Spatial Inequality in Access to Elite Public Schools: Evidence from India's Jawahar Navodaya Vidyalaya System

Amit Kr Eshore[1]

[1] *Indira Gandhi Institute of Development Research, Mumbai*

April 2026

Abstract

India's Jawahar Navodaya Vidyalaya system selects students for centrally funded residential schools through a district-level entrance examination whose cut-off scores are determined locally. Using a newly constructed district–year panel of JNVST cut-offs for 2025 and 2026, merged with UDISE^{+} school infrastructure data, we examine what drives variation in admission thresholds across districts and whether that variation is spatially organised.

Two patterns stand out. After conditioning on state and year effects, demographic composition matters far more than observable infrastructure: districts with higher SC/ST student shares face lower cut-offs and wider urban–rural OBC gaps, while infrastructure variables are weak and inconsistent predictors. Admission thresholds also exhibit strong spatial clustering (Moran's $I = 0.573$), and a spatial lag model reveals substantial dependence across neighbouring districts—local conditions are amplified, not contained.

The two dimensions of inequality diverge sharply. Gender gaps are small and geographically unstructured. Urban–rural OBC gaps, by contrast, are large, spatially clustered, and closely tied to demographic composition. These contrasting patterns trace back to a common mechanism: admission thresholds mirror geographically uneven applicant preparedness, so a uniform reservation rule produces systematically unequal access depending on where a child happens to live.

Keywords: educational inequality, spatial autocorrelation, India, caste, gender, Navodaya Vidyalaya

1 Introduction

Educational opportunity often depends less on institutional rules and more on the conditions students arrive with. We study this geography of opportunity in India's Jawahar Navodaya Vidyalaya (JNV) system, a network of centrally funded residential schools. Admission relies on a nationally uniform entrance exam (JNVST), but selection is based on district-specific cut-offs. Consequently, the score required for entry varies widely across the country, reflecting local disparities in applicant pools rather than deliberate policy design. Equal rules, applied to unequal geographies, yield highly unequal access.

We ask two main questions. First, are these varying thresholds driven more by observable school infrastructure (supply-side) or by the demographic composition of the applicant pool (demand-side)? Second, do these admission cut-offs cluster geographically? If disparities are spatially dependent and rooted in demographics, simple infrastructure investments or uniform national policies will struggle to close the gap.

To answer these questions, we construct a novel 2025–2026 district-level panel of JNVST cut-offs merged with UDISE$^+$ infrastructure data and spatial shapefiles. We employ two-way fixed effects and maximum likelihood spatial regressions.

Several patterns emerge consistently. Demographic composition substantially outweighs physical infrastructure in explaining cut-off variation: districts with higher SC/ST populations see lower thresholds and wider urban–rural gaps. Cut-offs also exhibit strong spatial dependence—educational disadvantage is regionally clustered rather than randomly scattered. Gender gaps are small and spatially unstructured; urban–rural gaps are large, geographically concentrated, and closely tied to demographics. Uniform reservation rules, applied to an uneven applicant pool, cannot offset these structural disparities.

Contributions and Related Literature.

The paper connects three strands of literature—education economics, affirmative action, and spatial econometrics—through the following contributions:

First, we document strong geographic clustering in JNV admission thresholds. This extends the affirmative action literature (Pande, 2003; Bertrand et al., 2010; Bagde et al., 2016) by showing that spatial heterogeneity fundamentally alters how uniform quotas operate across locations.

Second, we find that applicant demographics outperform observable school inputs in explaining equity gaps. This shifts the focus of the school inputs literature (Duflo et al., 2011; Muralidharan and Sundararaman, 2011; Andrabi et al., 2017) from supply-side constraints toward demand-side preparedness.

Third, applying spatial econometrics (Anselin, 1988; Rey and Montouri, 1999; Case et al., 1993; Gibbons and Overman, 2012) to educational access, we show that local disparities are amplified through a spatial multiplier. Spatial models substantially outperform non-spatial specifications, indicating that district-level gaps spill over and compound across neighbouring jurisdictions.

Roadmap.

Section 2 describes the JNV system. Section 3 presents the conceptual framework, and Section 4 the data. Sections 5–7 cover the empirical strategy, main results, and spatial analysis. Section 8 assesses robustness, and Sections 9–10 discuss policy implications and conclude.

2 Institutional Background

2.1 The Navodaya Vidyalaya System

The Jawahar Navodaya Vidyalaya system was set up under the 1986 National Policy on Education to offer high-quality residential schooling to talented rural children regardless of socioeconomic background. The network now covers over 650 schools across nearly every district in India (Classes VI–XII), fully funded by the central government and administered by the Navodaya Vidyalaya Samiti (NVS) under the Ministry of Education.

2.2 The JNVST and District-Level Cut-offs

Entry to Class VI is decided entirely by the JNVST, a 100-mark multiple-choice test covering mental ability, arithmetic, and language, held each December or January. The key institutional feature is that each JNV admits students only from its own district, so the cut-off in district d is set by the marginal applicant within that district's quota pool. This generates the cross-district variation in cut-offs we exploit. Block-level seat caps additionally ensure *within-district geographic representation*.

2.3 Reservation Structure and Expected Quota Hierarchy

The reservation rules create a nested quota system: at least 75 percent of seats must come from rural areas, at least one-third are reserved for girls, and SC/ST/OBC seats are allocated in proportion to district population shares. Within each gender, seats span four categories—Open-Unreserved (Open-UR), Rural-OBC, Rural-SC, and Rural-ST—each with its own NVS-published cut-off.

This design implies a theoretical ordering:

$$C_{\text{Open-UR}} \;\geq\; C_{\text{Rural-OBC}} \;\geq\; C_{\text{Rural-SC}} \;\geq\; C_{\text{Rural-ST}}, \tag{1}$$

where C denotes the cut-off within a quota cell. Any violation of this ordering is a *quota hierarchy inversion*, typically arising from thin applicant pools or local demand imbalances.

3 Conceptual Framework

3.1 A Spatial Equilibrium Model of Cut-off Determination

Let C_{dt} denote the cut-off score in district d at time t. We model C_{dt} as arising from a spatial equilibrium in applicant quality, where the preparedness of the local applicant pool depends on both district-specific factors and conditions in neighbouring districts. This interdependence can reflect the geographic diffusion of coaching resources, information spillovers about examination strategies, awareness of the benefits the JNV network offers, and the movement of students across nearby administrative boundaries.

