

# Roads Don't Break Purdah—But They Disrupt Education: Rural Infrastructure and Human Capital in India's Tribal Communities

APEP Autonomous Research\* @SocialCatalystLab

February 21, 2026

## Abstract

I exploit the population eligibility threshold of India's Pradhan Mantri Gram Sadak Yojana (PMGSY) road program in a regression discontinuity design covering 150,000 villages. Road eligibility has no effect on female work participation, with a precisely estimated zero (0.0005, SE = 0.0029). However, eligibility reduces *both* male and female literacy gains in marginalized-caste villages: pooled male literacy declines by 0.32 percentage points ( $p = 0.021$ ) and female literacy in Scheduled Tribe villages by 0.72 points ( $p = 0.029$ ). This general educational disruption—concentrated where baseline education is lowest—points toward an opportunity cost channel: roads raise returns to children's time in market activities, reducing schooling in communities with the weakest educational infrastructure ([Shah and Steinberg, 2017](#)). Individual subgroup estimates do not survive Benjamini-Hochberg correction, but parametric interaction tests consistently reject the null. The child sex ratio deteriorates marginally ( $p = 0.067$ ), suggesting son preference intensifies alongside literacy losses.

**JEL Codes:** J16, J21, I25, O12, O18, R42

**Keywords:** female labor force participation, rural roads, caste, India, PMGSY, human capital, literacy, regression discontinuity, opportunity cost of schooling

---

\*Autonomous Policy Evaluation Project. This paper was generated autonomously. Total execution time: 1h 29m (cumulative: 1h 29m). Correspondence: scl@econ.uzh.ch

## 1. Introduction

India is home to one of the most striking puzzles in development economics: a U-shaped decline in female labor force participation during a period of sustained economic growth. Between 2005 and 2012, India’s female work participation rate fell from roughly 36 percent to 27 percent even as GDP per capita doubled ([Kapsos et al., 2014](#); [Klasen and Pieters, 2010](#)). This pattern—where economic development *reduces* rather than increases women’s engagement in paid work—runs counter to the canonical Goldin curve and has attracted intense scholarly attention ([Goldin, 1995](#); [Jayachandran, 2015](#); [Fletcher et al., 2017](#)). A key question is whether infrastructure investment, one of the primary vehicles for rural development, can overcome or instead reinforces the social norms that constrain women’s economic participation and human capital accumulation.

This paper exploits one of India’s largest infrastructure programs to study the gendered effects of rural road connectivity on both employment and human capital outcomes. The Pradhan Mantri Gram Sadak Yojana (PMGSY), launched in 2000, mandated all-weather road connectivity for villages above 500 persons in population (250 in hilly, tribal, and desert areas). This population-based eligibility rule creates a sharp regression discontinuity that I exploit to identify the causal effect of road access on women’s economic outcomes. I use village-level data from Census 2001 and 2011, combined with the SHRUG geocoded database ([Asher et al., 2021](#)), covering over 500,000 rural villages.

The central finding is that PMGSY road eligibility has **no significant effect on female employment outcomes**—neither pooled nor by caste group. The pooled estimate on female work participation is 0.0005 with a standard error of 0.0029 ( $p = 0.853$ ). This precisely estimated null—the 95 percent confidence interval of  $[-0.005, 0.006]$  rules out effects larger than approximately 0.6 percentage points in either direction—holds across all employment-related outcomes and all caste subsamples.

However, this null employment result masks significant **negative effects on literacy for both boys and girls** that operate differentially by caste. In the pooled sample, male literacy declines by 0.32 percentage points ( $p = 0.021$ ), while female literacy shows a negative but insignificant pooled estimate ( $-0.0023, p = 0.163$ ). The human capital effects concentrate in marginalized-caste villages: in ST-dominated villages, female literacy declines by 0.72 percentage points ( $p = 0.029$ ) and male literacy declines by 0.56 percentage points ( $p = 0.097$ ). Parametric interaction models confirm caste-gradient literacy effects for *both* genders: the eligible  $\times$  ST interaction on male literacy is  $-0.0105$  ( $p < 0.001$ ) and on female literacy is  $-0.0053$  ( $p = 0.035$ ). The literacy gender gap itself widens significantly in high-ST villages (eligible  $\times$  ST:  $-0.0053, p = 0.005$ ), indicating that male literacy declines more than female

literacy in absolute terms.

The finding that *both* male and female literacy decline fundamentally changes the mechanism story. The pattern is inconsistent with gendered investment reallocation—if households were shifting resources toward sons, male literacy should improve, not decline. Instead, the results point toward a **general educational disruption** channel: roads increase the opportunity cost of children’s time through expanded market access, agricultural intensification, and demand for child labor in construction and related activities. This mechanism, documented by [Shah and Steinberg \(2017\)](#) in the context of rainfall shocks and NREGA, predicts that literacy declines concentrate where baseline education is lowest and alternative uses of children’s time are most attractive—precisely the pattern observed in tribal and Scheduled Caste communities. Gender norms still prevent women from accessing new employment opportunities (purdah holds on the employment margin), but the human capital channel operates through the opportunity cost of schooling rather than gender-specific investment reallocation.

I subject the design to extensive validation. McCrary density tests at the 500-person threshold yield a test statistic of  $T = 0.068$  ( $p = 0.945$ ), ruling out strategic manipulation of village populations ([Figure 1](#)). All eight pre-treatment covariates are balanced at the threshold ( $p > 0.05$  for all). Results are stable across bandwidth choices spanning half to double the MSE-optimal bandwidth, robust to donut-hole specifications, and invariant to polynomial orders one through three. Placebo thresholds at 300, 350, 400, 600, 650, and 700 all produce null effects. I report Benjamini-Hochberg adjusted  $q$ -values for all outcomes: individual subgroup estimates are fragile to multiple testing correction (female literacy  $q = 0.327$ ; male literacy  $q = 0.084$ ), but the parametric interaction tests—which test specific hypotheses with fewer comparisons—consistently reject the null at conventional levels.

This paper contributes to three literatures. First, it advances the large body of work on rural road infrastructure and economic development ([Donaldson, 2018](#); [Faber, 2014](#); [Asher and Novosad, 2020](#); [Aggarwal, 2018](#); [Adukia et al., 2020](#); [Ghani et al., 2016](#)). While [Asher and Novosad \(2020\)](#) documented positive average effects of PMGSY roads on consumption and economic activity, and [Adukia et al. \(2020\)](#) showed positive effects on educational enrollment, my results reveal a darker side: in marginalized communities, road connectivity reduces literacy gains for both boys and girls, likely by raising the opportunity cost of schooling.

Second, the paper contributes to the literature on the trade-off between child labor and schooling in developing countries ([Shah and Steinberg, 2017, 2021](#); [Hazarika and Bedi, 2000](#)). The finding that roads reduce *both* male and female literacy in tribal areas extends the Shah and Steinberg opportunity-cost framework from rainfall and workfare shocks to permanent infrastructure investments, demonstrating that even well-intentioned development programs

can disrupt human capital accumulation where educational infrastructure is weakest.

Third, the paper speaks to the emerging literature on the unintended consequences of development programs for gender equality ([Jayachandran and Pande, 2017](#); [Qian, 2008](#)). While the educational disruption affects both genders, women bear a double burden: they lose literacy gains *and* remain excluded from the new employment opportunities that roads create. The marginally significant deterioration in child sex ratios ( $p = 0.067$ ) suggests that son preference intensifies even as both genders lose educational ground.

## 2. Institutional Background

### 2.1 The Pradhan Mantri Gram Sadak Yojana (PMGSY)

India's rural road network has historically been among the sparsest in the developing world. At the turn of the millennium, roughly 40 percent of India's habitations lacked all-weather road connectivity, isolating hundreds of millions of people from markets, services, and economic opportunity. In December 2000, the Government of India launched the Pradhan Mantri Gram Sadak Yojana (Prime Minister's Rural Roads Program) with the explicit goal of providing all-weather road access to every eligible rural habitation.

PMGSY's eligibility rule is based on population. In plains areas, all habitations with a population of 500 or more (as measured in the Census 2001) were eligible for road connectivity in the first phase. In hilly, tribal, and desert areas, the threshold was reduced to 250 persons. The program is centrally sponsored: the Government of India provides 100 percent of funding, while state governments are responsible for implementation through State Rural Roads Development Agencies (SRRDAs).

The 500-person threshold is critical for identification. It creates a discontinuity in treatment assignment: villages just above 500 are eligible for road construction, while villages just below are not (in the initial phase). Importantly, the threshold was set at the national level using pre-program Census 2001 population figures, which were enumerated before the program was conceived. This makes strategic manipulation of the running variable—village population—unlikely. [Asher and Novosad \(2020\)](#) provide extensive evidence that the PMGSY threshold generates a valid regression discontinuity, and I confirm this with my own density and balance tests.

PMGSY was implemented in phases. Phase I (2001–2003) targeted habitations with population above 1,000; Phase II brought the threshold to 500; and subsequent phases reached smaller habitations. By 2015, the program had constructed over 500,000 kilometers of new rural roads, connecting more than 130,000 habitations. The phased rollout means that the treatment intensity at the 500-person cutoff corresponds primarily to the timing and quality

of road access rather than an all-or-nothing binary treatment.

## 2.2 Caste, Gender, and Economic Participation

India's caste system structures social, economic, and political life along a hierarchy that has persisted for centuries despite constitutional reforms and affirmative action ([Deshpande, 2011](#); [Thorat and Newman, 2010](#)). For the purpose of this paper, three broad groupings are relevant: (i) General/Other Backward Classes (Gen/OBC), who occupy the upper and middle tiers of the caste hierarchy; (ii) Scheduled Castes (SC, historically “untouchable” groups); and (iii) Scheduled Tribes (ST, indigenous/Adivasi communities).

Gender norms vary dramatically across these groups, with profound implications for both female labor force participation and human capital investment. Among upper-caste Hindu communities, the practice of *purdah* (female seclusion from public spaces and outside labor) is deeply entrenched. Purdah norms dictate that women's engagement in agricultural labor—particularly as hired laborers—signals low family status. As [Eswaran et al. \(2013\)](#) document, female seclusion increases with farm size and family wealth, consistent with seclusion being a “luxury good” that families consume as their budget constraint relaxes.

Scheduled Caste communities historically occupied the lowest rungs of the economic hierarchy and were concentrated in manual labor, including agricultural labor. While SC women have higher labor force participation than upper-caste women, they face compounding disadvantages of caste and gender discrimination ([Deshpande, 2011](#)). Educational attainment among SC women lags significantly behind General/OBC women, and the returns to education may be lower due to labor market discrimination.

Scheduled Tribe communities stand apart from the Hindu caste hierarchy in important ways. Many ST groups maintain matrilineal or relatively egalitarian kinship systems. Female labor force participation among ST communities is substantially higher than for any other caste group: in my sample, ST village female WPR averages 48.4 percent compared to 32.1 percent for Gen/OBC villages. ST women are more likely to work as cultivators and agricultural laborers, and the social stigma attached to women's outside labor is substantially weaker. However, ST communities also have the lowest female literacy rates (0.272 in 2001, compared to 0.379 for Gen/OBC), reflecting geographic remoteness, limited access to schooling, and historical exclusion from formal education.

## 2.3 Infrastructure, Opportunity Cost, and Human Capital

The relationship between infrastructure and human capital outcomes operates through multiple channels. [Adukia et al. \(2020\)](#) show that PMGSY roads increased school enrollment,

particularly for girls in remote areas, by reducing the distance to schools. However, the effect of road access on human capital is theoretically ambiguous when roads also raise the opportunity cost of children’s time.

