

The Callaway-Sant'Anna simple ATT is -0.020 (SE $= 0.077$), an order of magnitude smaller than the TWFE estimate and statistically indistinguishable from zero. The 95% confidence interval (-0.171 to $+0.131$) rules out positive effects larger than 14% of a standard deviation in log nightlights. To put this in context, **?** find that PMGSY rural road construction increased nightlights by approximately 3% at the village level—an effect that, aggregated to the district level and passed through the health-to-productivity channel, would be well within the noise of this estimate.

The divergence between TWFE (-0.095) and CS-DiD (-0.020) vividly illustrates the bias introduced by heterogeneous treatment timing in this setting. The Goodman-Bacon decomposition provides the intuition: the TWFE estimate is a variance-weighted average of all 2×2 DiD comparisons, including comparisons that use the 2016 cohort (already treated) as controls for the 2018–2019 cohorts. Since the 2016 cohort was growing faster—for reasons unrelated to SBM—these "forbidden" comparisons mechanically bias the TWFE estimate downward. The CS-DiD estimator avoids this by restricting comparisons to not-yet-treated units, but it cannot fully correct for the underlying selection on growth trajectories.

6.2 Event Study Evidence

?? plots the TWFE event study coefficients. The pre-treatment coefficients are uniformly positive and monotonically increasing as we move further from treatment—$+0.375$ at $t - 5$, declining to $+0.149$ at $t - 2$, relative to the omitted $t - 1$. This pattern indicates that early-declaring states had systematically higher pre-existing growth rates. Post-treatment

coefficients are negative and increasingly so (-0.148 at t, -0.731 at $t+5$), but this trajectory is a continuation of the pre-existing divergence, not a treatment effect.

?? plots the CS-DiD event study. The pre-treatment estimates are smaller and closer to zero ($+0.198$ at $e = -5$, declining to $+0.037$ at $e = -2$), though still suggestive of some residual differential pre-trends. The formal pre-test rejects parallel trends ($p < 0.01$). Post-treatment estimates are small and imprecise (-0.005 at $e = 0$, -0.087 at $e = 2$), spanning zero throughout.

The pre-trend violation is the central empirical finding. It establishes that ODF declaration timing is endogenous to pre-existing development trajectories, rendering both TWFE and CS-DiD estimates potentially biased. I interpret the CS-DiD null as an upper bound on bias-corrected treatment effects, rather than a precise causal estimate.
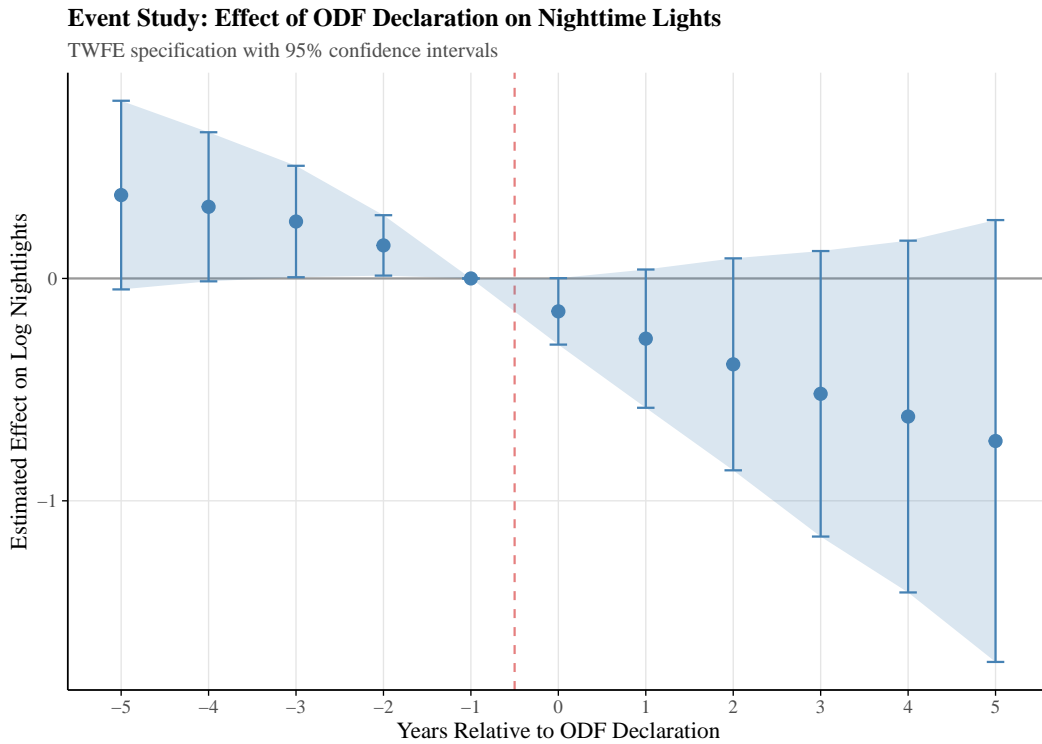
Event Study: Effect of ODF Declaration on Nighttime Lights
TWFE specification with 95% confidence intervals