We formalise this as a spatial autoregressive (SAR) model:

$$C_{dt} = \rho \sum_{j \neq d} w_{dj}\, C_{jt} + \mathbf{X}'_{dt}\boldsymbol{\beta} + \alpha_s + \gamma_t + \varepsilon_{dt}, \tag{2}$$

where w_{dj} are elements of a spatial weights matrix **W** capturing geographic contiguity between districts, and $\sum_j w_{dj} C_{jt}$ denotes the spatial lag of the dependent variable. The vector $\mathbf{X}_{dt}$ includes district-level controls, α_s are state fixed effects, and γ_t are year fixed effects. The error term ε_{dt} captures unobserved district-level shocks.

The parameter ρ measures the degree of spatial dependence: it captures how admission thresholds in a given district co-move with those in neighbouring districts, reflecting equilibrium interactions in applicant quality across space.[1]

[1] Under standard assumptions, the SAR model implies a reduced-form representation $C = (I - \rho W)^{-1}(X\beta + \alpha + \gamma + \varepsilon)$, so that local shocks propagate across districts through a spatial multiplier $(1 - \rho)^{-1}$.

3.2 *Testable Predictions*

Prediction 1 (Spatial dependence). If applicant quality is shaped by geographically correlated preparation environments, cut-off scores should exhibit positive spatial dependence, net of observable covariates.

Prediction 2 (Demographic dominance). If SC/ST share proxies for persistent structural differences in applicant preparedness not captured by school infrastructure, it should be a stronger and more robust correlate of both cut-off levels and equity gaps than infrastructure variables, conditional on fixed effects.

3.3 *Equity Gap Outcomes*

The *gender gap* is defined as:

$$G_{dt}^{\text{gender}} = C_{dt}^{\text{Boy-UR}} - C_{dt}^{\text{Girl-UR}}.$$

A positive value indicates boys face a higher cut-off, reflecting either higher male applicant preparedness or greater competitiveness in the male pool given institutional design.

The *urban–rural OBC gap* is:

$$G_{dt}^{\text{UR-OBC}} = C_{dt}^{\text{Open-UR}} - C_{dt}^{\text{Rural-OBC}}.$$

Under the expected hierarchy (1), this gap should be non-negative. Large positive values indicate substantial separation between unrestricted and reserved pools; values near zero signal hierarchy compression.

4 Data

4.1 *Sources and Construction*

The primary outcomes are district-level JNVST cut-off scores for 2025 and 2026, drawn from official NVS notifications and published separately by gender and reservation category (Open-UR, Rural-OBC, Rural-SC, Rural-ST, disability quota). Removing island territories and districts with incomplete records yields 978 district–year observations across 27 states and union territories, with 530 districts in both years (the balanced panel).

District-level controls come from UDISE$^+$: share of schools with functional electricity, share of teaching staff with recognised training, pupil–teacher ratio, and SC/ST student share.

The last of these is a demographic proxy for applicant pool composition rather than an infrastructure measure.

Merging the two sources required manual district-name harmonisation via a fuzzy-match crosswalk, hand-correction of boundary and naming discrepancies, and a deterministic fix for DBF encoding errors in several Himachal Pradesh districts. Districts with multiple JNV campuses had cut-offs averaged and infrastructure rates averaged across campuses. The final analytical sample covers over 560 districts. District boundary shapefiles (Survey of India, harmonised to 2011 Census delimitation) underpin the spatial weights matrix **W**, constructed using Queen contiguity and row-standardised so that each district's spatial lag equals the average cut-off of its neighbours. Island and non-contiguous territories are excluded from **W** but retained in non-spatial regressions.

4.2 Descriptive Statistics

Table 1 reports summary statistics. The mean Open-UR cut-off for boys is 93.5 (SD $= 4.3$ pp; range 65–100); the gap between the 5th and 95th percentile districts is large relative to the 100-point scale, signalling substantial cross-district heterogeneity.

The two dimensions of inequality look very different from one another. Gender gaps are tightly centred around zero (mean $= 0.27$ pp; median $= 0$), pointing to near parity between boys and girls. Urban–rural OBC gaps are a different matter: large and right-skewed (mean $= 10.5$ pp; SD $= 8.5$ pp; max $= 46$ pp), with SC gaps wider still (mean $= 13.2$ pp). The nominal 3 percent hierarchy inversion rate is an artefact of cross-year averaging; within each cross-section and state there are no inversions. Infrastructure variables are high on average with moderate spread, while SC/ST student share varies considerably (SD $= 19.4$ pp)—a pattern that anticipates its central role in what follows.

5 Empirical Strategy

5.1 Panel Fixed-Effects Regressions

For each outcome $Y_{dt} \in \{C_{dt}^{\text{Open-UR}},\ G_{dt}^{\text{gender}},\ G_{dt}^{\text{UR-OBC}}\}$, we estimate a two-way fixed-effects specification:

$$Y_{dt} = \mathbf{X}'_{dt}\boldsymbol{\beta} + \alpha_s + \gamma_t + u_{dt}, \tag{3}$$

Table 1: Descriptive Statistics

Variable	N	Mean	SD	Min	p25	Median	p75	Max
Cut-off: Boys, Open-UR (%)	978	93.54	4.26	65.00	91.00	94.00	97.00	100.00
Cut-off: Girls, Open-UR (%)	978	93.27	4.41	60.00	90.00	94.00	97.00	100.00
Gender Gap (Boys−Girls, pp)	978	0.27	2.27	−13.00	−1.00	0.00	2.00	15.00
Urban−Rural OBC Gap (pp)	978	10.47	8.49	−7.00	5.00	9.00	14.00	46.00
Urban−Rural SC Gap (pp)	978	13.17	9.52	−15.00	6.00	12.00	19.00	52.00
Hierarchy Inversion (0/1)	978	0.03	0.16	0.00	0.00	0.00	0.00	1.00
Schools with Electricity (%)	978	87.95	17.52	7.29	80.82	93.29	99.10	100.00
Trained Teachers (%)	978	82.32	15.48	17.87	73.42	85.51	94.45	100.00
Pupil–Teacher Ratio	978	22.71	8.23	3.75	17.02	22.06	27.25	73.68
SC/ST Student Share (%)	978	27.68	19.37	0.00	12.30	23.91	40.10	100.00

Notes: District–year panel, 2025–2026. Cut-off scores expressed as percentage marks out of 100. Island territories and non-contiguous union territories excluded from spatial analysis but included here. pp = percentage points.