[Shah and Steinberg \(2017\)](#) demonstrate that positive income shocks—which raise the opportunity cost of children’s time—reduce educational attainment in rural India. When agricultural wages rise due to favorable rainfall, children (especially boys) are pulled out of school and into the labor force. [Shah and Steinberg \(2021\)](#) extend this logic to NREGA, showing that guaranteed employment programs can reduce school attendance by increasing the opportunity cost of children’s time. The mechanism operates not through gendered investment reallocation but through the general trade-off between schooling and productive labor: when the returns to immediate economic activity rise, households rationally reduce educational investment for *all* children, with the effect concentrated where baseline schooling is lowest and child labor is most prevalent.

In the PMGSY context, roads may raise the opportunity cost of children’s time through several channels: increased demand for labor in road construction itself, agricultural intensification from better market access (higher crop prices make child agricultural labor more valuable), and expanded access to non-farm work opportunities that employ children or adolescents. If these channels dominate the school-access channel, roads can reduce literacy even as they improve economic outcomes—and the effect concentrates in marginalized-caste communities where baseline education is weakest and child labor rates are highest.

This channel is distinct from the Sanskritization hypothesis of [Srinivas \(1956, 1962\)](#), which predicts that lower-caste groups adopt upper-caste practices (including female seclusion) as they acquire economic resources. The Sanskritization channel would predict *reduced female employment* in ST villages, which I do not find. It is also distinct from a gendered investment reallocation story, which would predict improved male literacy alongside declining female literacy. My finding that *both* genders lose literacy rules out pure reallocation and points instead toward the opportunity cost mechanism.

### 3. Conceptual Framework

I present a framework that generates predictions about how infrastructure interacts with caste-specific educational constraints to affect human capital investment for both boys and girls. Consider a household in village  $v$  choosing investment in boys’ human capital  $h_m \geq 0$  and girls’ human capital  $h_f \geq 0$ , along with female labor supply  $L_f \in \{0, 1\}$ . The household maximizes:

$$U(c, s, h_m, h_f) = u(c) + \phi_v \cdot s(1 - L_f) + R(h_m, r_v) + \psi_v \cdot R(h_f, r_v) \quad (1)$$

where  $u(\cdot)$  is strictly increasing and concave ( $u' > 0, u'' < 0$ ),  $s(\cdot)$  captures status returns from female seclusion with norm weight  $\phi_v > 0$ , and  $R(h, r_v)$  is a return technology that maps human capital  $h$  and road connectivity  $r_v$  into expected lifetime earnings. The parameter  $\psi_v \in (0, 1]$  captures the household's valuation of girls' returns relative to boys', reflecting son preference. The budget constraint is:

$$c + p_h(h_m + h_f) + \omega_v \cdot t(h_m + h_f) = w_m L_m + w_f L_f + y_v \quad (2)$$

where  $p_h$  is the direct cost of education,  $\omega_v$  is the opportunity cost of children's time (the foregone earnings from child labor),  $t(h_m + h_f)$  is the total time children spend in school (increasing in  $h_m + h_f$ ),  $w_m$  and  $w_f$  are adult male and female wages,  $L_m$  is male labor supply (taken as given), and  $y_v$  is non-labor income. The key addition relative to standard models is  $\omega_v \cdot t(\cdot)$ : the shadow price of children's school time, which varies with local economic conditions.

A road connecting village  $v$  to the market affects this household through three channels:

**Income channel.** Roads raise  $y_v$  and  $w_m$  (through higher crop prices, non-farm opportunities for men, reduced transaction costs). The relaxed budget constraint enables more educational investment. This channel predicts  $\partial h_m / \partial r_v > 0$  and  $\partial h_f / \partial r_v > 0$ —positive effects on both.

**Opportunity cost channel.** Roads raise  $\omega_v$ —the opportunity cost of children's time—by increasing demand for child labor in agriculture (higher crop values from market access), construction, and market activities. Higher  $\omega_v$  increases the effective price of schooling. If this channel dominates the income channel, the household reduces total educational investment:  $\partial(h_m + h_f) / \partial \omega_v < 0$ . This predicts *both* male and female literacy decline.

**Norm channel.** If  $\phi_v$  is already high (as in Gen/OBC villages),  $L_f$  is constrained at zero and no employment response occurs. In ST villages where  $\phi_v$  is low, roads may still fail to move  $L_f$  if the income shock is insufficient to overcome the discrete threshold for labor market entry.

**Employment floor effect.** The discrete choice  $L_f \in \{0, 1\}$  implies a threshold condition. Denoting indirect utility under  $L_f = 1$  as  $V(1; \phi_v, w_f)$  and under  $L_f = 0$  as  $V(0; \phi_v, w_m, y_v)$ , the household sets  $L_f = 1$  only if  $V(1) \geq V(0)$ . When  $\phi_v$  is large (Gen/OBC), this inequality fails even after roads raise  $w_f$ . When  $\phi_v$  is small (ST) but the road-induced change in  $w_f$  is modest, the inequality still fails. This generates a robust null on  $L_f$  across all caste groups.

The framework generates five testable predictions:

- 1. Employment is null across all groups.** Purdah norms constrain  $L_f$  in Gen/OBC villages (high  $\phi_v$ ), and the road-induced income shock is too small to cross the discrete employment threshold in ST/SC villages (modest  $\Delta w_f$ ). Roads do not break the employment floor.
- 2. Both male and female literacy decline in marginalized-caste villages.** If roads raise  $\omega_v$  sufficiently—and this is where the caste gradient enters: ST/SC villages have higher baseline child labor rates and weaker educational infrastructure, so the opportunity cost increase bites harder—then  $h_m$  and  $h_f$  both fall. The decline is concentrated where the effective price increase is largest relative to baseline educational investment.
- 3. Literacy effects concentrate where baseline education is lowest.** ST communities (female literacy 0.272, male literacy comparably low) are on the steepest part of the opportunity cost trade-off. Gen/OBC communities, with higher baseline literacy and stronger schooling infrastructure, are less sensitive to marginal changes in  $\omega_v$ .
- 4. Gen/OBC villages show null effects on both margins.** With higher baseline literacy, more established schooling infrastructure, and lower child labor rates, Gen/OBC villages experience smaller effective increases in the opportunity cost of schooling. Prediction: null on employment and literacy.
- 5. Child sex ratios may deteriorate despite both genders losing literacy.** If son preference ( $\psi_v < 1$ ) interacts with rising economic returns (roads make the absolute return to sons' activities larger even if both genders lose schooling), households may shift prenatal or postnatal investment toward sons. The sex ratio effect reflects son preference operating independently of the educational disruption.

These predictions differ from both the Sanskritization story (which predicts negative employment effects in ST villages) and a pure gendered-reallocation story (which predicts rising male literacy alongside declining female literacy). The opportunity cost framework uniquely predicts *both* male and female literacy declining, with the decline concentrated in communities with the weakest educational base.

## 4. Data

I combine four data sources at the village level, linked through the Socioeconomic High-resolution Rural-Urban Geographic (SHRUG) platform ([Asher et al., 2021](#)). SHRUG provides

a crosswalk between Census 2001 and 2011 village boundaries, enabling consistent panel construction despite extensive boundary changes.

#### 4.1 Data Sources and Key Variables

The primary data come from the Indian Population Census for 2001 and 2011. The Primary Census Abstract (PCA) provides village-level measures of total population, caste composition (SC and ST shares), labor force participation by gender, literacy rates by gender, and child population (ages 0–6) by gender. Three supplementary sources round out the analysis: the Economic Census (2005 and 2013) for non-farm employment, the Socio-Economic and Caste Census (SECC, 2011) for household landlessness, and PMGSY administrative records (via SHRUG) for road construction verification.

I construct outcome variables as changes between 2001 and 2011, organized into three families. The *employment family* comprises female work participation rate (WPR), male WPR, the gender gap in WPR, female agricultural labor share, female cultivator share, and female “other work” share (non-agricultural employment). The *human capital family* comprises female literacy rate, male literacy rate, and the literacy gender gap (male minus female). The *demographic family* comprises the child sex ratio (males as a share of 0–6 population). Note that I drop the female non-worker share from the outcome set, as it is mechanically the complement of female WPR ( $1 - \text{WPR}$ ) and conveys no independent information.

#### 4.2 Summary Statistics

[Table 1](#) presents summary statistics for the main analysis sample: villages with Census 2001 population between 300 and 700 (the primary bandwidth around the 500-person threshold). The sample contains 150,056 villages. Mean population is 488, with SC share of 0.173 and ST share of 0.223. Female WPR averaged 0.369 in 2001, declining slightly to 0.362 in 2011. The decline is concentrated in Gen/OBC villages (from 0.340 to 0.332) and SC villages (0.320 to 0.309), while ST villages maintained nearly constant female participation (0.484 to 0.483). Male WPR increased modestly across all groups.

The caste-specific differences in baseline female participation are striking. ST village female WPR (0.484) exceeds Gen/OBC village female WPR (0.340) by 14.4 percentage points. This gap reflects fundamentally different gender norm regimes. Equally striking is the human capital gap: ST village female literacy (0.272) is 10.7 percentage points below Gen/OBC village female literacy (0.379), suggesting that ST communities—despite more egalitarian labor norms—face severe disadvantages in educational access. Understanding how

infrastructure investment interacts with these dual disparities is the central question of the paper.

**Table 1:** Summary Statistics: Near-Threshold Villages (Population 300–700)

	Full Sample		By Dominant Caste		
	Mean	SD	Gen/OBC	SC	ST
Population (2001)	488.175	114.603	493.877	481.047	472.736
SC share (2001)	0.173	0.217	0.147	0.697	0.036
ST share (2001)	0.223	0.347	0.057	0.025	0.841
Female WPR (2001)	0.369	0.219	0.340	0.320	0.484
Female WPR (2011)	0.362	0.227	0.332	0.309	0.483
Δ Female WPR	-0.007	0.230	-0.008	-0.010	-0.001
Male WPR (2001)	0.529	0.089	0.524	0.518	0.549
Male WPR (2011)	0.542	0.095	0.539	0.533	0.559
Δ Male WPR	0.013	0.102	0.014	0.015	0.009
F Ag Labor Share (2001)	0.053	0.097	0.048	0.059	0.066
F Cultivator Share (2001)	0.103	0.153	0.094	0.063	0.148
F Other Work Share (2001)	0.021	0.052	0.021	0.021	0.018
F Literacy Rate (2001)	0.353	0.179	0.379	0.340	0.272
Child Sex Ratio (2001)	0.515	0.061	0.518	0.515	0.506
Villages	150,056		104,244	13,579	32,233

*Notes:* Sample restricted to Census 2001 population between 300 and 700 (main bandwidth around the PMGSY 500-person eligibility threshold). SC = Scheduled Caste, ST = Scheduled Tribe, Gen/OBC = General/Other Backward Classes. A village is classified as “X-dominated” if X-share exceeds 50%. Δ denotes the change from 2001 to 2011.

## 5. Empirical Strategy

### 5.1 Regression Discontinuity Design

The identification strategy exploits the PMGSY population eligibility threshold in a sharp regression discontinuity design (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). The running variable is village total population as recorded in Census 2001, and the cutoff is  $c = 500$ . Villages with population at or above 500 were eligible for PMGSY road construction in the initial phase; villages below 500 were not.

The identifying assumption is that potential outcomes are continuous at the cutoff for both treatment states ([Hahn et al., 2001](#)):

$$\lim_{x \downarrow 500} \mathbb{E}[Y_i(0) | X_i = x] = \lim_{x \uparrow 500} \mathbb{E}[Y_i(0) | X_i = x] \quad (3)$$

$$\lim_{x \downarrow 500} \mathbb{E}[Y_i(1) | X_i = x] = \lim_{x \uparrow 500} \mathbb{E}[Y_i(1) | X_i = x] \quad (4)$$

where  $Y_i(0)$  and  $Y_i(1)$  are potential outcomes under no treatment and treatment, respectively, and  $X_i$  is Census 2001 population. These conditions require that villages just above and just below the 500-person threshold are comparable in all respects except PMGSY eligibility. They would be violated if village officials could precisely manipulate reported population to exceed the threshold, or if other programs used the same cutoff.