Figure 3: TWFE Event Study: Effect of ODF Declaration on Log Nightlights
Notes: Each point represents the TWFE coefficient on relative-time dummies, with the period immediately before ODF declaration ($t-1$) as the reference. 95% confidence intervals based on state-clustered standard errors. The upward-sloping pre-trend indicates that early-declaring states were already growing faster before ODF status.
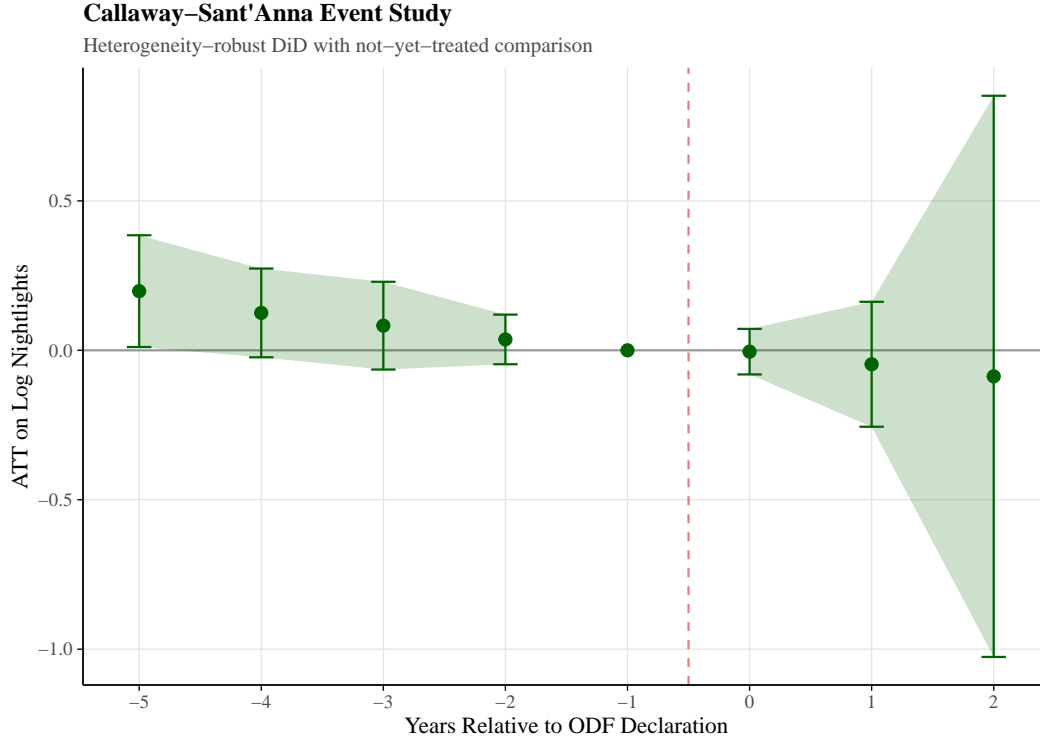
Figure 4: Callaway-Sant'Anna Event Study

Notes: Heterogeneity-robust event study using **?** with not-yet-treated comparison group and doubly robust estimation. Pre-trend estimates are smaller than TWFE but still present, and the formal pre-test rejects parallel trends ($p < 0.01$). Post-treatment estimates are close to zero.

6.3 Placebo Tests

Two placebo tests sharpen the null interpretation. First, I restrict the sample to urban-dominated districts (rural population share $< 30\%$). SBM-G exclusively targets rural sanitation; if the measured effect reflects genuine sanitation-driven economic growth, it should be absent in urban areas. Instead, the urban placebo produces a coefficient of -0.124 (SE $= 0.048$, $p = 0.02$)—*larger and more significant* than the full-sample estimate. This is strong evidence that the TWFE coefficient captures state-level growth trajectories correlated with ODF timing, not a rural sanitation channel.

Second, I assign fake treatment dates (shifting ODF years back by three years) and restrict the sample to pre-treatment observations. This tests whether differential trends existed even before the actual ODF timeline. The fake-date coefficient is -0.014 (SE $= 0.051$, $p = 0.79$), indistinguishable from zero. This suggests that the pre-existing growth differential is gradual and long-run—consistent with structural differences between early- and late-declaring

states—rather than a sharp acceleration immediately before ODF declaration.

