Inter-Community Educational Participation in Indian Districts and Stochasticity of Wealth

Arnab Samanta

Introduction

- What explains the within- and between-district variation in educational attainment/participation in India?
- Inter-community (caste, religion) variation
- What can we say about the influence of wealth on inter-community educational participation?

We attempt to answer these questions using data from the fifth round(2019-21) of NFHS (National Family Health Survey). The econometric tools employed include a district level mixed-effect multistage regression model and a simple Markov chain process to model community perceptions about social transition.

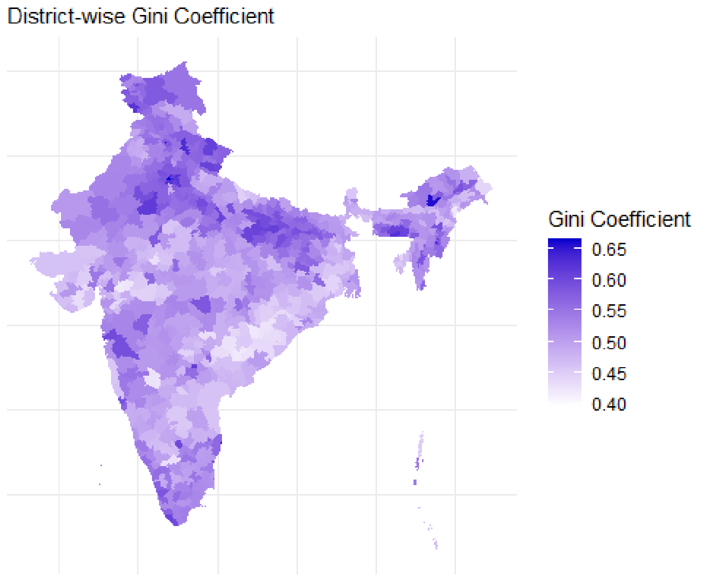
Inequality in Educational Attainment : Within Districts

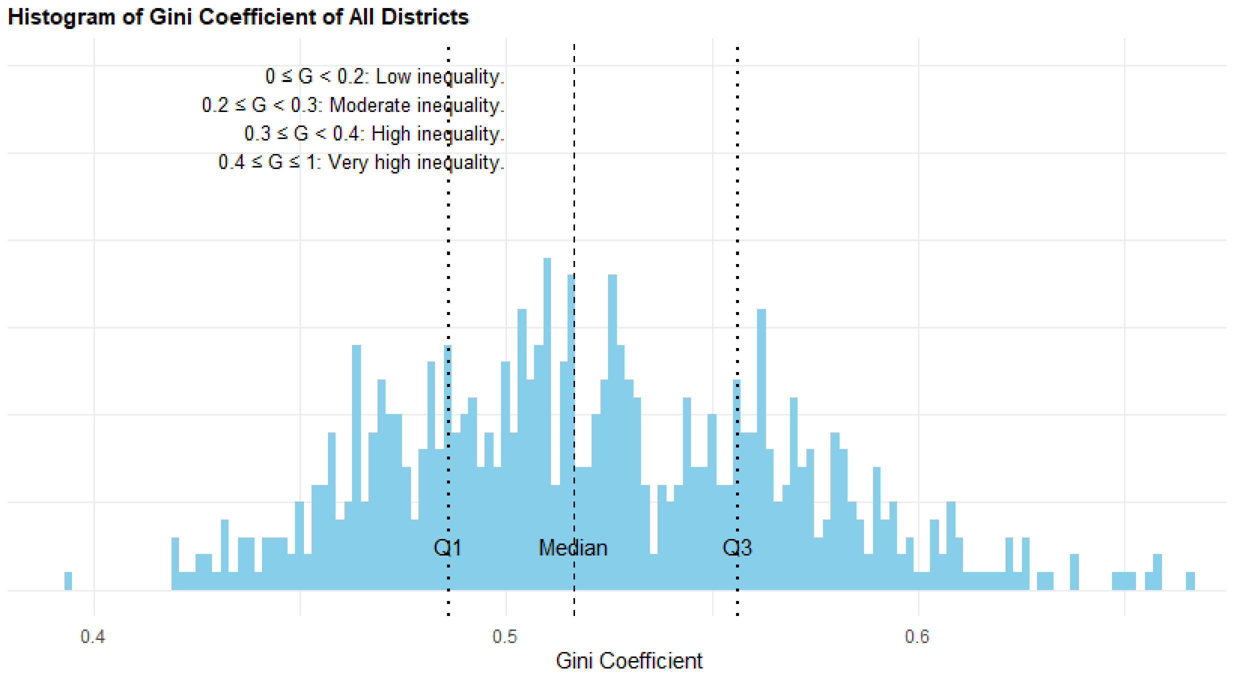
The Gini Coefficients of Educational Attainment within Indian Districts are extremely high for every Indian District