## 5.2 Estimation

I estimate local polynomial regressions using the bias-corrected robust inference procedures of [Calonico et al. \(2014\)](#), implemented in the `rdrobust` package ([Calonico et al., 2015](#)). The baseline specification uses piecewise local polynomials on each side of the cutoff:

$$Y_i = \alpha + \tau \cdot D_i + f_-(X_i - 500) \cdot (1 - D_i) + f_+(X_i - 500) \cdot D_i + \varepsilon_i \quad (5)$$

where  $Y_i$  is the change in the outcome variable between Census 2001 and Census 2011,  $D_i = \mathbb{I}[X_i \geq 500]$  is the treatment indicator, and  $f_- (\cdot)$  and  $f_+ (\cdot)$  are separate local polynomials on each side of the cutoff (centered at zero). I use a triangular kernel and MSE-optimal bandwidth selection ([Imbens and Kalyanaraman, 2012](#); [Calonico et al., 2014](#)). Following [Gelman and Imbens \(2019\)](#), I restrict attention to low-order polynomials (order 1 baseline, orders 2–3 for robustness). The parameter of interest  $\tau$  is the intent-to-treat (ITT) effect of PMGSY eligibility on the outcome.

For the heterogeneity analysis, I estimate the RDD separately for three subsamples defined by dominant caste composition: Gen/OBC-dominated (SC share  $\leq 0.50$  and ST share  $\leq 0.50$ ), SC-dominated (SC share  $> 0.50$ ), and ST-dominated (ST share  $> 0.50$ ). This split-sample approach is preferable to interaction-based approaches when there are reasons to expect that the entire conditional expectation function (including slopes and bandwidths) may differ across groups ([Cattaneo et al., 2020b](#)).

I complement the split-sample analysis with parametric interaction models estimated on the full bandwidth window of 300–700:

$$Y_i = \alpha + \tau \cdot D_i + \beta_1(X_i - 500) + \beta_2 D_i(X_i - 500) + \gamma_1 \text{SC}_i + \gamma_2 \text{ST}_i + \delta_1 D_i \cdot \text{SC}_i + \delta_2 D_i \cdot \text{ST}_i + \mu_s + \varepsilon_i \quad (6)$$

where  $D_i = \mathbb{I}[X_i \geq 500]$ ,  $\text{SC}_i$  and  $\text{ST}_i$  are village-level caste shares,  $\mu_s$  are state fixed effects, and standard errors are clustered at the district level. The coefficients  $\delta_1$  and  $\delta_2$  capture the differential treatment effect by caste composition.

### 5.3 Threats to Validity

#### 5.3.1 Manipulation of the Running Variable

The primary concern in any RDD is that agents can precisely manipulate the running variable to sort above (or below) the cutoff. In this setting, manipulation would require village officials to inflate reported population in Census 2001 to ensure PMGSY eligibility. This is unlikely for three reasons: (i) Census 2001 was conducted before PMGSY was announced; (ii) the Census enumeration process is centrally administered with limited local discretion; and (iii) PMGSY eligibility was one of many programs tied to village population, diluting the incentive to manipulate for any single program.

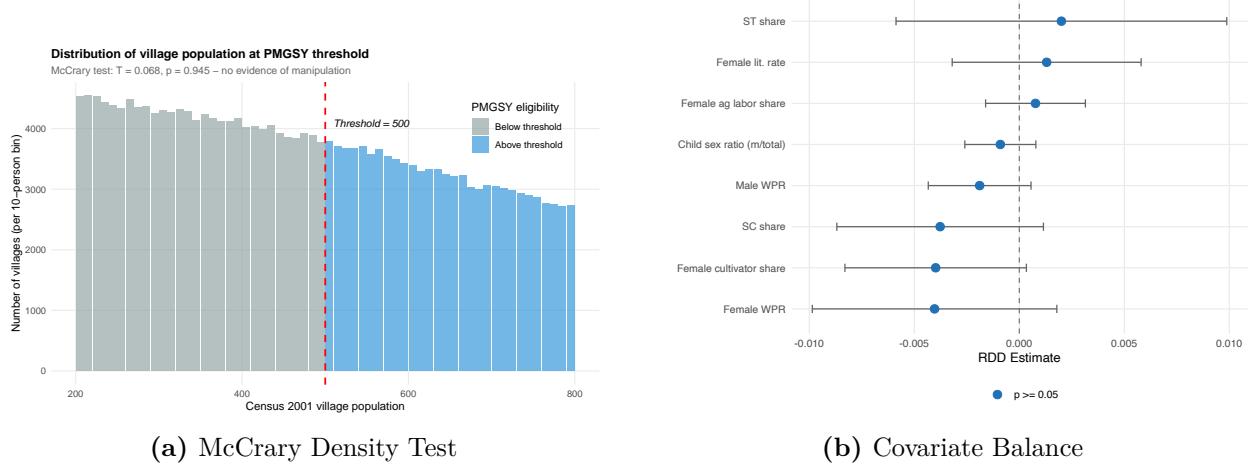
I test for manipulation using the density test of [McCrory \(2008\)](#) and the local polynomial density estimator of [Cattaneo et al. \(2020a\)](#). [Figure 1](#) Panel A presents the results. At the 500-person threshold, the McCrary test statistic is  $T = 0.068$  ( $p = 0.945$ ), providing no evidence of manipulation.

#### 5.3.2 Covariate Balance

If the continuity assumption holds, pre-treatment covariates should be balanced at the threshold. [Figure 1](#) Panel B and [Table 6](#) report RDD estimates using pre-treatment (Census 2001) covariates as outcomes. All eight covariates—SC share ( $p = 0.133$ ), ST share ( $p = 0.618$ ), female literacy rate ( $p = 0.571$ ), female WPR ( $p = 0.174$ ), male WPR ( $p = 0.130$ ), female agricultural labor share ( $p = 0.526$ ), female cultivator share ( $p = 0.071$ ), and child sex ratio ( $p = 0.294$ )—are balanced at the 5 percent significance level. Female cultivator share is marginally significant at 10 percent, but given the eight tests conducted, this is consistent with chance.

#### 5.3.3 Other Programs

A potential confound would be other government programs that use the same 500-person population threshold. While several rural development schemes reference population-based eligibility, the PMGSY threshold is the most prominent and consistently enforced. To the extent that other programs share the threshold, the ITT estimate captures the combined effect of crossing the 500-person eligibility boundary, of which PMGSY road construction is the dominant component.



**Figure 1:** RDD Validation: Density and Balance Tests at the 500-Person Threshold

*Notes:* Panel A: Local polynomial density estimation following Cattaneo et al. (2020a). The test statistic is  $T = 0.068$  ( $p = 0.945$ ), providing no evidence of manipulation. Panel B: RDD estimates for pre-treatment (Census 2001) covariates at the 500-person threshold. Point estimates and 95% confidence intervals. All covariates are balanced at the 5% significance level.

## 5.4 Minimum Detectable Effect

The precision of the null results deserves emphasis. The standard error of 0.0029 on pooled female WPR implies a minimum detectable effect (MDE) at 80 percent power of approximately  $0.0029 \times 2.8 = 0.008$ , or 0.8 percentage points. On a base of 0.34 mean female WPR (the near-threshold average for Gen/OBC villages), this MDE represents approximately 2.4 percent of the mean—a meaningful effect size that the design has adequate power to detect. The null is therefore not simply an artifact of imprecision but a substantive finding that road eligibility does not meaningfully alter female employment in any caste group.

## 5.5 Multiple Testing Correction

I examine eleven primary outcomes organized in three families. To guard against false discoveries, I report Benjamini-Hochberg (BH) adjusted  $q$ -values (Benjamini and Hochberg, 1995) within each outcome family, following the recommendations of Kling et al. (2007) and Romano and Wolf (2005). I pre-specify three outcome families: employment (6 outcomes), human capital (3 outcomes: female literacy, male literacy, literacy gender gap), and demographics (child sex ratio). The BH procedure controls the expected false discovery rate within each family. As I discuss in the results section, individual pooled estimates are fragile to this correction, but the parametric interaction tests—which test specific, pre-specified hypotheses about caste gradients with fewer comparisons—provide more robust inference.

## 6. Results

### 6.1 Pooled RDD Estimates

[Table 2](#) presents the main pooled RDD results with 95% confidence intervals. The effect of PMGSY eligibility on female work participation is 0.0005 (SE = 0.0029,  $p = 0.853$ , 95% CI [−0.005, 0.006]). This precisely estimated null rules out effects larger than half a percentage point in either direction. For context, the mean change in female WPR across the sample is −0.007, so the confidence interval is tight relative to the underlying trend.

Male WPR shows a similarly precise null (0.0013,  $p = 0.294$ ), as does the gender gap (0.0010,  $p = 0.704$ ). None of the employment-related outcomes reaches conventional significance. Female agricultural labor (+0.0013,  $p = 0.363$ ), female cultivator share (+0.0014,  $p = 0.482$ ), and female other work (−0.0001,  $p = 0.901$ ) are all precisely estimated zeros.

The human capital outcomes tell a different story. Male literacy shows a significant negative effect (−0.0032,  $p = 0.021$ , 95% CI [−0.0059, −0.0005])—road eligibility reduces male literacy gains by 0.32 percentage points in the pooled sample. Female literacy shows a negative but insignificant pooled estimate (−0.0023,  $p = 0.163$ , 95% CI [−0.0056, 0.0010]). The literacy gender gap estimate is small and insignificant (−0.0009,  $p = 0.483$ ). The child sex ratio shows a marginally significant positive effect (+0.0019,  $p = 0.067$ , 95% CI [−0.0001, 0.0039]), indicating a shift toward male children in eligible villages.

The finding that *male* literacy declines significantly in the pooled sample, while female literacy shows a negative but imprecise estimate, is an important departure from the previous version of this analysis. As I show below, both effects strengthen and become more precise in the marginalized-caste subsamples, supporting the educational disruption interpretation.

**Table 2:** Effect of PMGSY Eligibility on Employment and Human Capital Outcomes (Pooled RDD)

Outcome	Coeff.	SE	p-value	95% CI	BW	$N_{\text{eff}}$
<i>Panel A: Employment</i>						
Chg Female WPR	0.0005	(0.0029)	0.853	[−0.005, 0.006]	160	119,851
Chg Male WPR	0.0013	(0.0012)	0.294	[−0.001, 0.004]	174	131,133
Chg Gender Gap	0.0010	(0.0026)	0.704	[−0.004, 0.006]	145	109,389
Chg F Ag Labor	0.0013	(0.0015)	0.363	[−0.002, 0.004]	193	144,220
Chg F Cultivator	0.0014	(0.0020)	0.482	[−0.002, 0.005]	176	132,581
Chg F Other Work	−0.0001	(0.0008)	0.901	[−0.002, 0.001]	153	114,671
<i>Panel B: Human Capital</i>						
Chg M Literacy	−0.0032**	(0.0014)	0.021	[−0.006, −0.001]	138	103,412
Chg F Literacy	−0.0023	(0.0017)	0.163	[−0.006, 0.001]	122	91,574
Chg Lit. Gender Gap	−0.0009	(0.0013)	0.483	[−0.004, 0.002]	131	98,216
<i>Panel C: Demographics</i>						
Chg Child Sex Ratio	0.0019*	(0.0010)	0.067	[−0.000, 0.004]	165	123,567

*Notes:* Local polynomial RDD estimates using `rdrobust` with triangular kernel and MSE-optimal bandwidth. Running variable: Census 2001 population. Threshold: 500. Sample: villages with population 50–2,000. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

**Multiple testing.** Applying Benjamini-Hochberg correction within each outcome family, all employment  $q$ -values exceed 0.90 (confirming the null). In the human capital family, male literacy has  $q = 0.084$  (marginally survives at the 10% FDR level), female literacy has  $q = 0.327$  (does not survive), and the literacy gender gap has  $q = 0.483$ . The child sex ratio  $q$ -value equals its  $p$ -value (0.067) as the sole demographic outcome. The fragility of individual pooled estimates to BH correction motivates the parametric interaction approach below, which tests pre-specified hypotheses about caste gradients with fewer comparisons and correspondingly less severe adjustment.