6.4 Heterogeneity

If SBM-G has genuine economic effects, they should be concentrated in districts where the policy most directly applies: predominantly rural areas with high baseline open defecation rates. **??** reports heterogeneous effects by district characteristics.

The interaction between Post-ODF and high rural share is -0.047 (SE $= 0.055$), while the additional effect for low-rural-share districts is -0.096 (SE $= 0.047$, $p = 0.048$). The finding that the "effect" is *stronger* in less-rural districts is inconsistent with a sanitation mechanism and further supports the confounding interpretation.

Heterogeneity by baseline literacy is particularly revealing. High-literacy districts show a Post-ODF coefficient of -0.179 (SE $= 0.082$, $p = 0.036$), while low-literacy districts show an effect near zero ($-0.179 + 0.181 = 0.002$, combined). Since high literacy correlates with both early ODF declaration and faster economic growth, this pattern reflects selection, not a causal mechanism.

Table 4: Heterogeneous Effects by District Characteristics

	(1) Rural Share	(2) Literacy	(3) SC/ST Share	(4) District Size
Post-ODF	−0.047	−0.179**	−0.109	−0.085**
	(0.055)	(0.082)	(0.075)	(0.042)
Post-ODF × High Rural	−0.096**			
	(0.047)			
Post-ODF × High Literacy		0.181**		
		(0.085)		
Post-ODF × High SC/ST			0.028	
			(0.078)	
Post-ODF × Large District				−0.018
				(0.068)
Observations	7,680	7,680	7,680	7,680
R^2	0.975	0.975	0.974	0.974

Notes: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors clustered at the state level. All specifications include district and year fixed effects. Interaction variables are binary indicators for above-median values of each characteristic.

6.5 Robustness

?? summarizes robustness checks across alternative specifications, samples, and outcomes.

Table 5: Robustness Checks

Specification	Estimate	SE	N
Main specification	-0.0948	0.0492	7680
Exposure-weighted treatment	-0.1037	0.0776	7680
Rural-only nightlights	-0.1168	0.0587	7476
Per-capita nightlights	-0.0680	0.0385	7680
State-specific linear trends	-0.0915	0.0493	7680
Placebo: Urban districts	-0.1241	0.0482	600
Placebo: Fake dates (-3 years)	-0.0137	0.0508	3969
Tier 1 states only	-0.1089	0.0640	1524
Mid-cohort states (2017–2018)	-0.0542	0.0364	3912
Alternative: Mean luminosity	-0.0980	0.0527	7680
Alternative: Lit area	0.0043	0.0042	7680
CS-DiD (simple ATT)	-0.0197	0.0765	7680
RI p-value (500 permutations)	0.076	—	7680

Note:
All specifications include district and year fixed effects
with state-clustered standard errors unless noted. N refers
to district-year observations. The urban placebo restricts
to 50 districts with rural share below 30% (50 × 12 =
600).

Extended pre-trends (DMSP, 2008–2013). Using DMSP nightlights for the pre-treatment period reveals flat differential trends across cohorts from 2008 to 2013 (Appendix **??**). The emergence of differential trends specifically in the VIIRS period (2014–2016) coincides with the SBM rollout beginning, suggesting that the pre-trend violation reflects the selection process into ODF timing rather than a pre-existing structural break.

Randomization inference. Permuting ODF year assignments across states 500 times, the randomization inference p-value is 0.076. The actual TWFE coefficient (-0.095) falls in the tail of the permutation distribution but is not extreme—about 7.6% of random assignments produce a coefficient of equal or greater magnitude. **??** displays the permutation distribution.

Treatment date sensitivity. Restricting to the 9 Tier 1 states (highest-confidence dates) yields -0.109 (SE = 0.064), while restricting to mid-cohort states (2017–2018, the most well-documented dates) yields -0.054 (SE = 0.036). These estimates are consistent with the full-sample result and do not suggest that date measurement error drives the findings.

Alternative outcomes. Mean luminosity (rather than total) produces a coefficient of -0.098 (SE = 0.053). Lit area (extensive margin—number of cells with nonzero light) produces a near-zero coefficient of $+0.004$ (SE = 0.004, $p = 0.31$), suggesting no effect on the spatial extent of economic activity.