- There is no Indian District that does not have very high inequality level for educational attainment.
- Across 700+ Indian districts, the Gini Coefficient ranged from a minimum of 0.394 in Devbhumi Dwarka to a Q1 value of approximately 0.47 . In the capital of India, it was 0.694 in South-West Delhi and 0.63 in New Delhi.

Age-adjusted Number of Years of Education

- From the raw data on no. of years of education, we derived an age adjusted number of years of education indicator on a scale of 0-1 where 0 indicates no education.





Data Source : NFHS 2019-21

Inequality in Educational Attainment : Between Districts

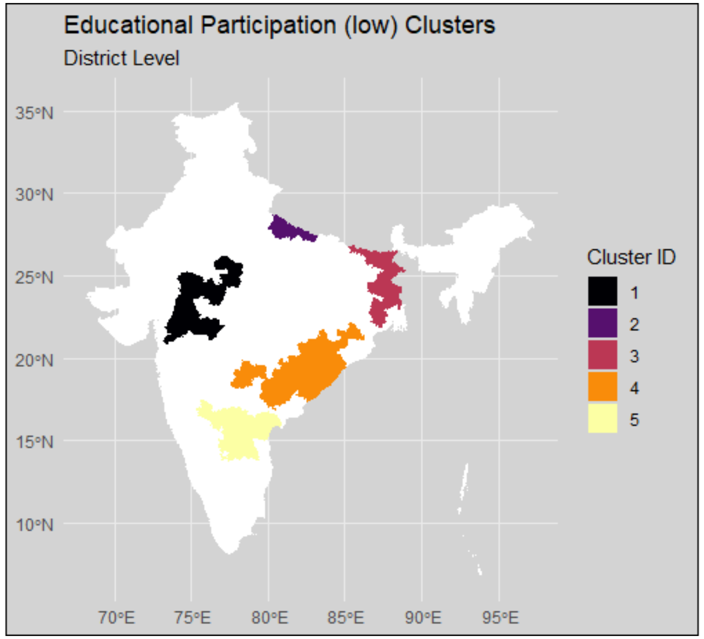
Local Indicator of Spatial Autocorrelation (LISA)