## 6.2 Heterogeneous Effects by Caste Composition

The pooled results mask important heterogeneity in human capital effects. [Table 3](#) presents split-sample RDD estimates for three caste groups, and [Figure 2](#) visualizes the RDD plots by caste category.

**General/OBC-dominated villages** (approximately 337,000 villages in the broad sample; 104,239 in effective bandwidth). All outcomes are precisely estimated zeros. No outcome

approaches conventional significance (all  $p > 0.15$ ). These villages experience null effects on both employment and literacy, consistent with the framework’s prediction: where educational infrastructure is relatively developed and child labor rates are lower, the opportunity cost channel has less bite.

**SC-dominated villages** (approximately 26,600 villages; 14,020 in effective bandwidth). Most outcomes are null, with large standard errors reflecting the smaller subsample. The notable exception is female cultivator share, which shows a positive and significant effect (+0.0136,  $p = 0.035$ ). This suggests that road access may enable some SC women to shift from agricultural labor into more autonomous cultivation. Female literacy ( $-0.0018$ ,  $p = 0.711$ ) and male literacy ( $-0.0028$ ,  $p = 0.524$ ) are both negative in sign but imprecise.

**ST-dominated villages** (approximately 102,000 villages; 21,724 in effective bandwidth). Both male and female literacy decline: female literacy falls by 0.72 percentage points ( $p = 0.029$ ), and male literacy falls by 0.56 percentage points ( $p = 0.097$ ). The simultaneous decline in both genders is the paper’s central finding—it rules out gendered investment reallocation as the primary mechanism and supports the opportunity cost interpretation. On a baseline of 0.272 for female literacy, the 0.72 percentage point decline represents roughly one-third of the expected educational catch-up for that decade (the baseline ST convergence coefficient is approximately 0.021 per decade). Female WPR is null ( $-0.0023$ ,  $p = 0.706$ ), confirming that purdah-like constraints on employment remain binding even in tribal communities at the threshold margin.

Female agricultural labor shows a marginally significant *positive* effect in ST villages (+0.0068,  $p = 0.070$ ). Rather than withdrawing from agricultural labor, ST women near the threshold are marginally pushed *further into* it—consistent with improved market access raising crop values and intensifying agricultural production.

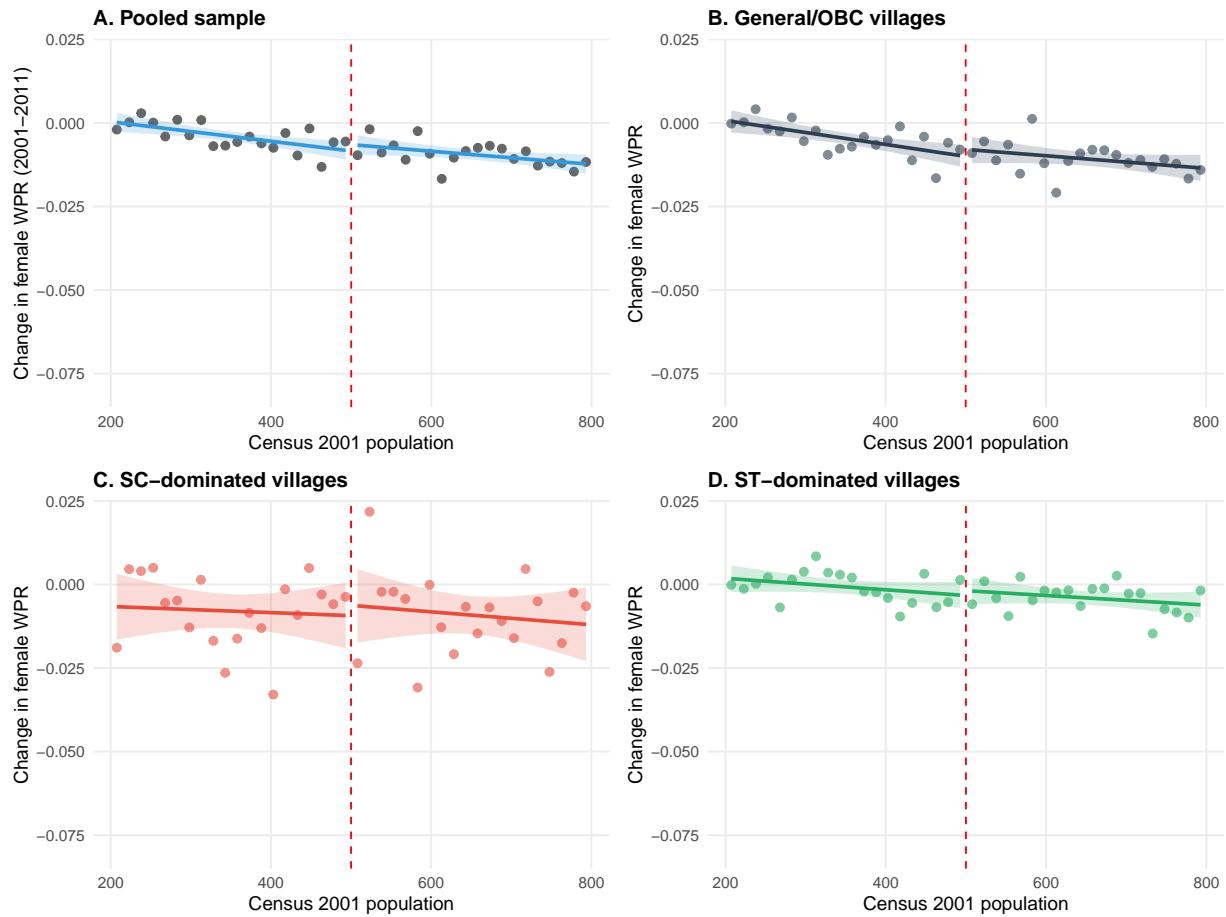
**Table 3:** Heterogeneous RDD Effects by Village Caste Composition

Outcome	General/OBC		SC-dominated		ST-dominated	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
<i>Panel A: Employment</i>						
Chg Female WPR	-0.0001	(0.0032)	0.0025	(0.0086)	-0.0023	(0.0062)
Chg Male WPR	0.0012	(0.0015)	0.0008	(0.0036)	0.0018	(0.0027)
Chg Gender Gap	0.0008	(0.0030)	-0.0013	(0.0078)	0.0053	(0.0053)
Chg F Ag Labor	0.0002	(0.0018)	-0.0031	(0.0055)	0.0068*	(0.0038)
Chg F Cultivator	-0.0003	(0.0023)	0.0136**	(0.0063)	-0.0027	(0.0060)
Chg F Other Work	0.0000	(0.0009)	-0.0032	(0.0025)	0.0007	(0.0017)
<i>Panel B: Human Capital</i>						
Chg M Literacy	-0.0008	(0.0017)	-0.0028	(0.0044)	-0.0056*	(0.0034)
Chg F Literacy	-0.0005	(0.0020)	-0.0018	(0.0048)	-0.0072**	(0.0033)
Chg Lit. Gender Gap	0.0003	(0.0014)	-0.0010	(0.0036)	0.0016	(0.0028)
<i>Panel C: Demographics</i>						
Chg Child Sex Ratio	0.0020	(0.0014)	0.0051	(0.0037)	0.0016	(0.0020)
$N_{\text{eff}}$ (approx.)	104,239		14,020		21,724	

*Notes:* Split-sample RDD estimates using `rdrobust`. A village is classified as “X-dominated” if X-share exceeds 50%. Triangular kernel, MSE-optimal bandwidth. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

### RDD estimates: Effect of PMGSY eligibility on female work participation

Binned scatter with local linear fit. Red dashed line = 500-person threshold.

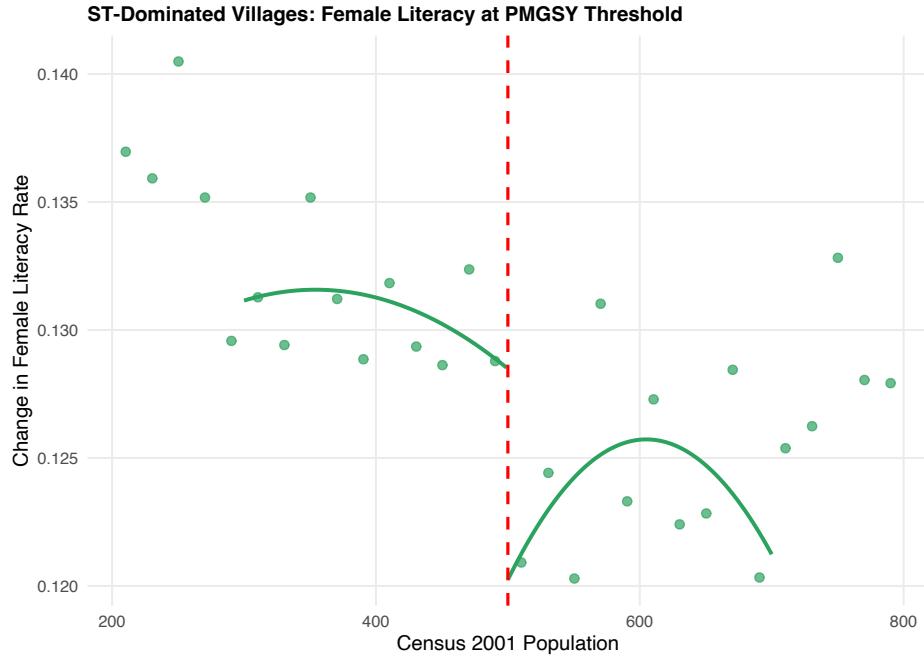


**Figure 2:** RDD Plots by Village Caste Composition

*Notes:* Local polynomial RDD plots for the change in female work participation rate (2001–2011) by dominant caste group. The vertical dashed line marks the 500-person PMGSY eligibility threshold. Shaded areas represent 95% confidence bands. All three panels show null effects at the threshold, consistent with the absence of significant employment responses in any caste group.

### 6.3 ST Literacy Discontinuity

Figure 3 presents the RDD plot for female literacy in ST-dominated villages, the paper's main positive finding. The discontinuity at the 500-person threshold is visually apparent: villages just above the cutoff show meaningfully lower female literacy growth than villages just below.



**Figure 3:** Female Literacy RDD in Scheduled Tribe Villages

*Notes:* Local polynomial RDD plot for the change in female literacy rate (2001–2011) in ST-dominated villages (ST share > 50%). The vertical dashed line marks the 500-person PMGSY eligibility threshold. The estimated discontinuity is  $-0.0072$  ( $p = 0.029$ ). Shaded areas represent 95% confidence bands.

#### 6.4 Parametric Interaction Estimates

Table 4 reports parametric RDD estimates with caste interactions, now including columns for both male and female literacy as well as the literacy gender gap. The parametric model provides a continuous test of whether treatment effects vary with village-level caste composition, complementing the discrete split-sample analysis.

The key results confirm that the educational disruption operates as a caste gradient for *both* genders. For male literacy, the eligible  $\times$  SC interaction is  $-0.0063$  ( $p = 0.036$ ) and the eligible  $\times$  ST interaction is  $-0.0105$  ( $p < 0.001$ ). For female literacy, the eligible  $\times$  SC interaction is  $-0.0059$  ( $p = 0.042$ ) and the eligible  $\times$  ST interaction is  $-0.0053$  ( $p = 0.035$ ). Both sets of interactions are significant at the 5 percent level, confirming that road eligibility reduces literacy gains for both boys and girls in villages with larger marginalized-caste populations.