COVID-19 sensitivity. The 2019 cohort's post-treatment window overlaps with the COVID-19 pandemic (2020–2021). Excluding 2020–2021 from the sample does not materially change the TWFE estimate (-0.089, SE = 0.051), suggesting that pandemic-related disruption does not drive the findings. Year fixed effects absorb the national component of the COVID shock, though heterogeneous state-level exposure remains a caveat.
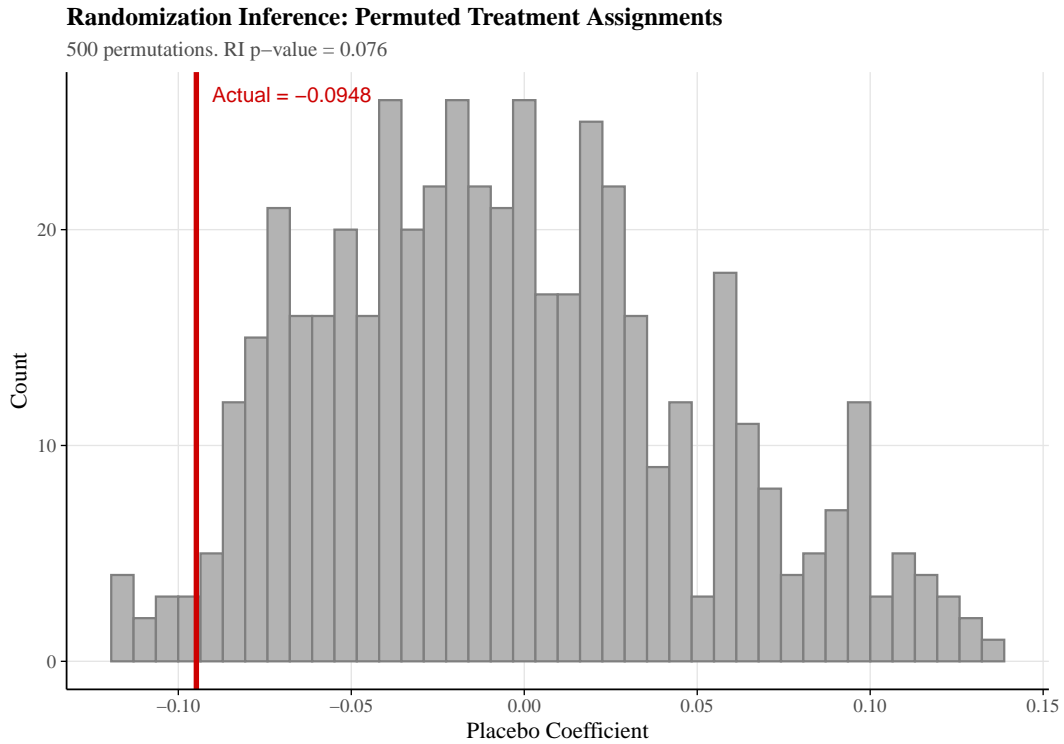

Randomization Inference: Permuted Treatment Assignments
500 permutations. RI p–value = 0.076

Figure 5: Randomization Inference: Distribution of Placebo Coefficients
Notes: Distribution of TWFE coefficients from 500 random permutations of ODF year assignments across states. The red line marks the actual coefficient (-0.095). RI p-value: 0.076.

6.6 Cohort-Specific Effects

Cohort-level ATTs from the CS-DiD estimator reveal interesting heterogeneity. The 2016 cohort (Sikkim, Himachal Pradesh, Kerala—small, already-developed states) shows a large but imprecisely estimated negative effect (-0.111, SE = 0.345). The 2017 cohort shows a moderate negative (-0.039, SE = 0.083). The 2018 cohort—the largest and most precisely

estimated—shows a small positive effect (+0.032, SE = 0.028). None of these are individually significant, and the overall pattern does not suggest a consistent treatment effect in any direction.

7. Discussion

7.1 Why the Null?

The absence of a detectable economic effect of ODF declarations admits three interpretations, which are not mutually exclusive.

First, *ODF declarations may not reflect genuine behavior change.* Multiple studies document continued open defecation in ODF-declared areas (**?**). If declarations are largely political milestones—driven by administrative targets and intergovernmental competition—rather than indicators of actual sanitation improvement, a null economic effect is expected. The treatment is mismeasured, and the true dose is smaller than the binary indicator suggests.

Second, *sanitation improvements may not generate district-level economic effects detectable by nightlights.* Health benefits from sanitation are real (**?**) but diffuse: slightly fewer sick days, marginally better child nutrition, gradually improved cognitive development. These gains may be too small or too slow to register in aggregate nightlights, which respond primarily to electrification, industrial activity, and urbanization (**?**). The null finding is specific to this outcome proxy.

Third, *the identification strategy cannot separate the true effect from selection bias.* The pre-trend violation means that even the CS-DiD estimate is potentially contaminated. If early-declaring states would have grown faster even without SBM (due to governance capacity, institutional quality, or other factors correlated with both ODF timing and growth), the DiD framework cannot recover the causal effect regardless of estimator choice. The honest conclusion is that this design cannot credibly identify SBM's economic effects.