Source: Author's calculations using JNVST cut-off data, UDISE^{+} infrastructure data, and district boundary shapefiles.

where α_s are state fixed effects and γ_t are year fixed effects. Standard errors are clustered at the district level to account for serial correlation within districts over time.[2]

The covariate vector $\mathbf{X}_{dt}$ includes SC/ST student share, pupil–teacher ratio, share of trained teachers, and electricity access. State and year fixed effects absorb time-invariant heterogeneity and common shocks; identification comes from within-state variation across districts over time.

All estimates are *descriptive conditional correlations*, not causal effects. Conditional on the fixed effects, the identifying assumption is that included covariates are orthogonal to unobserved drivers of cut-offs.[3]

5.2 Spatial Diagnostics

To assess geographic clustering in admission thresholds, we compute Global Moran's I, defined as:

$$I = \frac{n}{\sum_d \sum_j w_{dj}} \cdot \frac{\sum_d \sum_j w_{dj}(C_d - \bar{C})(C_j - \bar{C})}{\sum_d (C_d - \bar{C})^2}, \tag{4}$$

where w_{dj} are elements of the spatial weights matrix **W**. Positive values of I indicate geographic clustering of similar cut-offs; values near zero are consistent with spatial randomness.[4]

[2] Clustering at the district level allows for arbitrary within-district correlation in the error term across years, which is important given the short panel structure.

[3] This assumption may be violated in the presence of unobserved, time-varying district-level factors (e.g., changes in local schooling quality or coaching intensity) that are correlated with both covariates and outcomes. Results should therefore be interpreted as reduced-form associations rather than structural effects.

[4] Inference is based on permutation tests using 999 randomisations of the data, yielding pseudo-p-values under the null of spatial independence.

The Moran scatter plot visualises spatial dependence by plotting the standardised outcome against its spatial lag; the slope of the fitted line equals Moran's I by construction.

To examine local patterns of clustering, we compute Local Indicators of Spatial Association (LISA) (Anselin, 1995), defined as:

$$I_d = z_d \sum_j w_{dj} z_j, \tag{5}$$

where z_d denotes the standardised value of the outcome in district d. LISA statistics identify the contribution of each district to global spatial dependence and allow classification into High–High (HH) and Low–Low (LL) clusters, as well as High–Low (HL) and Low–High (LH) spatial outliers.[5]

5.3 *Spatial Model Selection and Estimation*

We select the spatial specification using the Anselin LM diagnostic framework on OLS residuals. When both LM-Lag and LM-Error tests reject spatial independence, model choice follows the robust statistics, preferring the specification with the smaller p-value (Anselin, 1988).[6]

Guided by these diagnostics, we estimate a spatial autoregressive (SAR) model of the form:

$$\mathbf{C} = \rho\,\mathbf{W}\mathbf{C} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{6}$$

where $\mathbf{W}$ is a row-standardised spatial weights matrix.[7]

The SAR model captures outcome dependence across neighbouring districts, estimated by maximum likelihood.

Following LeSage and Pace (2009), total effects are decomposed into direct (within-district) and indirect (spillover) components.[8]

[5] Cluster significance is assessed at the 5% level using 999 permutations. HH (LL) clusters indicate districts surrounded by similarly high (low) values, while HL and LH denote spatial outliers.

[6] The decision rule distinguishes between spatial dependence in the dependent variable (lag dependence) and in the error term (error dependence), helping to identify the appropriate structural form.

[7] Row-standardisation ensures that spatial lags are comparable across districts and facilitates interpretation of the spatial autoregressive parameter.

[8] In the SAR model, coefficients cannot be interpreted directly as marginal effects due to feedback through the spatial multiplier $(I - \rho W)^{-1}$. Reported effects are therefore computed using average partial derivatives.

6 Results

6.1 Gender Gaps in Cut-off Scores

Figure 1 plots state fixed effects from a two-way FE regression of the gender gap on state and year indicators, with 95% confidence intervals clustered at the district level, for 30 states and union territories. The omitted category (shown at zero) is Andaman & Nicobar by alphabetical ordering.

The picture is clear. Almost all state fixed effects are statistically indistinguishable from parity: the confidence intervals overlap substantially, and there is no systematic tendency for boys or girls to face higher cut-offs at the state level. Lakshadweep is the notable exception—a pronounced negative outlier at around -5 pp—likely reflecting the island territory's thin applicant pool rather than any structural feature of gender competition. Telangana sits at the other end at approximately $+1$ pp. Everywhere else, the estimates cluster closely around zero.

The FE-based approach is important here: stripping out state-level fixed factors and the common time trend before plotting reveals that apparent geographic patterns in gender gaps largely dissolve. What remains is remarkably flat.

6.2 Correlates of Cut-off Levels and Equity Gaps

Table 2 reports estimates of equation (3) for three outcomes: the Open-UR cut-off level (Column 1), the gender gap (Column 2), and the urban–rural OBC gap (Column 3). All specifications include state and year fixed effects, with standard errors clustered at the district level.

The dominant story is demographic composition, not infrastructure. SC/ST student share is the most robust and economically significant correlate across all three columns. A 10 percentage point increase in SC/ST share is associated with a 0.58 pp reduction in the Open-UR cut-off (Column 1, $p < 0.001$) and a 1.48 pp increase in the urban–rural OBC gap (Column 3, $p < 0.001$). Its association with the gender gap, however, is small and statistically insignificant (Column 2).

The asymmetry suggests that SC/ST share is picking up structural differences in applicant pool composition that affect competition across quota categories, but not competition between boys and girls. Where the pool is disproportionately SC/ST, the unreserved category becomes harder to enter relative to reserved categories—widening the OBC gap—but there is no analogous mechanism operating along gender lines.