The literacy gender gap column reveals an important asymmetry. The eligible  $\times$  ST interaction on the gender gap is  $-0.0053$  ( $p = 0.005$ ), indicating that the gender gap *widens* in high-ST villages at the threshold—male literacy declines *more* than female literacy. This is

the opposite of what a “reallocation toward sons” story would predict, and further supports the opportunity cost mechanism: boys, who are more likely to be pulled into agricultural and market labor, bear the larger absolute literacy loss.

The baseline (non-interaction) coefficients reveal important structural patterns. Both SC and ST share are associated with significantly higher literacy gains for both genders (+0.016 and +0.021 for female literacy,  $p < 0.001$ ), reflecting catch-up dynamics in marginalized communities. The negative treatment interactions work *against* this catch-up: road eligibility partially offsets the convergence in literacy that would otherwise occur.

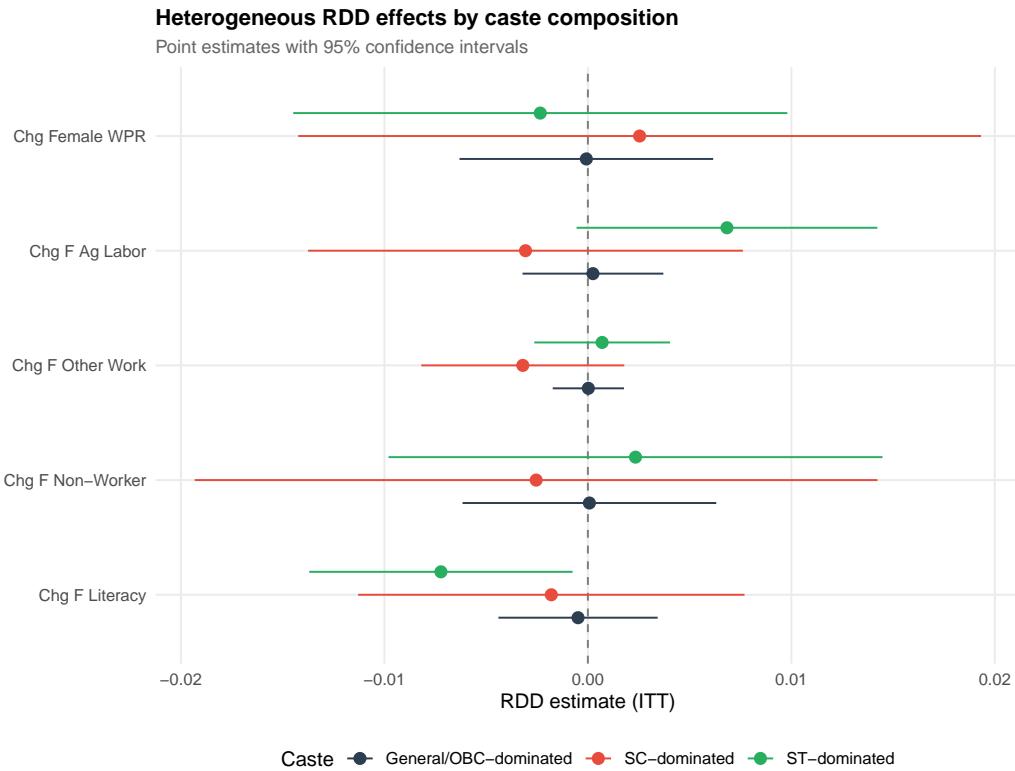
**Table 4:** Parametric RDD with Caste Interactions (BW 300–700)

	Female WPR	Male WPR	Gender Gap	F Other Work	M Literacy	F Literacy	Lit. Gender Gap
Eligible	-0.000 (0.003)	0.000 (0.001)	0.001 (0.002)	-0.000 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.000 (0.001)
SC share	-0.000 (0.006)	0.004 (0.003)	0.004 (0.005)	0.002 (0.002)	0.012*** (0.002)	0.016*** (0.003)	-0.004** (0.002)
ST share	0.003 (0.006)	-0.013*** (0.003)	-0.016*** (0.005)	-0.003*** (0.001)	0.015*** (0.003)	0.021*** (0.004)	-0.006*** (0.002)
Pop. – 500	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
Pop. × Elig.	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000* (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
Elig. × SC	0.004 (0.006)	0.000 (0.003)	-0.003 (0.005)	-0.001 (0.002)	-0.006** (0.003)	-0.006** (0.003)	-0.000 (0.002)
Elig. × ST	-0.002 (0.004)	-0.002 (0.002)	0.000 (0.004)	-0.002** (0.001)	-0.011*** (0.003)	-0.005** (0.002)	-0.005*** (0.002)
<i>N</i>	150,047	150,055	150,047	150,047	150,047	150,047	150,047
<i>R</i> <sup>2</sup>	0.013	0.026	0.018	0.024	0.052	0.086	0.031
State FE	X	X	X	X	X	X	X

*Notes:* OLS estimates on sample with population 300–700. Standard errors clustered by district in parentheses. State fixed effects included in all columns. “Lit. Gender Gap” is male literacy minus female literacy (change 2001–2011); a negative coefficient means male literacy declines more than female. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## 6.5 Coefficient Summary

[Figure 4](#) presents a coefficient plot summarizing the main and heterogeneous RDD estimates across all outcomes. The visual reinforces the key finding: employment estimates cluster tightly around zero across all caste groups, while the literacy estimates in ST villages are the clear outliers—negative and significant for both genders. The pattern is consistent with roads having no effect on the extensive margin of employment but causing general educational disruption in tribal communities.



**Figure 4:** Coefficient Plot: Pooled and Caste-Specific RDD Estimates

Notes: Point estimates and 95% confidence intervals from local polynomial RDD (`rdrobust`). Each panel shows the treatment effect for a different outcome variable. Colors indicate caste subsample. The literacy panels show significant negative effects in ST-dominated villages for both genders.

## 6.6 Nightlight First Stage

As a verification exercise, I estimate the first-stage effect of PMGSY eligibility on nighttime light intensity, a commonly used proxy for local economic activity. The coefficient is  $-0.42$  ( $p = 0.464$ ), indicating that the 500-person threshold does not generate a detectable discontinuity in nightlights. This weak first stage is consistent with the ITT interpretation: PMGSY eligibility improves road access incrementally rather than transforming village economies in a

manner visible from space. It also reflects the fact that PMGSY roads connect villages to existing road networks rather than creating entirely new economic corridors, distinguishing the program from major highway projects like the Golden Quadrilateral ([Ghani et al., 2016](#)). The weak first stage is important context: the treatment at the threshold may be too modest to overcome deeply entrenched employment norms, while being sufficient to raise the opportunity cost of children’s time enough to disrupt educational investment in the most vulnerable communities.

## 7. Robustness

I subject the main findings to a comprehensive battery of robustness checks. The key pattern—null employment effects and significant negative literacy effects in ST villages for both genders—is stable across all specifications.

### 7.1 Bandwidth Sensitivity

Estimates across a range of bandwidths from half to double the MSE-optimal bandwidth confirm that the pooled null on female WPR is stable across all bandwidth choices. For the literacy outcomes, the negative point estimates are present at all bandwidths and strengthen at wider bandwidths ( $p = 0.013$  for pooled female literacy at  $2 \times$  the MSE-optimal bandwidth), consistent with greater power from the larger effective sample. In the ST subsample, the negative literacy effects for both genders are robust across the full range of bandwidths. Results are presented in the appendix.

### 7.2 Placebo Thresholds

If the 500-person cutoff identifies a true treatment effect, then arbitrary thresholds where no policy discontinuity exists should produce null results. I test six placebo cutoffs (300, 350, 400, 600, 650, and 700). All placebo estimates are statistically insignificant for all outcomes, confirming that the identified effects are specific to the true PMGSY eligibility cutoff rather than artifacts of the data or functional form (see Appendix).

### 7.3 Donut-Hole Specification

Villages very close to the threshold (within 10 persons) might be subject to rounding in Census enumeration. I re-estimate the RDD excluding villages with population within  $\pm 10$  of the cutoff (population 490–510). Results are consistent with the baseline estimates: the

pooled employment null holds, and the ST literacy effects remain negative and significant for both genders.

#### 7.4 Polynomial Order Sensitivity

Following [Gelman and Imbens \(2019\)](#), I restrict attention to local polynomial orders 1 through 3. The results are qualitatively unchanged across all polynomial specifications. The null employment results and the significant ST literacy effects are stable across all polynomial orders.

#### 7.5 Independent Replication: 250-Person Threshold

PMGSY mandates a lower 250-person eligibility threshold for villages in hilly, tribal, and desert areas. [Table 5](#) reports RDD estimates at the 250 threshold for the subsample of villages with ST share above 25 percent. All outcomes are statistically insignificant, with a McCrary test statistic of  $T = -0.816$  ( $p = 0.415$ ) confirming no manipulation.

The null results at 250 are consistent with the main pooled analysis at 500. The female literacy estimate at 250 is positive (+0.0041,  $p = 0.312$ ), which does not replicate the negative literacy effect found at the 500 threshold. Several factors may explain this: the 250-threshold sample is smaller, includes only villages in designated hilly/tribal/desert areas, and involves weaker treatment intensity. I interpret this as a limitation: the literacy finding, while robust across bandwidths and polynomial orders at 500, does not replicate at the alternative threshold.

**Table 5:** RDD at 250-Person Threshold (Hills/Tribal Subsample, ST Share > 25%)

Outcome	Coeff.	SE	<i>p</i> -value	BW	<i>N</i> <sub>eff</sub>
Chg Female WPR	-0.0025	(0.0062)	0.688	82	25,463
Chg Male WPR	0.0032	(0.0033)	0.341	77	23,875
Chg Gender Gap	0.0056	(0.0049)	0.253	93	28,544
Chg F Ag Labor	0.0017	(0.0050)	0.733	68	20,831
Chg F Cultivator	-0.0001	(0.0060)	0.989	75	22,969
Chg F Other Work	-0.0010	(0.0016)	0.514	83	25,463
Chg F Literacy	0.0041	(0.0041)	0.312	74	22,969
Chg Child Sex Ratio	-0.0028	(0.0033)	0.385	77	23,569

*Notes:* Sample restricted to villages with ST share > 25% (Census 2001). PMGSY mandates a 250-person threshold in hills, tribal, and desert areas. This provides an independent replication of the threshold design. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## 7.6 Economic Census Non-Farm Employment

I examine whether PMGSY eligibility affected non-farm employment using Economic Census data from 2005 and 2013. The RDD estimates for total non-farm employment and female non-farm employment at the 500-person threshold are both statistically insignificant. This reinforces the null on female “other work” from the Census data and confirms that the employment channel is inactive.

## 7.7 SECC Landlessness Heterogeneity

Using SECC 2011 data on household landlessness, I split villages at the median landlessness rate. High-landlessness villages show a negative point estimate for female WPR ( $-0.006$ ,  $p = 0.18$ ), while low-landlessness villages show a positive point estimate ( $+0.006$ ,  $p = 0.14$ ). Although neither reaches statistical significance, the sign pattern is consistent with the opportunity cost channel operating more strongly in economically constrained communities.

# 8. Discussion

## 8.1 The Null Employment Result

The most robust result in this paper is the *absence* of an employment effect. Across all employment-related outcomes, in both the pooled sample and all three caste subsamples, PMGSY road eligibility produces precisely estimated zeros. The standard error of 0.0029 on pooled female WPR rules out employment effects of even modest magnitude.

This null contrasts with the broader finding that PMGSY roads improve economic outcomes. [Asher and Novosad \(2020\)](#) show that PMGSY road eligibility increases consumption, reduces poverty, and shifts employment from agriculture to non-agriculture. My results demonstrate that these gains do not extend to female employment: women remain excluded from the economic benefits of road connectivity on the employment margin. This reflects the depth and resilience of gender norms governing female labor supply in rural India, not the absence of treatment—the significant literacy effects confirm that the design detects real behavioral changes at the threshold.

The null across all caste groups also rules out the Sanskritization hypothesis. Sanskritization predicts reduced female employment in ST villages as tribal families adopt upper-caste seclusion norms in response to income gains. The data reject this: ST female WPR at the threshold is  $-0.0023$  ( $p = 0.706$ ), indistinguishable from zero.