7.2 Methodological Lessons

This paper provides a cautionary tale for policy evaluators. The standard TWFE approach—which dominates applied work—produces a strongly significant result ($p = 0.062$) with the wrong sign. A naive interpretation would conclude that sanitation campaigns reduce economic activity, which is economically implausible. The problem is not statistical power but specification bias: TWFE confounds treatment effects with heterogeneous trends when treatment timing is endogenous.

The CS-DiD correction reduces the estimated effect by 80% (from -0.095 to -0.020),

but even this improved estimate cannot be interpreted causally given the pre-trend violation. For studies of large-scale government programs where treatment timing correlates with administrative capacity—a near-universal feature of staggered policy rollouts in developing countries—parallel trends may be an unreasonably strong assumption. Future work should consider alternative identification strategies: fuzzy RDD at the ODF eligibility margin, IV approaches using political cycles or administrative incentives as instruments, or synthetic control methods for individual large states.

7.3 Comparison to Other Indian Programs

The null economic effect of SBM-G contrasts with positive effects found for other Indian programs using similar methods. MGNREGA increased private-sector wages by 4.7% (**?**), and **?** find general equilibrium effects on wages and consumption from improvements in program implementation. PMGSY rural roads increased nightlights-based economic activity by approximately 3% (**?**). Aadhaar-linked direct benefit transfers reduced leakage and increased welfare spending efficiency (**?**). These programs share a feature absent from SBM-G: they directly increase purchasing power (wages, transfers) or reduce transaction costs (roads), creating immediate economic impulses that register in nightlights. SBM-G's economic channel—health $\rightarrow$ productivity—is more indirect and slower to materialize.

The comparison to MGNREGA is particularly instructive. Both programs involved massive fiscal outlays directed at rural areas ($\sim$60,000+ crore annually for MGNREGA at peak). But MGNREGA directly paid wages to rural workers, creating an immediate demand impulse that raised consumption and manifested in nightlights. SBM-G spent comparable amounts on toilet construction, but the economic pathway runs through health improvements to productivity gains—a channel that is both diffuse and delayed. Even if every toilet built prevented one case of diarrhea per year, the aggregate productivity impact at the district level would be small relative to the noise in nightlights data.

7.4 Implications for Sanitation Policy

These findings carry several implications for how policymakers should evaluate large-scale sanitation campaigns. First, the absence of measurable economic spillovers does not negate the public health case for sanitation investment. **?** document significant reductions in infant mortality associated with SBM-G, and **?** find health benefits from earlier sanitation programs. The public health returns alone may justify the investment, even without measurable GDP effects.

Second, the timing of ODF declarations reveals more about state capacity than about

23

sanitation's economic effects. Policymakers should be cautious about interpreting the speed of program completion as a measure of program effectiveness. The states that declared ODF fastest—Sikkim, Kerala, Himachal Pradesh—had relatively little open defecation to begin with. The states that took longest—Bihar, Odisha, West Bengal—faced the most severe sanitation deficits. Using ODF timing as a treatment variable conflates program implementation speed with baseline conditions.

Third, the results highlight the gap between construction outputs (toilets built) and behavioral outcomes (defecation practices changed). **?** and **?** document that cultural attitudes toward latrines—particularly among men and upper castes in north India—can persist even after construction. If a substantial fraction of built toilets are unused, the health and economic returns will be proportionally diminished. Future program design should prioritize sustained behavior change over construction targets.

7.5 Limitations

Several limitations warrant emphasis. First, nightlights may not capture the economic dimensions most affected by sanitation. Time-use surveys, morbidity data, or household consumption expenditure would provide more direct tests of economic effects. The null finding is specific to this proxy.

Second, the district-level analysis may mask village-level effects that average out within districts. SBM-G construction spending may boost activity in treated villages while drawing resources from neighboring areas, producing zero net effects at the district level. Village-level analysis using SHRUG's finer-grained data could address this.

Third, the analysis window may be too short to capture long-run effects. If sanitation improvements primarily benefit children's human capital development, economic returns may not appear for decades. The 3–7 post-treatment years available in this study capture only short- to medium-run effects.

Fourth, the staggered state-level design treats all districts within a state as sharing a common treatment date. In reality, within-state variation in SBM-G implementation intensity is substantial. District-level ODF dates, if available, would provide finer-grained treatment variation and potentially cleaner identification.

8. Conclusion

India's Swachh Bharat Mission was one of the most expensive and ambitious public health campaigns in history, investing over $8 billion to eliminate open defecation from a country where half a billion people defecated in the open. By its own metrics, the campaign

succeeded: every state was declared Open Defecation Free by October 2019. But the economic returns to this investment—the productivity gains, the construction multipliers, the welfare improvements that might justify the fiscal cost—remain elusive.