Infrastructure variables, by contrast, add little. The pupil–teacher ratio enters positively for

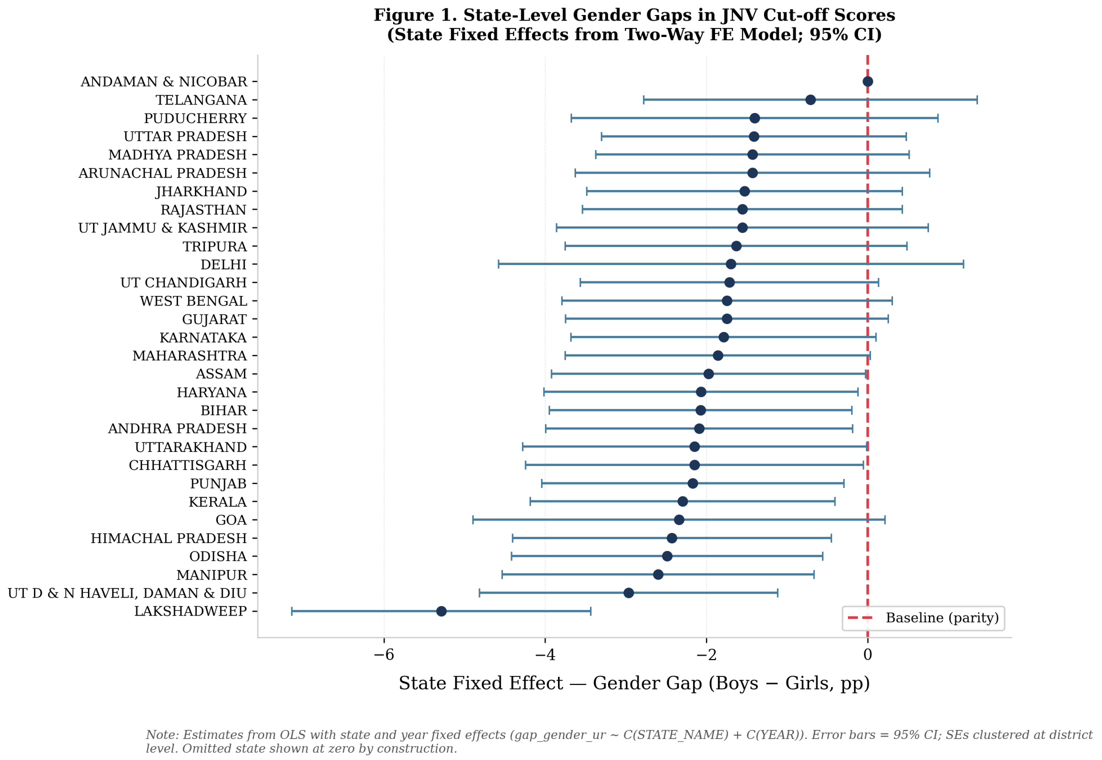


Figure 1: State Fixed Effects for Gender Gap in JNV Cut-off Scores (Boys − Girls, Open-UR). Estimates from OLS with state and year fixed effects; SEs clustered at district level; 95% CIs shown. Positive values = boys face a higher threshold relative to the omitted baseline. Omitted state (Andaman & Nicobar) shown at zero by construction. $N = 978$ observations, 30 states/UTs.

Source: Author's calculations using JNVST cut-off data, UDISE+ infrastructure data, and district boundary shapefiles.

Table 2: Correlates of JNV Selectivity and Equity Gaps

	Cut-off level (1)	Gender gap (2)	Urban–rural gap (3)
SC/ST student share (%)	-0.058^{***}	-0.007	0.148^{***}
	(0.010)	(0.005)	(0.017)
Pupil–teacher ratio	0.085^{**}	0.000	-0.127^{**}
	(0.033)	(0.017)	(0.053)
Trained teachers (%)	0.031	0.010	0.064^{*}
	(0.024)	(0.013)	(0.036)
Schools with electricity (%)	-0.016	-0.028^{**}	-0.014
	(0.023)	(0.014)	(0.033)
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Clustered SE	District	District	District
Observations	978	978	978
R^2	0.677	0.147	0.621
Adjusted R^2	0.667	0.120	0.609
Residual SE	2.843	2.266	4.179
F statistic	1405.541^{***}	219.706^{***}	40.831^{***}

Notes: OLS with state and year fixed effects; standard errors clustered by district in parentheses. Estimates are descriptive conditional correlations. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Source: Author's calculations using JNVST cut-off data, UDISE+ infrastructure data, and district boundary shapefiles.

cut-off levels and negatively for the OBC gap, but these associations likely reflect a mix of crowding and underlying demand conditions and are difficult to interpret causally. Electricity access has a small negative coefficient on the gender gap that does not survive sample restrictions (Section 8). Trained teacher share shows no consistent pattern across outcomes.

Threshold variation, in short, reflects the composition of the applicant pool more than the characteristics of local schools. Year fixed effects further confirm a system-wide upward shift in cut-off levels (+2.02 pp) alongside modest compression in both gaps between 2025 and 2026.

Figure 2 visualises the estimated coefficients and 95% confidence intervals across all specifications.

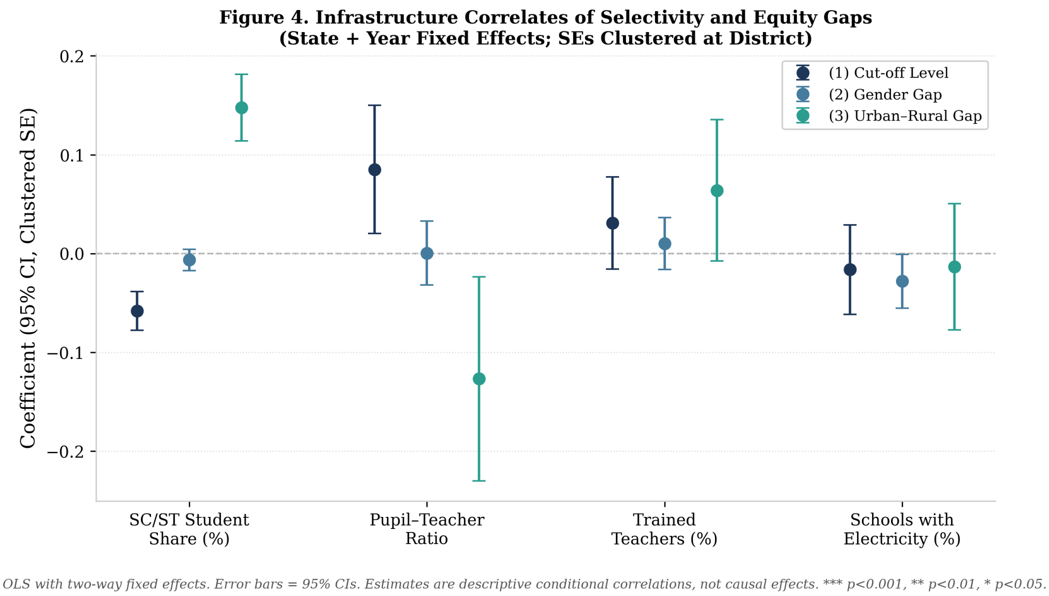


Figure 2: Correlates of JNV Selectivity and Equity Gaps. OLS coefficients with 95% CIs (clustered SEs at district level). State and year fixed effects included. Estimates are descriptive conditional correlations, not causal effects. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Source: Author's calculations using JNVST cut-off data, UDISE^{+} infrastructure data, and district boundary shapefiles.