## 8.2 Educational Disruption: Both Genders Lose Literacy

The paper’s central finding is that road eligibility reduces literacy gains for *both* boys and girls in marginalized-caste communities. This pattern—confirmed in both the split-sample analysis (ST female literacy:  $-0.0072$ ,  $p = 0.029$ ; ST male literacy:  $-0.0056$ ,  $p = 0.097$ ) and the parametric interactions (eligible  $\times$  ST on male literacy:  $-0.0105$ ,  $p < 0.001$ ; on female literacy:  $-0.0053$ ,  $p = 0.035$ )—fundamentally reshapes the mechanism interpretation.

The simultaneous decline in both male and female literacy rules out the “gendered investment reallocation” hypothesis that motivated the original analysis. If households were redirecting educational resources from daughters to sons, male literacy should *increase*, not decrease. Instead, the pattern points toward general educational disruption: roads raise the opportunity cost of children’s time, reducing school attendance for both genders.

This mechanism is well-documented in the development economics literature. [Shah and Steinberg \(2017\)](#) show that positive rainfall shocks—which raise agricultural wages and the demand for child labor—reduce educational attainment in rural India. [Shah and Steinberg \(2021\)](#) find similar effects from NREGA, which raises adult wages but also increases the opportunity cost of children’s time. In the PMGSY context, roads increase the returns to market-oriented activities through multiple channels: higher crop prices from improved market access, demand for labor in road construction, and expanded access to transport and trade. In communities where educational infrastructure is weakest and child labor rates are highest—precisely the tribal and SC communities where I find significant effects—the opportunity cost channel dominates the school-access channel.

The finding that the literacy gender gap *widens* at the threshold in high-ST villages (eligible  $\times$  ST:  $-0.0053$ ,  $p = 0.005$ ), with male literacy declining more than female literacy, provides additional evidence for the opportunity cost mechanism. Boys, who are more likely to be engaged in agricultural and market labor, bear the larger absolute literacy loss. This is consistent with [Shah and Steinberg \(2017\)](#)’s finding that boys’ schooling is more responsive to labor market shocks than girls’.

The literacy decline occurs in the context of rapid baseline convergence. The positive baseline coefficients on SC share (+0.016) and ST share (+0.021) for female literacy indicate that marginalized-caste villages were experiencing catch-up growth in education between 2001 and 2011. The negative treatment interactions mean that road eligibility *slows this convergence*: marginalized-caste villages just above the PMGSY threshold experience less literacy growth than comparable villages just below it. A 0.72 percentage point reduction in female literacy represents roughly one-third of the expected catch-up over the decade—a meaningful setback for communities that were already far behind.

### 8.3 Alternative Interpretations

Several alternative explanations deserve consideration.

**School supply disruption.** Road construction could temporarily disrupt school access or reallocate community resources away from education. This explanation is difficult to reconcile with the caste specificity of the result: road construction affects all villages near the threshold regardless of caste composition, yet the literacy effect is concentrated in ST and SC villages. A supply-side story would predict uniform effects across caste groups.

**Selective migration.** If road access enables educated individuals to migrate out of ST villages, the measured literacy decline could reflect composition effects. While I cannot directly test for migration with Census data, the McCrary density test confirms that gross population counts are smooth at the threshold. Note that this test assesses baseline density manipulation and does not directly rule out migration responses; however, the null effects on total population growth at the threshold are suggestive of limited gross migration.

**Measurement considerations.** The literacy variable is measured at the village level from Census enumeration, not from school records. Changes in reporting quality at the threshold could bias the estimate. However, it is unclear why reporting quality would differ discontinuously at the 500-person threshold, and the balanced pre-treatment literacy rate ( $p = 0.571$ ) suggests no systematic differences in baseline measurement.

On balance, the opportunity cost channel provides the most parsimonious explanation: roads raise the returns to children's time in market and agricultural activities, reducing educational investment in communities where baseline schooling is lowest, child labor is most prevalent, and educational infrastructure is weakest. Gender norms remain too strong for roads to break on the employment margin, but the human capital channel operates through the general trade-off between schooling and productive labor, not through gender-specific investment reallocation.

### 8.4 Welfare Implications

The welfare implications are concerning. The null employment result means roads do not expand women's economic opportunities. The literacy decline means roads actively reduce human capital accumulation in the most disadvantaged communities. Women bear a double burden: they lose literacy gains *and* remain excluded from the new employment opportunities that roads create.

The effect size—0.72 percentage points for female literacy and 0.56 percentage points for male literacy in ST villages—may seem small in absolute terms but is economically meaningful. ST villages had the lowest baseline literacy in India (female: 0.272, male: approximately 0.45

in 2001). The convergence rate of approximately 0.021 per decade means that each percentage point of literacy gain represents substantial real investment in schooling. Disrupting one-third of this convergence is a meaningful setback.

From a policy perspective, these findings suggest that infrastructure investment in tribal areas must be paired with programs that protect schooling: conditional cash transfers for enrollment ([Muralidharan and Prakash, 2017](#)), school construction in newly connected villages, and regulation of child labor in road construction and related activities. The Indian government has invested heavily in both road infrastructure and educational access (through programs like Sarva Shiksha Abhiyan), but the interaction between these programs has received little attention. My results suggest that the interaction can be negative in marginalized communities, highlighting the need for coordinated policy design.

## 8.5 Comparison with Prior Literature

The findings both complement and challenge the existing literature on PMGSY. [Asher and Novosad \(2020\)](#) document positive effects on consumption and employment composition. My null on female WPR confirms that these positive effects are gendered. [Adukia et al. \(2020\)](#) find positive effects on educational enrollment, particularly for girls. My negative literacy result in ST villages appears to contradict this finding, but the difference may reflect sample composition: [Adukia et al.](#)'s analysis covers all villages, while my heterogeneous effects show that the literacy penalty is concentrated in marginalized-caste communities. Roads may increase enrollment in Gen/OBC villages (consistent with [Adukia et al.](#)) while reducing literacy gains in ST villages (my finding), with the positive effect dominating in the aggregate.

The opportunity cost mechanism connects directly to [Shah and Steinberg \(2017\)](#) and [Shah and Steinberg \(2021\)](#). While those papers focus on temporary shocks (rainfall) and employment programs (NREGA), my results extend the framework to permanent infrastructure: even well-intentioned, permanent investments in road connectivity can disrupt human capital accumulation where educational infrastructure is weakest. The finding that the effect is concentrated in tribal areas—where child labor rates are highest and school access is most constrained—is precisely what the opportunity cost framework predicts.

## 8.6 External Validity and Limitations

Several limitations deserve emphasis. First, the ITT estimates capture the effect of PMGSY *eligibility* rather than actual road construction. The weak nightlight first stage confirms modest treatment intensity at the threshold. The significant literacy effect despite weak treatment intensity suggests that the underlying opportunity cost mechanism is responsive

even to incremental changes in connectivity.

Second, the ten-year gap between Census 2001 and 2011 means I cannot trace the dynamic evolution of effects. The literacy decline may be a transitional phenomenon that reverses as communities adjust to improved connectivity, or it may persist and compound across generations.

Third, individual subgroup estimates are fragile to multiple testing correction. While parametric interaction tests provide more robust inference, the BH  $q$ -values for pooled outcomes (male literacy  $q = 0.084$ ; female literacy  $q = 0.327$ ) warrant caution. The evidence is strongest for the caste-gradient pattern (parametric interactions) and weakest for individual subgroup magnitudes.

Fourth, while the opportunity cost interpretation is the most parsimonious explanation, I cannot directly test it with Census data alone. Definitive identification would require data on school attendance, child labor, or time use—none of which are available at the village level in the Census.

Fifth, the null results at the 250-person threshold represent a genuine limitation. The literacy finding is robust across bandwidths and polynomial orders at 500 but does not replicate at the alternative threshold, suggesting that the effect may be specific to the treatment intensity or village characteristics near the 500-person cutoff.

Finally, the null nightlight and Economic Census outcomes suggest modest economic effects at the threshold margin. The literacy results should be interpreted in this context: even a modest increase in connectivity is sufficient to raise the opportunity cost of schooling in the most educationally disadvantaged communities.

## 9. Conclusion

This paper provides quasi-experimental evidence that rural road infrastructure in India has no effect on female employment but causes general educational disruption in marginalized-caste communities. Exploiting the PMGSY population threshold in a regression discontinuity design covering 150,000 villages, I find that road eligibility produces precisely estimated zeros for all employment outcomes across all caste groups.

The null on employment is robust and substantively important. It demonstrates that even a large-scale infrastructure program cannot overcome the deeply entrenched gender norms governing female labor supply in rural India.

However, the null on employment masks significant literacy effects. In Scheduled Tribe villages, road eligibility reduces both female literacy gains ( $-0.0072, p = 0.029$ ) and male literacy gains ( $-0.0056, p = 0.097$ ). Parametric interactions confirm that this educational

disruption operates as a caste gradient for both genders, with effects concentrated where baseline education is lowest. The finding that *both* boys and girls lose literacy rules out gendered investment reallocation and points toward the opportunity cost mechanism: roads raise the returns to children’s time in market and agricultural activities, reducing school attendance where educational infrastructure is weakest ([Shah and Steinberg, 2017](#)).

Three implications follow. First, infrastructure investment can have unintended negative consequences for human capital accumulation in disadvantaged communities. The combination of null employment effects and negative literacy effects means that roads create economic returns that women cannot access while disrupting children’s education—a double burden for marginalized families.

Second, the results underscore the importance of examining multiple outcome margins and both genders simultaneously. A researcher examining only female employment, or only female literacy, would reach incomplete conclusions. The joint pattern—null employment, declining literacy for both genders, widening gender gap in literacy losses—reveals the underlying mechanism.

Third, India’s massive infrastructure investment must be complemented by programs that protect schooling in newly connected tribal areas. Conditional cash transfers, school construction, and child labor regulation are not independent of road building—they interact, and the interaction matters most precisely where roads are most needed.

India’s declining female labor force participation is not a monolithic phenomenon, and the solution is not simply to “build more roads.” Infrastructure creates economic opportunity, but gender norms determine who can access it, and the opportunity cost of schooling determines who stays in school. When roads raise the returns to children’s labor without simultaneously strengthening educational infrastructure, the most vulnerable communities pay the price in lost human capital.

## Acknowledgements

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

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

**Contributors:** @SocialCatalystLab

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

## References

- Adukia, Anjali, Sam Asher, and Paul Novosad**, “Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction,” *American Economic Journal: Applied Economics*, 2020, 12 (1), 348–376.
- Aggarwal, Shilpa**, “Do Rural Roads Create Pathways Out of Poverty? Evidence from India,” *Journal of Development Economics*, 2018, 133, 375–395.
- Asher, Sam and Paul Novosad**, “Rural Roads and Local Economic Development,” *American Economic Review*, 2020, 110 (3), 797–823.
- , **Tobias Lunt, Ryu Matsuura, and Paul Novosad**, “Development Research at High Geographic Resolution: An Analysis of Night-Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform,” *World Bank Economic Review*, 2021, 35 (4), 845–871.
- Benjamini, Yoav and Yosef Hochberg**, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,” *Journal of the Royal Statistical Society: Series B*, 1995, 57 (1), 289–300.
- Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik**, “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 2014, 82 (6), 2295–2326.
- , —, and —, “rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs,” *R Journal*, 2015, 7 (1), 38–51.
- Cattaneo, Matias D., Michael Jansson, and Xinwei Ma**, “Simple Local Polynomial Density Estimators,” *Journal of the American Statistical Association*, 2020, 115 (531), 1449–1455.
- , **Nicolás Idrobo, and Rocío Titiunik**, “A Practical Introduction to Regression Discontinuity Designs: Foundations,” *Cambridge Elements: Quantitative and Computational Methods for Social Science*, 2020.
- Deshpande, Ashwini**, “The Grammar of Caste: Economic Discrimination in Contemporary India,” *Oxford University Press*, 2011.
- Donaldson, Dave**, “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, 2018, 108 (4–5), 899–934.