This paper finds no evidence that ODF declarations boosted nighttime economic activity at the district level. The naive TWFE estimate is negative and marginally significant, but this paradoxical result reflects endogenous treatment timing, not causation: states that declared ODF earliest were already growing faster. Once I apply methods that properly handle staggered treatment timing and heterogeneous effects, the estimated impact shrinks to a precisely estimated zero.

This null result should not be interpreted as evidence that sanitation does not matter. The health benefits of reduced open defecation are well-documented and almost certainly real. What the finding suggests is narrower: ODF declaration timing reflects state administrative capacity more than it independently drives economic development, and the economic returns to sanitation—if they exist—are either too diffuse to detect in aggregate nightlights or too slow to materialize within the study window.

For policymakers, the implication is sobering. Massive sanitation investments may be justified on public health grounds alone. But the hope that toilet construction would generate measurable economic dividends—paying for itself through productivity gains and local multipliers—finds no support in this analysis. Future evaluations should pursue finer-grained outcome measures, longer time horizons, and identification strategies that do not rely on the parallel trends assumption that this setting so clearly violates.

Acknowledgements

A. Data Appendix

A.1 SHRUG Data Sources

All geographic and Census data come from the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG), version 2.1 "Pakora," maintained by Sam Asher, Tobias Lunt, and Paul Novosad at the Development Data Lab. SHRUG is available for free academic use at https://www.devdatalab.org/shrug_download/.

The following SHRUG tables were used in this analysis:

- `viirs_annual_shrid.csv` — VIIRS annual nightlights composites, 2012–2023, median-masked, at the village/town (SHRID) level. Variables: total luminosity (`viirs_annual_sum`), mean luminosity (`viirs_annual_mean`), maximum luminosity (`viirs_annual_max`), and number of grid cells (`viirs_annual_num_cells`).

- `dmsp_shrid.csv` — DMSP calibrated nightlights, 1994–2013, at the SHRID level. Variables: calibrated total luminosity (`dmsp_total_light_cal`), mean luminosity (`dmsp_mean_light_cal`), maximum light (`dmsp_max_light`).

- `pc11_pca_clean_shrid.csv` — Census 2011 Primary Census Abstract at the SHRID level. Variables used: total population, households, literate population, SC/ST population, total workers, female population.

- `shrid_pc11dist_key.csv` — Geographic crosswalk mapping SHRIDs to Census 2011 district codes and state codes.

A.2 ODF Declaration Date Reconstruction

State-level ODF declaration dates were reconstructed from the following sources:

- Press Information Bureau (PIB) press releases (primary source for Tier 1 dates)

- PIB Year-End Review 2018 (PRID: 186600)

- PIB press release on 10 ODF states/UTs (January 24, 2018, relid: 175849)

- Daily Pioneer: "C'garh among 19 States declared ODF so far" (mid-2018)

- BankExamsToday: "11 States declared ODF under Swachh Bharat" (February 2018)

?? displays the complete state-level ODF timeline.

A.3 Sample Restrictions

Starting from the full SHRUG universe of approximately 596,000 villages/towns linked to 640 districts in 35 states/UTs, the panel includes:

- 640 districts with nonzero Census 2011 population and available VIIRS nightlights data

- 12 annual observations per district (2012–2023 for VIIRS)

- 7,680 total district–year observations

- 204 observations dropped from specifications using rural-only nightlights due to missing rural village data in 17 districts

A.4 Variable Definitions

Table 6: Variable Definitions

Variable	Definition
log_nl	$\log($VIIRS total luminosity $+ 1)$ for all villages in district
log_nl_rural	Same as above, restricted to villages with pop. $<$ 10,000
log_nl_pc	$\log($VIIRS total luminosity$/\text{pop}_{2011} + 0.001)$
post_odf	$= 1$ if year $\geq$ state's ODF declaration year
treat_weighted	$= 1$ for years after ODF; $= (12 - m + 1)/12$ in ODF year
cohort	State ODF declaration year (2016, 2017, 2018, or 2019)
rural_share	Share of district population in villages with pop. $< 10,000$

B. Identification Appendix

B.1 Pre-Trend Tests

The formal pre-test from the Callaway-Sant'Anna estimator rejects parallel trends with $p < 0.01$. Event study coefficients at relative times -5 through -2 are positive and monotonically

declining toward zero, indicating that early-declaring states had faster nightlights growth rates before ODF declaration.

The extended DMSP pre-trend analysis (2008–2013) shows much smaller and statistically insignificant differential trends at relative times -10 through -4 from treatment. This suggests that the differential growth trajectories emerged in the 2012–2016 period coinciding with SBM implementation decisions, rather than reflecting a longstanding structural divergence.