6.3 Demographic Composition and Equity Gaps

Figure 3 plots SC/ST student share against the two equity gaps in the 2026 cross-section; the contrast between panels is sharp.

Panel A shows no systematic relationship between SC/ST share and the gender gap. The LOWESS smoother is essentially flat, and the Spearman correlation is negligible ($\rho = 0.015$, n.s., $N = 476$). However the local demographic profile is composed, boys and girls compete on roughly equal terms.

Panel B tells the opposite story. There is a strong positive association between SC/ST share and the urban–rural OBC gap ($\rho = 0.380$, $p < 0.001$, $N = 552$): districts with more SC/ST students tend to show much wider category-based gaps. The relationship is mildly non-linear—somewhat compressed at intermediate SC/ST shares of around 20–35 percent—and steepens at higher values.

SC/ST share likely proxies for differences in the depth and competitiveness of applicant pools within quota categories. Where SC/ST representation is high, the competitive landscape within reserved categories becomes more uneven, and the gap between unreserved and OBC cut-offs widens accordingly. Gender competition operates through a different mechanism and is unaffected. The asymmetry reinforces the regression evidence and points consistently toward a demand-side account of how inequality in access is generated.

(a) SC/ST Share vs. Gender Gap. Spearman $\rho = 0.015$ (n.s.), $N = 476$.

(b) SC/ST Share vs. Urban–Rural OBC Gap. Spearman $\rho = 0.380^{***}$, $N = 552$.

Figure 3: SC/ST Student Share and Equity Gaps, 2026 Cross-section. Each point = one district. LOWESS smoother (bandwidth = 0.4) shown in red. Spearman ρ reported. Association is descriptive. *Source:* Author's calculations using JNVST cut-off data, UDISE^{+} infrastructure data, and district boundary shapefiles.

7 Spatial Analysis

7.1 Geographic Distribution

Figure 4 maps Boys Open-UR cut-off scores across 562 districts in 2026, with the colour scale clipped to the 5th–95th percentile (raw range 65.0–100.0).

Even before any formal testing, a geographic logic is visible. Districts in central and peninsular India cluster at higher cut-offs; much of the northwest and northeast sits lower. The patterning is contiguous—high-cut-off districts neighbour other high-cut-off districts rather than being scattered at random. The spatial diagnostics and models that follow ask whether this clustering is statistically meaningful and whether it survives covariate controls.

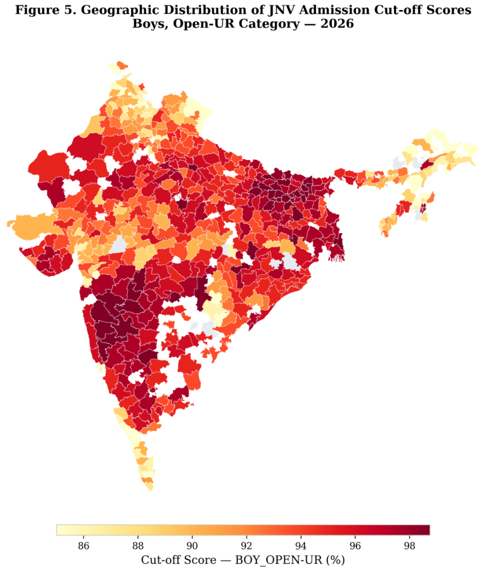


Figure 4: Geographic Distribution of JNV Admission Cut-off Scores (Boys, Open-UR), 2026. Colour clipped to 5th–95th percentile (raw range 65.0–100.0). Grey = outside analytical sample. $N = 562$.

Source: Author's calculations using JNVST cut-off data, UDISE+ infrastructure data, and district boundary shapefiles.

7.2 *Global Spatial Autocorrelation*

Table 3 reports Global Moran's I statistics for the primary outcomes. The results mirror the asymmetry already visible in the regression analysis.

Table 3: Global Moran's I Statistics for Key Outcomes (2026 Cross-section)

Fig.	Variable	N	Moran's I	$\mathbb{E}[I]$	z-score	p-value	
5	Boys Open-UR Cut-off	562	0.5725	−0.0018	20.46	0.001	**
7A	Gender Gap	482	0.0009	−0.0021	0.10	0.476	
7B	Urban–Rural OBC Gap	560	0.5560	−0.0018	19.09	0.001	**

p-values from 999 pseudo-random permutations. Queen contiguity **W**, row-standardised. ** $p < 0.01$.

Source: Author's calculations using JNVST cut-off data, UDISE+ infrastructure data, and district boundary shapefiles.

Boys Open-UR cut-offs show strong positive clustering ($I = 0.573$, $z = 20.46$, $p = 0.001$): high-cut-off districts are systematically surrounded by other high-cut-off districts, and the same pattern holds at the lower end of the distribution. The urban–rural OBC gap is nearly as clustered ($I = 0.556, z = 19.09, p = 0.001$). The gender gap, by contrast, is spatially random ($I = 0.001$, $p = 0.476$)—its distribution across the map is indistinguishable from noise.

The divergence between these three statistics is central to the paper's argument. Spatial clustering is a property of caste-based and aggregate selectivity patterns, but not of gender-based differences. The two dimensions of inequality do not simply reflect the same underlying process; they operate through distinct mechanisms.