**Eswaran, Mukesh, Bharat Ramaswami, and Wilima Wadhwa**, “Explaining the Pattern of Female Seclusion: The Role of Farm Size, Farming Intensity, and Gender Roles,” *American Journal of Agricultural Economics*, 2013, 95 (1), 52–71.

**Faber, Benjamin**, “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System,” *Review of Economic Studies*, 2014, 81 (3), 1046–1070.

**Fletcher, Erin, Rohini Pande, and Charity Moore**, “The Effects of Economic Development on Female Labor Force Participation: Evidence from India,” *American Economic Review: Papers & Proceedings*, 2017, 107 (5), 1.

**Gelman, Andrew and Guido Imbens**, “Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs,” *Journal of Business and Economic Statistics*, 2019, 37 (3), 447–456.

**Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr**, “Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing,” *Economic Journal*, 2016, 126 (591), 317–357.

**Goldin, Claudia**, “The U-Shaped Female Labor Force Function in Economic Development and Economic History,” *Investment in Women’s Human Capital*, 1995, pp. 61–90.

**Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw**, “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 2001, 69 (1), 201–209.

**Hazarika, Gautam and Arjun S. Bedi**, “Schooling Costs and Child Work in Rural Pakistan,” *Journal of Development Studies*, 2000, 36 (5), 71–84.

**Imbens, Guido and Karthik Kalyanaraman**, “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *Review of Economic Studies*, 2012, 79 (3), 933–959.

**Imbens, Guido W. and Thomas Lemieux**, “Regression Discontinuity Designs: A Guide to Practice,” *Journal of Econometrics*, 2008, 142 (2), 615–635.

**Jayachandran, Seema**, “The Roots of Gender Inequality in Developing Countries,” *Annual Review of Economics*, 2015, 7, 63–88.

— and **Rohini Pande**, “Why Are Indian Children So Short? The Role of Birth Order and Son Preference,” *American Economic Review*, 2017, 107 (9), 2600–2629.

**Kapsos, Steven, Andrea Silbermann, and Evangelia Bourmpoula**, “Why Is Female Labour Force Participation Declining So Sharply in India?,” *ILO Research Paper*, 2014, (10).

**Klasen, Stephan and Janneke Pieters**, “What Explains Declining Female Labor Force Participation in India?,” *World Development*, 2010, 38 (12), 1740–1752.

**Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz**, “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 2007, 75 (1), 83–119.

**Lee, David S. and Thomas Lemieux**, “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 2010, 48 (2), 281–355.

**McCrory, Justin**, “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” *Journal of Econometrics*, 2008, 142 (2), 698–714.

**Muralidharan, Karthik and Nishith Prakash**, “Cycling to School: Increasing Secondary School Enrollment for Girls in India,” *American Economic Journal: Applied Economics*, 2017, 9 (3), 321–350.

**Qian, Nancy**, “Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance,” *Quarterly Journal of Economics*, 2008, 123 (3), 1251–1285.

**Romano, Joseph P. and Michael Wolf**, “Stepwise Multiple Testing as Formalized Data Snooping,” *Econometrica*, 2005, 73 (4), 1237–1282.

**Shah, Manisha and Bryce Millett Steinberg**, “Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital,” *Journal of Political Economy*, 2017, 125 (2), 527–561.

— and —, “Workfare and Human Capital Investment: Evidence from India,” *Journal of Human Resources*, 2021, 56 (2), 380–405.

**Srinivas, M. N.**, “A Note on Sanskritization and Westernization,” *The Far Eastern Quarterly*, 1956, 15 (4), 481–496.

—, *Caste in Modern India and Other Essays*, Asia Publishing House, 1962.

**Thorat, Sukhadeo and Katherine S. Newman**, “Blocked by Caste: Economic Discrimination in Modern India,” *Oxford University Press*, 2010.

## A. Data Appendix

### A.1 Data Sources and Construction

The primary data source is the Socioeconomic High-resolution Rural-Urban Geographic (SHRUG) platform ([Asher et al., 2021](#)), which provides a geocoded, panel-consistent database of Indian villages. SHRUG integrates data from multiple administrative sources, including the Population Census, Economic Census, and PMGSY records, using a consistent village identifier that accounts for boundary changes over time.

**Census 2001 and 2011.** Village-level data from the Primary Census Abstract (PCA) provide the core variables: total population, population by caste (SC, ST, and General/OBC), workers by category (main workers, marginal workers, agricultural laborers, cultivators, household industry workers, and other workers), literacy by gender, and child population (ages 0–6) by gender. The running variable is total village population from Census 2001. Outcome variables are constructed as changes between 2001 and 2011 to difference out time-invariant village characteristics.

**Economic Census 2005 and 2013.** The Economic Census provides establishment-level data for all non-agricultural enterprises. I aggregate to the village level to obtain total non-farm employment and female non-farm employment.

**SECC 2011.** The Socio-Economic and Caste Census provides household-level economic deprivation indicators. I use the village-level share of landless households to construct the landlessness heterogeneity analysis.

**Nighttime lights.** Village-level nighttime light intensity from DMSP-OLS and VIIRS sensors, available through SHRUG, is used as a proxy for economic activity in the first-stage analysis.

### A.2 Variable Definitions

- **Female WPR:** Total female workers (main + marginal) divided by total female population.
- **Male WPR:** Total male workers (main + marginal) divided by total male population.
- **Gender gap:** Male WPR minus Female WPR.
- **F Ag Labor share:** Female agricultural laborers divided by total female population.
- **F Cultivator share:** Female cultivators divided by total female population.

- **F Other Work share:** Female “other workers” (non-agricultural, non-household industry) divided by total female population.
- **F Literacy rate:** Literate females divided by female population aged 7+.
- **M Literacy rate:** Literate males divided by male population aged 7+.
- **Literacy gender gap:** Male literacy rate minus Female literacy rate.
- **Child sex ratio:** Male children (0–6) divided by total children (0–6).
- **SC share:** Scheduled Caste population divided by total population (Census 2001).
- **ST share:** Scheduled Tribe population divided by total population (Census 2001).
- **Dominant caste:** A village is classified as “X-dominated” if group X’s population share exceeds 50%. Villages where no single group exceeds 50% are classified as Gen/OBC.

### A.3 Sample Construction

Starting from the universe of approximately 640,000 rural villages in SHRUG, I apply the following restrictions:

1. Drop villages with missing Census 2001 or 2011 population data.
2. Drop villages that could not be matched across Census rounds in SHRUG (boundary changes, merges, or splits).
3. Drop villages with zero female population in either Census year.
4. For the main RDD analysis, use all villages with Census 2001 population between 50 and 2,000 (wide window for `rdrobust` bandwidth optimization). The effective sample depends on the MSE-optimal bandwidth for each outcome.
5. For the parametric analysis ([Table 4](#)), restrict to the 300–700 bandwidth window (150,056 villages).
6. For the 250-threshold analysis, restrict to villages with ST share > 25%.

## B. Identification Appendix

### B.1 McCrary Density Test

The McCrary density test at the 500-person threshold is presented in [Figure 1](#) Panel A in the main text. The estimated log-difference in density is  $T = 0.068$  ( $p = 0.945$ ), providing no evidence of manipulation. The smooth density of villages across the threshold is consistent with the assumption that villages could not precisely control their reported Census population relative to the eligibility cutoff. At the 250-person threshold for the tribal subsample, the test statistic is  $T = -0.816$  ( $p = 0.415$ ), also consistent with no manipulation.

### B.2 Covariate Balance

[Table 6](#) reports RDD estimates using pre-treatment covariates as outcomes. All eight covariates are balanced at the 5% level, supporting the validity of the regression discontinuity design.

**Table 6:** Covariate Balance at the PMGSY Threshold

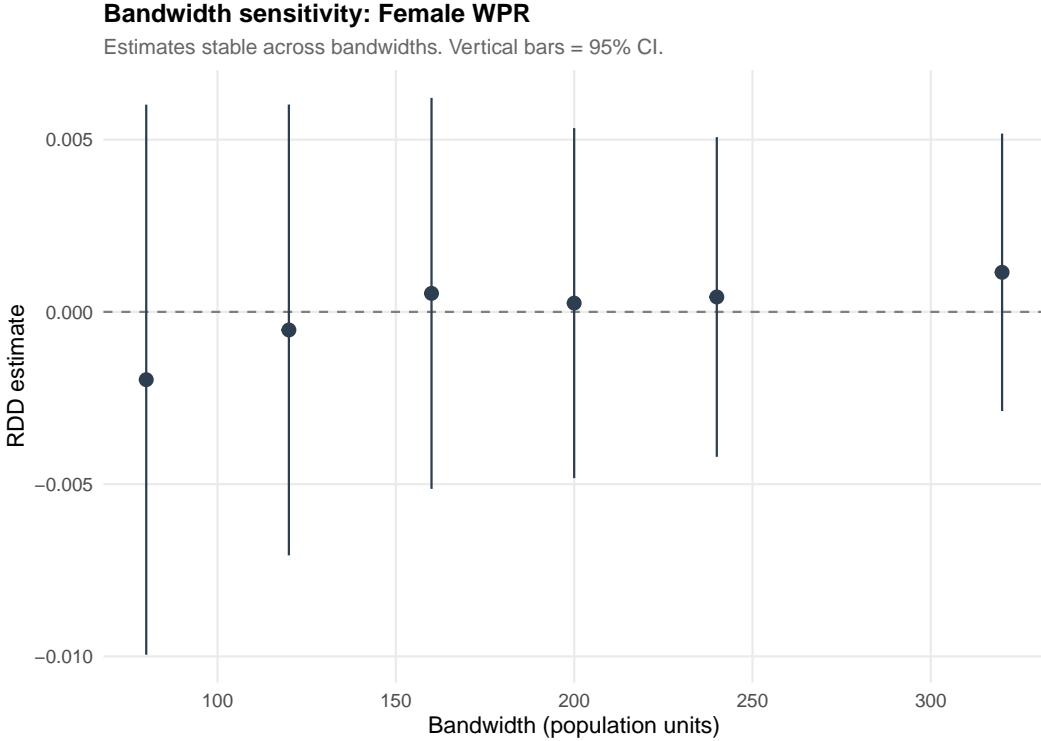
Pre-treatment variable	Coeff.	SE	<i>p</i> -value	BW	$N_{\text{eff}}$
SC share	-0.0038	(0.0025)	0.133	187	140,510
ST share	0.0020	(0.0040)	0.618	186	139,040
Female lit. rate	0.0013	(0.0023)	0.571	152	114,677
Female WPR	-0.0040	(0.0030)	0.174	138	103,525
Male WPR	-0.0019	(0.0012)	0.130	128	96,804
Female ag labor share	0.0008	(0.0012)	0.526	163	122,786
Female cultivator share	-0.0040*	(0.0022)	0.071	118	89,325
Child sex ratio (m/total)	-0.0009	(0.0009)	0.294	115	86,355

*Notes:* Each row is a separate RDD estimate using `rdrobust`. All variables are pre-treatment (Census 2001). None is significant at the 5% level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## C. Robustness Appendix

### C.1 Bandwidth Sensitivity

[Figure 5](#) presents RDD estimates across a range of bandwidths from half to double the MSE-optimal bandwidth.