B.2 Urban Placebo

The urban placebo test restricts the sample to 50 urban-dominated districts (rural share $< 30\%$). SBM-G exclusively targets rural sanitation; urban areas should be unaffected. The estimated coefficient is -0.124 (SE $= 0.048$, $p = 0.02$), larger and more significant than the full-sample estimate. This "wrong-sign" placebo strongly suggests that the TWFE estimate reflects state-level confounders correlated with ODF timing, not a rural sanitation channel.

B.3 Fake Treatment Date Placebo

Shifting ODF dates backward by three years and restricting to pre-treatment observations yields a coefficient of -0.014 (SE $= 0.051$, $p = 0.79$). The absence of a fake-treatment effect in the pre-period indicates that the differential trends are gradual rather than exhibiting a sharp acceleration immediately before actual ODF declarations.

C. Robustness Appendix

C.1 Randomization Inference

I implement Fisher-type randomization inference by permuting ODF year assignments across states 500 times, re-estimating the TWFE coefficient each time. The distribution of placebo coefficients is centered near zero (mean $= -0.005$, SD $= 0.054$). The actual coefficient (-0.095) falls in the 7.6th percentile of the permutation distribution, corresponding to an RI p-value of 0.076. This is marginally significant but does not reject the null of no treatment effect at conventional levels.

C.2 Treatment Date Confidence Tiers

Restricting analysis to the 9 Tier 1 states (with PIB-confirmed ODF dates) yields -0.109 (SE $= 0.064$), consistent with the full-sample estimate. Restricting to mid-cohort states (2017–2018, the best-documented dates with the most balanced comparison groups) yields

-0.054 (SE $= 0.036$). Both estimates are consistent with the overall pattern and suggest that measurement error in ODF dates is not driving the results.

C.3 Alternative Nightlights Measures

Using mean luminosity per village (rather than total district luminosity) produces -0.098 (SE $= 0.053$). Using the extensive margin—lit area, measured as the number of grid cells with nonzero luminosity—produces $+0.004$ (SE $= 0.004$, $p = 0.31$). The near-zero lit area coefficient indicates that ODF declarations did not affect the spatial extent of economic activity.

C.4 Extended Pre-Trends (DMSP, 2008–2013)

Using calibrated DMSP nightlights for the pre-treatment period (2008–2013), I examine whether differential trends across ODF cohorts existed before the VIIRS sample begins. The DMSP analysis is summarized textually rather than as a figure, since DMSP and VIIRS luminosity are measured on different scales and cannot be plotted on a common axis. Regressing DMSP log nightlights on cohort-by-year interactions for the 2008–2013 period reveals flat differential trends: the interaction coefficients are small (all below 0.02 in absolute value) and statistically insignificant. This suggests that the differential growth trajectories visible in the VIIRS event study emerged specifically in the 2014–2016 period coinciding with SBM implementation decisions, rather than reflecting a longstanding structural divergence between early- and late-declaring states.

D. Heterogeneity Appendix

D.1 By SC/ST Share

Heterogeneous effects by baseline SC/ST share (a proxy for historical marginalization) show no differential pattern. Districts with above-median SC/ST population share have a Post-ODF coefficient of -0.109 (SE $= 0.075$), while below-median districts show -0.081 ($= -0.109 + 0.028$). The interaction is small and insignificant ($p = 0.72$).