Figure 5 presents the Moran scatter plot for Boys Open-UR cut-offs, plotting the standardised cut-off z_d against its spatial lag Wz_d. The slope of the regression line equals Moran's $I = 0.573$ by construction. Observations concentrate in the High–High and Low–Low quadrants, confirming that clustering reflects large, contiguous regional blocks of similar values rather than isolated pockets; spatial outliers (HL and LH) are relatively scarce.

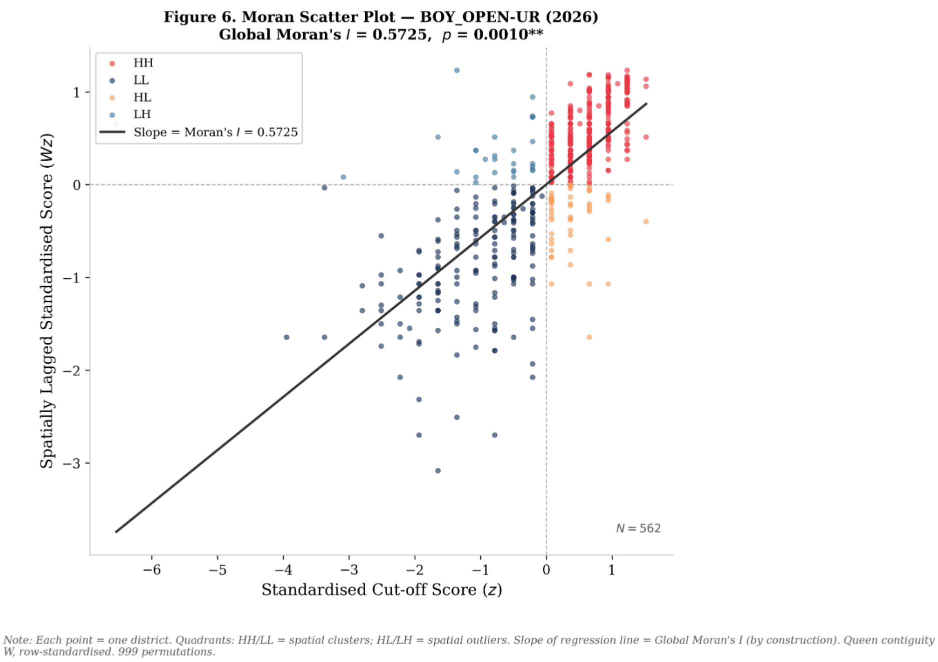


Figure 5: Moran Scatter Plot, Boys Open-UR Cut-off (2026). Global Moran's $I = 0.573$, $p = 0.001$. Each point = one district. Slope of regression line equals Moran's I by construction. Colour: HH (red), LL (blue), HL (orange), LH (light blue). Queen contiguity $\mathbf{W}$, 999 permutations. $N = 562$.

Source: Author's calculations using JNVST cut-off data, UDISE^{+} infrastructure data, and district boundary shapefiles.

7.3 LISA Cluster Maps

Figure 6 presents LISA cluster maps for the gender gap (Panel A) and the urban–rural OBC gap (Panel B), based on the 2026 cross-section with Queen contiguity weights and $p < 0.05$ (999 permutations).

Panel A confirms what the Moran statistic already suggested: gender gaps show no meaningful geographic organisation. The overwhelming majority of districts are not significantly clustered, and the few that are do not form contiguous regions—they scatter across the map. The near-zero Moran's I and the LISA map tell the same story.

Panel B is strikingly different. High–High clusters (large OBC gaps) are concentrated in northwestern India and parts of the northeast, while Low–Low clusters (compressed gaps) dominate the peninsular south. These form coherent regional blocks—not scattered individual observations—indicating that caste-based inequality operates at a geographic scale larger than the individual district. Districts with similar demographic profiles and educational environments cluster together, and the resulting differences in effective selectivity are regionally entrenched.

The contrast between the two panels encapsulates the paper's central finding: spatial structure is the signature of caste-based inequality in this system, not of gender-based inequality.

7.4 Maximum Likelihood Spatial Regression

Anselin LM diagnostics applied to OLS residuals strongly reject the null of no spatial dependence (Robust LM-Lag $p < 0.001$, Robust LM-Error $p = 0.087$), pointing clearly toward a spatial lag specification.

Table 4 reports maximum likelihood estimates of the spatial lag (SAR) and spatial error (SEM) models.

Table 4: ML Spatial Estimates

	SAR		SEM	
Variable	Coef.	SE	Coef.	SE
Constant	32.725^{***}	3.485	96.056^{***}	2.064
Electricity (%)	-0.008	0.012	-0.012	0.015
Trained teachers (%)	-0.016	0.011	-0.042^{**}	0.017
Pupil–teacher ratio	0.087^{***}	0.021	0.090^{***}	0.031
SC/ST share (%)	-0.026^{***}	0.006	-0.039^{***}	0.009
Spatial lag, $\hat{\rho}$	0.655^{***}	0.035	—	—
Spatial error, $\hat{\lambda}$	—	—	0.691^{***}	0.034
Observations	555		555	
Log likelihood	-1358.92		-1362.77	
AIC	2729.85		2735.54	
Pseudo-R^2	0.590		0.216	

Queen **W** row-standardised; 2026 cross-section. $^{***}p < 0.001$, $^{**}p < 0.01$.

Source: Author's calculations using JNVST cut-off data, UDISE+ infrastructure data, and district boundary shapefiles.

The headline result is the spatial autoregressive coefficient: $\hat{\rho} = 0.655$ ($p < 0.001$, SE $= 0.035$)—a substantial estimate. Admission thresholds in a given district co-move strongly with those in neighbouring districts, reflecting direct interdependence rather than merely shared covariates.

The practical implication follows from the spatial multiplier, $(1-\rho)^{-1} \approx 2.90$. A local improvement in conditions does not stay local: it propagates through the system and is amplified by a factor of nearly three. This shows up clearly in the effect decomposition—the direct effect of the pupil–teacher ratio is 0.087, but the total effect, once spatial feedback is accounted for, rises to 0.251. Treating districts as independent units would substantially understate how much local conditions matter.

The SC/ST student share remains negative and highly significant in both specifications—SAR ($\hat{\beta} = -0.026$) and SEM ($\hat{\beta} = -0.039$)—confirming that demographic composition is a robust predictor of admission thresholds even after controlling for spatial dependence. The SAR model also fits substantially better than non-spatial OLS (pseudo-$R^2 = 0.590$ vs. 0.246) and outperforms the SEM on AIC (2729.85 vs. 2735.54), reinforcing the interpretation that spatial dependence operates through interactions in outcomes rather than through spatially correlated unobservables.