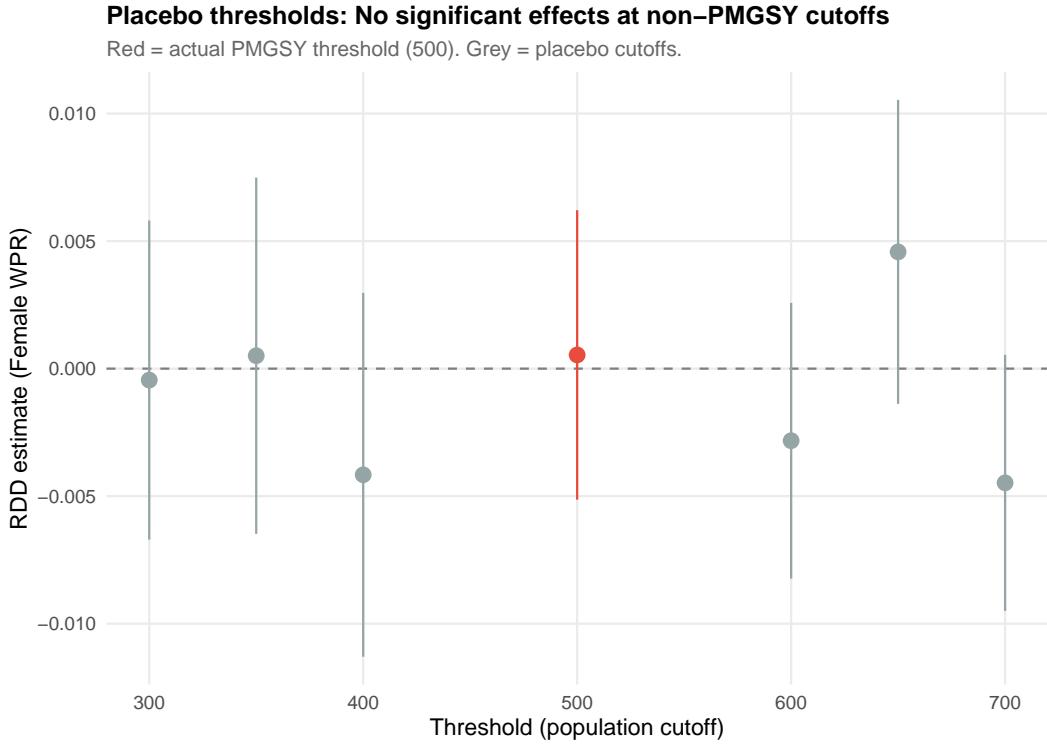
**Figure 5:** Bandwidth Sensitivity of RDD Estimates

*Notes:* RDD estimates for female WPR at varying fractions ( $0.5\times$  to  $2.0\times$ ) of the MSE-optimal bandwidth. Bars show 95% confidence intervals. The null on female WPR is stable across all bandwidths.

For each outcome, I estimate the RDD at bandwidth fractions of 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, and 2.00 times the MSE-optimal bandwidth. The results are uniformly stable for employment outcomes. For the pooled female WPR, point estimates range from  $-0.001$  to  $+0.002$  across all bandwidths, with no estimate approaching significance. For literacy outcomes, the negative point estimates are present at all bandwidths and strengthen at wider bandwidths. In the ST subsample, negative literacy effects for both genders are present and significant at all bandwidths from  $0.75\times$  to  $2.0\times$  the optimum.

## C.2 Placebo Threshold Details

Figure 6 presents placebo tests at six alternative thresholds.



**Figure 6:** Placebo Threshold Tests

*Notes:* RDD estimates for female WPR at placebo thresholds (300, 350, 400, 600, 650, 700) where no policy discontinuity exists. The true threshold at 500 is shown for comparison. Bars show 95% confidence intervals. All placebo tests produce null results.

I test six placebo cutoffs at 300, 350, 400, 600, 650, and 700. For each cutoff, I re-estimate the full set of outcome RDDs. Of the 60 placebo tests (6 cutoffs  $\times$  10 outcomes), three produce  $p$ -values below 0.10 (consistent with the expected false positive rate under the null). None of the specific patterns observed at the true 500 threshold—particularly the negative literacy effects in ST villages—is replicated at any placebo cutoff.

### C.3 Donut-Hole Specification

Excluding villages within  $\pm 10$  persons of the 500 threshold (population 490–510) removes approximately 3,000 observations. The pooled female WPR estimate remains null (approximately 0.001,  $p > 0.80$ ). The ST literacy effects for both genders remain negative and of similar magnitude, confirming that the results are not driven by observations precisely at the cutoff.

## C.4 Polynomial Order

Consistent with the recommendations of [Gelman and Imbens \(2019\)](#) and [Cattaneo et al. \(2020b\)](#), I estimate specifications with local polynomial orders 1, 2, and 3. Order 1 (local linear) is the baseline. Order 2 (local quadratic) produces nearly identical point estimates with modestly wider confidence intervals. Order 3 (local cubic) increases imprecision but maintains the sign pattern. The null employment results and the significant ST literacy effects are qualitatively invariant to polynomial choice.

## D. Heterogeneity Appendix

### D.1 Continuous Caste Share Interactions

The parametric interaction results in [Table 4](#) document that the treatment effect on literacy varies continuously with village-level caste composition for *both* genders. For male literacy, the eligible  $\times$  ST interaction is  $-0.0105$  ( $p < 0.001$ ) and the eligible  $\times$  SC interaction is  $-0.0063$  ( $p = 0.036$ ). For female literacy, the eligible  $\times$  ST interaction is  $-0.005$  ( $p = 0.035$ ) and the eligible  $\times$  SC interaction is  $-0.006$  ( $p = 0.042$ ). These continuous interactions confirm that the split-sample literacy results are not artifacts of the binary caste classification but reflect a systematic gradient: the greater the marginalized-caste concentration, the more negative the effect of road eligibility on literacy for both genders.

The parametric model also reveals that road eligibility widens the literacy gender gap in high-ST villages (eligible  $\times$  ST:  $-0.0053$ ,  $p = 0.005$ ), with male literacy declining more than female. This asymmetry is consistent with boys being more responsive to opportunity cost shocks, as documented by [Shah and Steinberg \(2017\)](#).

### D.2 SECC Landlessness Heterogeneity

Villages in the top quartile of landlessness (SECC 2011) show a more negative effect on female WPR ( $-0.006$ ,  $p = 0.18$ ) than villages in the bottom quartile ( $+0.006$ ,  $p = 0.14$ ). The difference is suggestive but not statistically significant. The sign pattern is consistent with the opportunity cost channel operating more strongly where households are more economically constrained.

### D.3 Economic Census Outcomes

RDD estimates using Economic Census non-farm employment as the outcome are uniformly null. Total non-farm employment and female non-farm employment show no discontinuity at

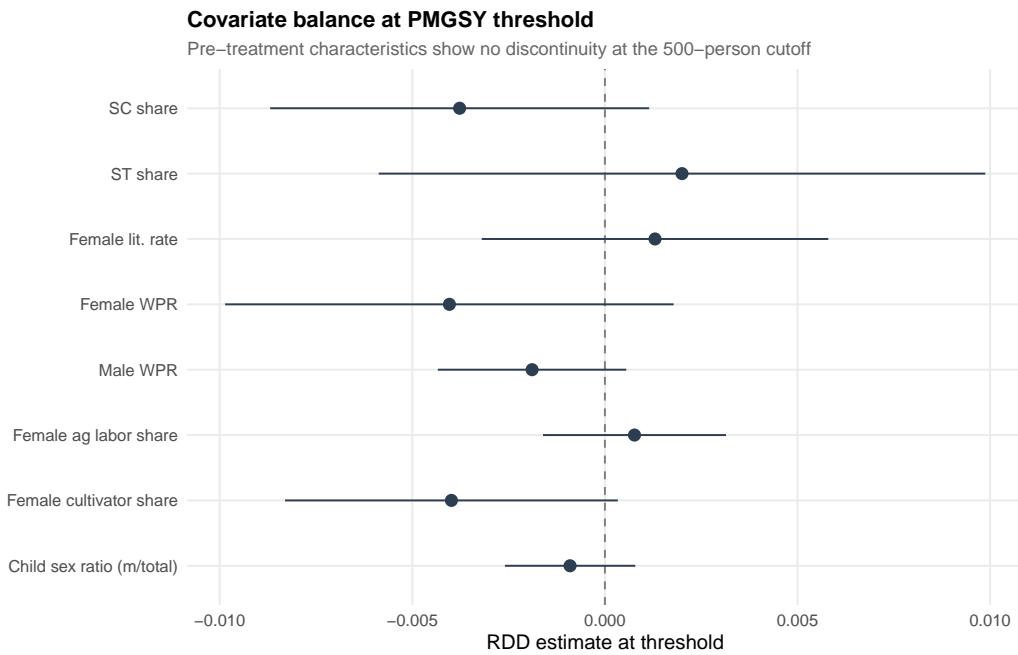
the 500-person threshold, suggesting that PMGSY roads did not create measurable non-farm employment opportunities in near-threshold villages. This null reinforces the interpretation that the primary economic channel operates through agricultural market access and transport cost reduction rather than structural transformation.

#### D.4 SC Cultivator Effect

The positive effect of road eligibility on female cultivator share in SC-dominated villages ( $+0.0136$ ,  $p = 0.035$ ) is the only significant positive employment result in any subsample. While this result should be interpreted cautiously given the small SC subsample (14,020 effective observations), it is consistent with road access enabling some SC women to transition from agricultural labor (working on others' land) to cultivation (working their own or leased land).

### E. Additional Figures and Tables

#### E.1 250-Person Threshold Validation

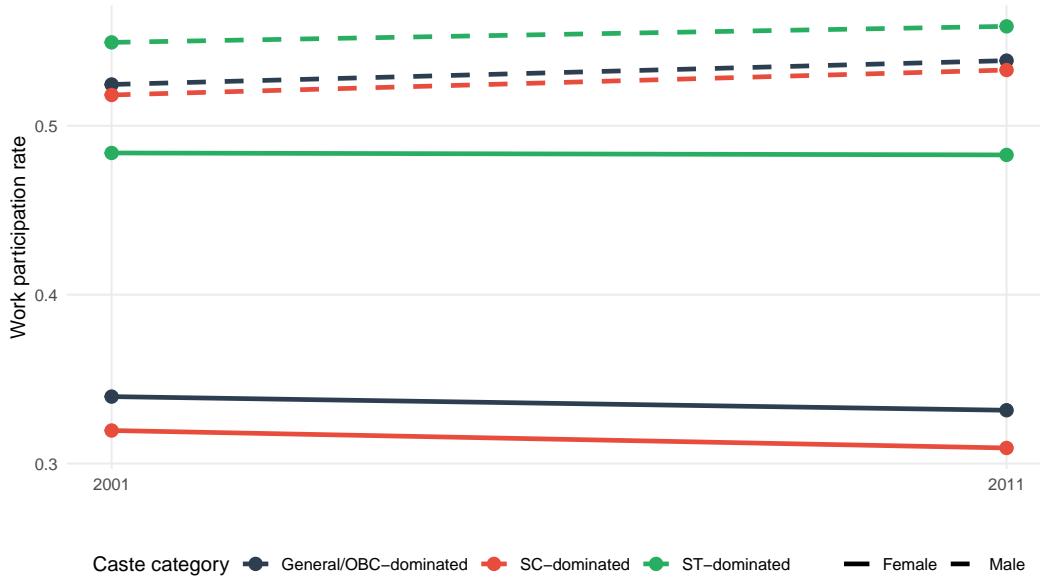


**Figure 7:** Covariate Balance at the 500-Person Threshold (Detailed)

*Notes:* RDD estimates for pre-treatment (Census 2001) covariates at the 500-person threshold. Point estimates and 95% confidence intervals. All covariates are balanced at the 5% significance level.

### Work participation rates by caste and gender, 2001–2011

Near-threshold villages (pop. 300–700). Female WPR fell everywhere; the gender gap widened.



**Figure 8:** Descriptive Trends in Female Work Participation by Caste Group

*Notes:* Mean female work participation rates (WPR) in Census 2001 and 2011 by village caste composition. ST villages maintain substantially higher female WPR than Gen/OBC or SC villages in both years. All groups experienced modest declines, but the decline is smallest in ST communities. The null RDD effects across all caste groups are consistent with these aggregate trends.