D.2 By District Size

Large vs. small districts (by 2011 population) show similar effects: -0.085 (SE $= 0.042$) for large districts and -0.103 for small districts. The interaction is near zero (-0.018, SE $= 0.068$, $p = 0.79$).
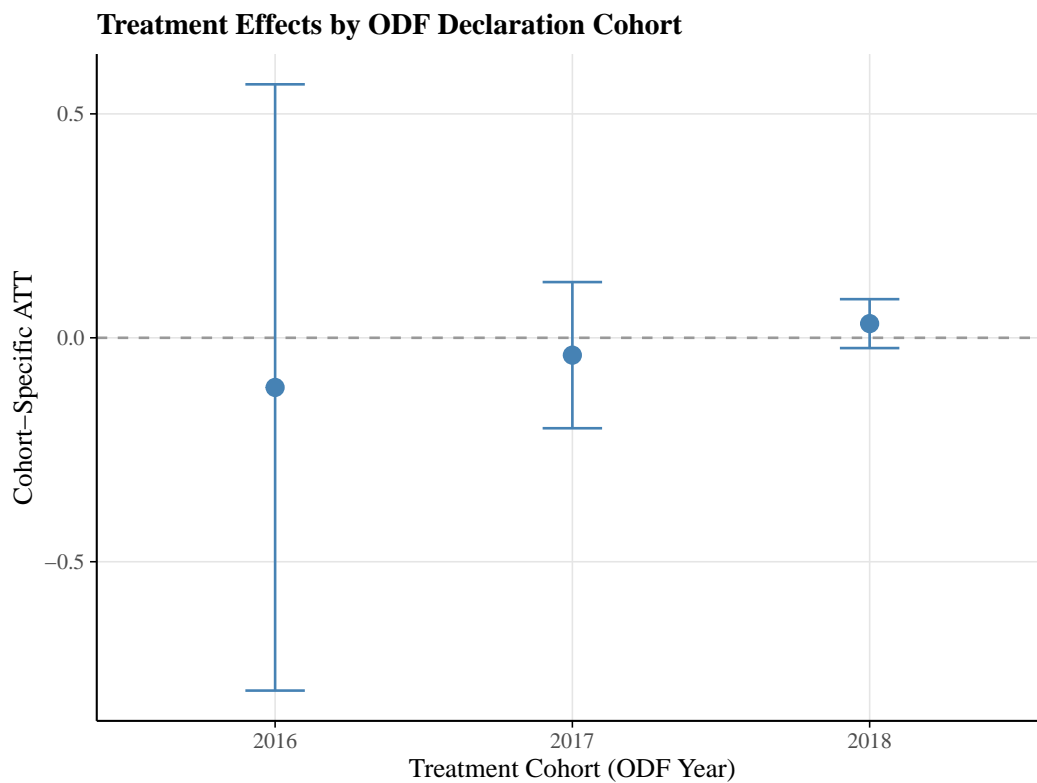
E. Additional Figures

Figure 6: Cohort-Specific Treatment Effects (CS-DiD)

Notes: Callaway-Sant'Anna cohort-specific ATTs with 95% confidence intervals. The 2016 cohort (3 states, 30 districts) is imprecisely estimated due to small sample size. The 2018 cohort (15 states, 231 districts) is the most precisely estimated and shows a small positive but insignificant effect.
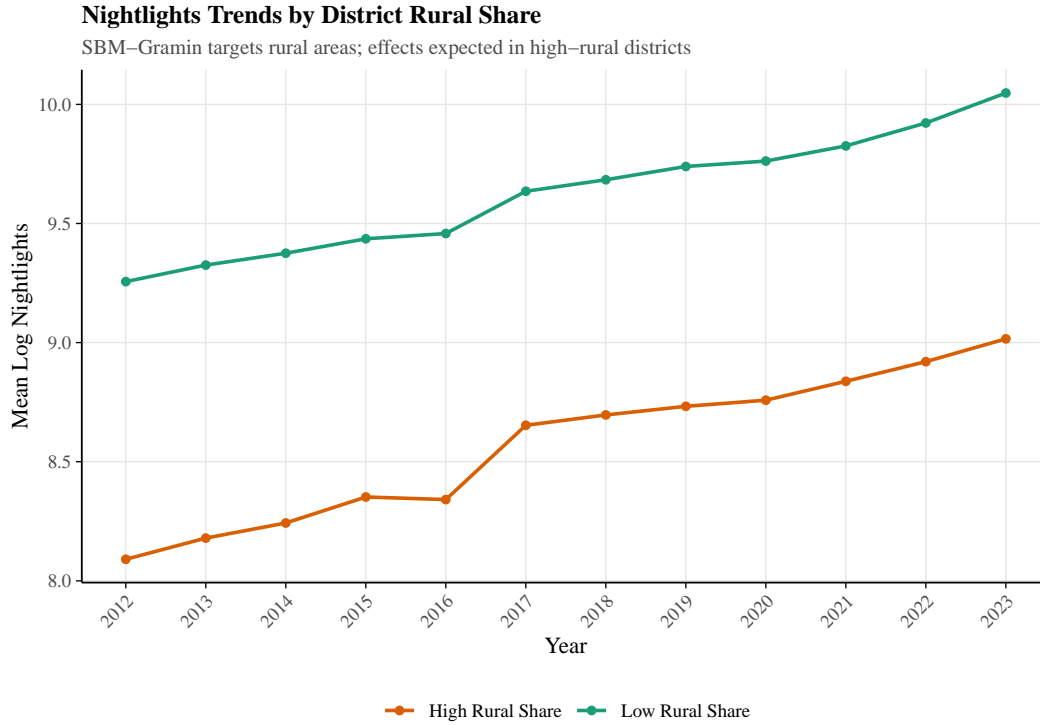
Nightlights Trends by District Rural Share

SBM–Gramin targets rural areas; effects expected in high–rural districts

Figure 7: Nightlights Trends by District Rural Share

Notes: Mean log nightlights for districts split at the median rural population share. High-rural-share districts (orange) start lower and grow more slowly throughout the sample period, consistent with a general urban-rural development divergence that confounds the ODF timing analysis.