8 Robustness

8.1 *Alternative Cut-off Outcomes*

Spatial autocorrelation is positive and statistically significant at the 1% level across all 16 cut-off categories in the 2026 cross-section—geographic clustering is not specific to the headline Open-UR variable but is a pervasive feature of admission thresholds across the board.

The magnitude varies informatively. Clustering is strongest for Rural-OBC cut-offs, followed by Open-UR and Open-OBC, and progressively weaker for SC, ST, and disability quota groups. Categories with broader, more competitive applicant pools cluster more strongly while thinner pools do not—consistent with clustering reflecting geographically correlated preparedness rather than a small-sample artefact.

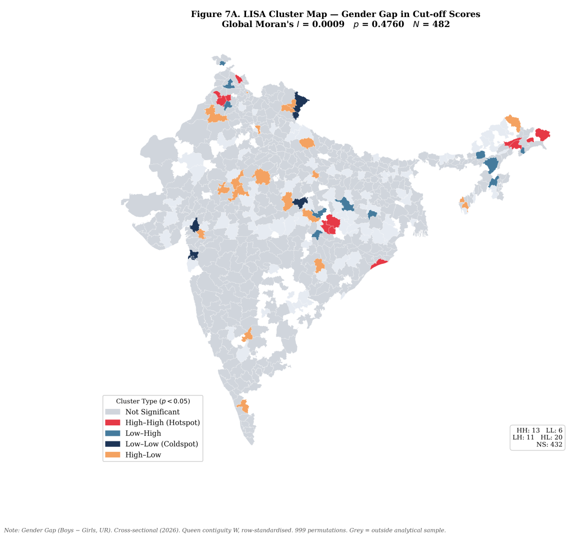
8.2 *Alternative Spatial Weight Matrices*

We re-estimate the SAR model replacing the Queen contiguity matrix with a 5-nearest-neighbour (KNN-5) matrix to check whether the results are sensitive to how neighbourhood is defined.

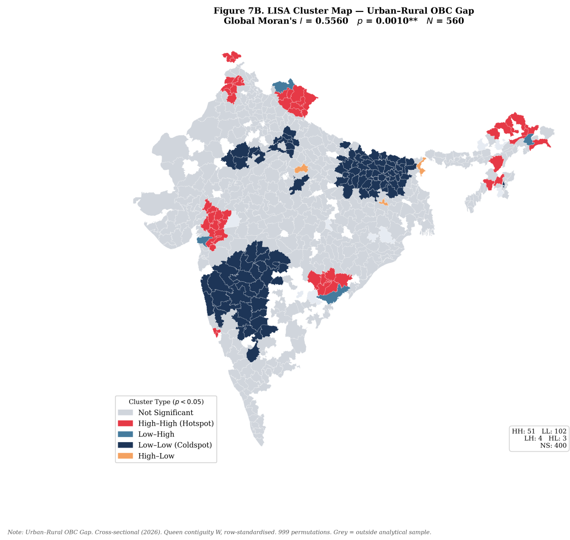
They are not. The spatial autoregressive coefficient is virtually unchanged ($\hat{\rho} = 0.655$ under Queen contiguity, 0.657 under KNN-5), and the SC/ST coefficient is identical ($\hat{\beta} = -0.026$) across both. Queen contiguity fits marginally better (AIC $= 2729.85$ vs. 2745.61), suggesting that geographic adjacency is a slightly better representation of spatial interactions here, but the substantive conclusions are the same under either specification.

8.3 *Year-by-Year Stability*

Estimating the SAR model separately for 2025 ($N = 535$) and 2026 ($N = 555$) yields $\hat{\rho} = 0.613$ and $\hat{\rho} = 0.655$, respectively—both significant at the 1% level. The modest increase in $\hat{\rho}$ between years is small relative to the overall level of spatial dependence and should not be over-interpreted; what matters is that the structure is stable. Spatial dependence in this system is a persistent feature, not a 2026-specific artefact.



(a) Gender Gap. $\hat{I} = 0.001$, $p = 0.476$, $N = 482$. HH: 13, LL: 6, LH: 11, HL: 20, NS: 432.



(b) Urban–Rural OBC Gap. $\hat{I} = 0.556$, $p = 0.001$, $N = 560$. HH: 51, LL: 102, LH: 4, HL: 3, NS: 400.

Figure 6: LISA Cluster Maps. Cluster classification at $p < 0.05$ (999 permutations). Queen contiguity **W**, row-standardised. Grey = outside analytical sample. Year = 2026. Panel A: Gender Gap. Panel B: Urban–Rural OBC Gap.

Source: Author's calculations using JNVST cut-off data, UDISE^{+} infrastructure data, and district boundary shapefiles.

Table 5: Robustness: Alternative Weights and Year-by-Year Estimates

Specification	N	$\hat{\rho}$	AIC	SC/ST Coef.
Alternative weight matrices (2026 cross-section):				
Queen Contiguity (baseline)	555	0.655***	2729.85	−0.026
5-Nearest Neighbours	555	0.657***	2745.61	−0.026
Year-by-year estimates (Queen contiguity):				
2025	535	0.613***	2846.23	—
2026	555	0.655***	2729.85	−0.026

*** $p < 0.001$. ML Spatial Lag (SAR). Controls: SC/ST share, pupil–teacher ratio, trained teachers, electricity share.

Source: Author's calculations using JNVST cut-off data, UDISE^{+} infrastructure data, and district boundary shapefiles.

8.4 *Sample Restrictions*

Dropping the six states with the most districts (Uttar Pradesh, Madhya Pradesh, Rajasthan, Maharashtra, Bihar, and Karnataka) reduces the sample to $N = 520$ observations. The main results are essentially unchanged. SC/ST student share retains a strong negative association with cutoff levels ($\hat{\beta} = -0.062$, $p < 0.001$, vs. -0.058 in the full sample) and a positive association with the urban–rural OBC gap ($\hat{\beta} = 0.138$, $p < 0.001$, vs. 0.148). Year fixed effects continue to show a system-wide upward shift in levels, and infrastructure variables remain weak predictors. The findings are not driven by India's largest states.

9 Policy Implications

Two policy-relevant implications emerge from the empirical findings.

First, spatial targeting is likely to matter. The estimated spatial multiplier implies that local conditions do not stay local—improvements in one district propagate to neighbours and are amplified in equilibrium. Interventions concentrated in geographically contiguous clusters of low-performing districts may therefore generate larger aggregate returns than isolated, district-level efforts. The corollary is sobering: districts embedded in low-performing regions face a structural headwind that a nationally uniform policy cannot easily offset.

Second, infrastructure investment alone is probably not enough. The weak and inconsistent relationship between school infrastructure and admission thresholds stands in sharp contrast to the strong, stable association between demographic composition and equity gaps. This points toward demand-side constraints—unequal access to information, preparatory support, and mentoring—as the more proximate determinants of who gets in. Targeted interventions addressing these constraints, particularly in districts with high SC/ST shares and wide urban–rural gaps, are likely to be a more effective complement to infrastructure spending than infrastructure alone.

These implications are informed by the evidence rather than mechanically derived from it. The analysis is descriptive, and the absence of student-level data means the underlying mechanisms remain inferred rather than directly identified. That said, the consistency of the patterns across specifications, geographies, and years makes the demand-side interpretation difficult to dismiss.

10 Conclusion

This paper offers the first systematic spatial analysis of access to the JNV system using district-level admission cut-off scores. Two findings stand out.

Demographic composition, not observable infrastructure, is the primary correlate of both selectivity and equity gaps. Districts with higher SC/ST student shares face lower cut-offs and wider urban–rural OBC gaps, while school infrastructure measures—electricity, trained teachers, pupil–teacher ratios—add little once state and year effects are controlled for. Admission thresholds are better understood as a reflection of who is competing than of what schools look like.

Admission thresholds are also strongly spatially clustered. Districts with similar cut-offs tend to neighbour one another, and spatial econometric estimates point to substantial interdependence across district boundaries. The spatial multiplier of roughly 2.9 means that local conditions are considerably amplified in equilibrium.

A key asymmetry connects both findings. Gender gaps are small, spatially unstructured, and unrelated to demographic composition—the gender quota appears to achieve broadly equal access regardless of local context. Urban–rural OBC gaps are large, spatially organised, and closely tied to SC/ST share, indicating that category-based inequality reflects persistent differences in applicant pool characteristics that uniform reservation rules alone cannot correct.

Several limitations apply. The panel covers only two years, which limits the ability to study longer-run dynamics, and the absence of student-level data means that the demand-side mechanisms inferred here—variation in coaching access, information, household resources—cannot be directly tested. Future work combining longer-run JNVST records with household survey data such as NFHS-5 would be well-placed to unpack these mechanisms.

The broader lesson is simple but consequential: equal rules do not produce equal outcomes when the applicant pool is geographically unequal. Policies designed without attention to spatial clustering and demographic heterogeneity may raise average performance while leaving the distribution of access largely unchanged.

References

Andrabi, T., Das, J., and Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, 107(6):1535–1563.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2):93–115.

Bagde, S., Epple, D., and Taylor, L. (2016). Does affirmative action work? Caste, gender, college quality, and academic success in India. *American Economic Review*, 106(6):1495–1521.

Bertrand, M., Hanna, R., and Mullainathan, S. (2010). Affirmative action in education: Evidence from engineering college admissions in India. *Journal of Public Economics*, 94(1–2):16–29.

Case, A. C., Rosen, H. S., and Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, 52(3):285–307.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–1774.

Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191.

Khanna, G. (2023). Large-scale education reform in general equilibrium: Regression discontinuity evidence from India. *Journal of Political Economy*, 131(2):405–444.

LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton.

Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1):39–77.

Pande, R. (2003). Can mandated political representation increase policy influence for disadvantaged minorities? *American Economic Review*, 93(4):1132–1151.

Rey, S. J. and Montouri, B. D. (1999). US regional income convergence: A spatial econometric perspective. *Regional Studies*, 33(2):143–156.

A Supplementary Tables

A.1 State-Level Gender Gap Fixed Effects

Table 6 reports state fixed effects from the two-way FE regression of the gender gap on state and year indicators. The omitted category is Andaman & Nicobar; all estimates are relative to this baseline. With the exception of Lakshadweep and, to a lesser degree, Telangana, every state is statistically indistinguishable from parity—reinforcing the finding that gender differences in admission thresholds are not geographically structured in any meaningful way.

Table 6: State Fixed Effects: Gender Gap in Open-UR Cut-off (Boys − Girls, pp)

State	Coef. (pp)	SE (pp)	95% CI
Andaman & Nicobar (baseline)	0.00	—	—
Arunachal Pradesh	+0.12	0.86	[−1.57, +1.81]
Assam	+0.14	0.57	[−0.97, +1.25]
Bihar	+0.18	0.45	[−0.70, +1.07]
Chhattisgarh	+0.11	0.66	[−1.20, +1.41]
Delhi	+0.88	1.33	[−1.74, +3.49]
Gujarat	+0.55	0.50	[−0.42, +1.53]
Haryana	+0.05	0.40	[−0.74, +0.83]
Himachal Pradesh	−0.29	0.47	[−1.22, +0.63]
Jharkhand	+0.74	0.59	[−0.41, +1.89]
Karnataka	+0.48	0.43	[−0.37, +1.33]
Kerala	−0.20	0.38	[−0.94, +0.54]
Lakshadweep	−2.59	0.48	[−3.52, −1.65]
Madhya Pradesh	+0.53	0.41	[−0.26, +1.33]
Maharashtra	+0.10	0.42	[−0.73, +0.93]
Manipur	−0.90	0.65	[−2.17, +0.37]
Odisha	−0.13	0.47	[−1.06, +0.79]
Punjab	+0.03	0.30	[−0.57, +0.62]
Rajasthan	+0.40	0.45	[−0.48, +1.27]
Telangana	+1.30	0.53	[+0.25, +2.34]
Tripura	+0.37	0.73	[−1.06, +1.80]
Uttar Pradesh	+0.51	0.45	[−0.37, +1.40]
Uttarakhand	−0.19	0.62	[−1.41, +1.03]
West Bengal	+0.56	0.60	[−0.62, +1.75]

Source: Author's calculations using JNVST cut-off data, UDISE^{+} infrastructure data, and district boundary shapefiles.