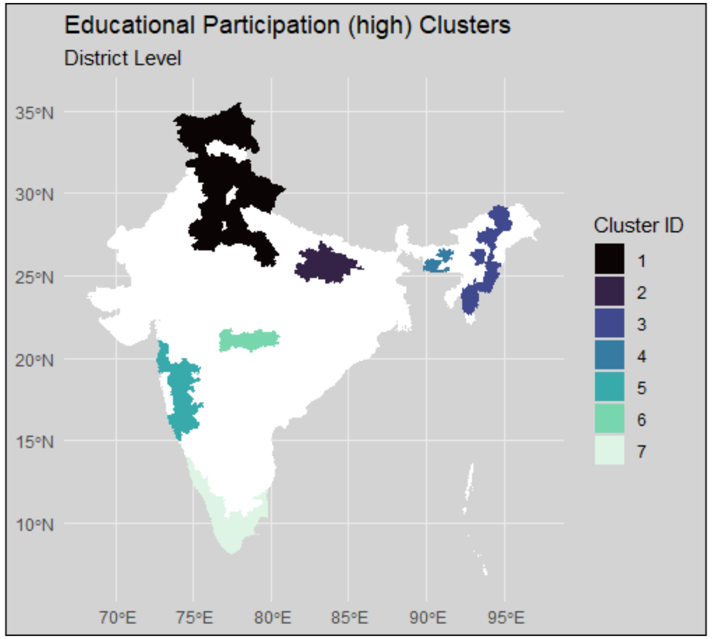
$$\mathrm{LISA_i} = \frac{\mathrm{x_i} - \bar{\mathrm{x}}}{\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}}(\mathrm{x_i} - \bar{\mathrm{x}})^2} \sum_{\mathrm{j \neq i}} \mathrm{w_{ij}}\left(\mathrm{x_j} - \bar{\mathrm{x}}\right) \quad \text{where,} \quad \mathrm{w_{ij}} \text{ (spatial weight modelled over distance)} = \exp\left(-\frac{\text{distance in Km}}{\text{threshold}}\right)$$

- The magnitude of $\mathrm{LISA_I}$ indicates how strongly, positively or negatively the neighboring x values (in our case educational attainment) are correlated to the i^{th} spatial entity (in our case the i^{th} district)
- As , distance increases $\mathrm{w_{ij}}$(spatial weight) diminishes to zero and thus very distant distrtcts contrubiute nothing to the Autocorrelation measuure

Spatial Clusters

- A high negative $\mathrm{LISA_i}$ indicates special heterogeneity whereas a high positive $\mathrm{LISA_I}$ indicates spatial similarity. Spatial similarity may indicate a cluster of spatial entities (in our case districts) - A) a hot spot or B) a cold spot, with respect to the X variable (in our case educational attainment). (Anselin, 1995)





Data Source : NFHS 2019-21

Caste-Religion Segregations in Indian Society:

Definition
Characterization
Effect

Definition :

Two concepts frequently equated with caste are varna (a system delineated within Hindu sacred texts. There can be only four varnas – that is, Brahmin, Kshatriya, Vaishya, and Shudra) and jāti (closely linked to the typical western concept of caste, originating as it does from the word jana (birth) and to the social identity ascribed by birth). Within each varna, there are many jātis forming a complex hierarchy of occupational communities (Sahoo, 2017)

Articles 341 and 342 of the Indian Constitution include a list of Scheduled Castes and Scheduled Tribes (broadly constituting India's 'untouchable' castes) who were to receive positive discrimination in education and political representation. That was as per Indian constitutional system we have four castes : general/upper cast, Scheduled Caste (SC), Scheduled Tribes (ST) , Other Backward Castes (OBC) .

Characterization :

One's caste can be known to neighbors if they still live in the ancestral locality or if they self-report it.

But, unlike race caste can not be determined by appearance or skin color definitively except for some cases. Last names can be the most definitive way but there many exceptions and complexities as surnames may vary between provinces.

Finally, interaction with other socioeconomic phenomena over time has made it appear more complex as to what extent caste matters. There are certain features that are historical and certain aspects that are contemporary.

Question such as whether non-Hindu backward households/communities can also be brought under caste-based definitions.

Effect :

In recent studies we often observe it is almost impossible to interpret empirical sociopolitical data drawn from Indian population studies without considering caste and religion. There is a huge literature that conducted both qualitative and quantitative work.

In this study as part of our exploratory work before describing the models to understand the association of different demographic features and educational attainment, we will present some examples to provide some understanding about to what extent caste positively or negatively influence social demographic phenomena. (please see the main paper for long tables)

Baseline Model

Number of Years Education_{ijd} =

$$\beta_{0d} + \sum_{k} \beta_{kd} \cdot (\text{demographic\_factor}_k) + \sum_{p} \beta_{pd} \cdot \left(\text{Gender} \times \text{demographic\_factor}_p\right) + \beta_{6d} \cdot \underbrace{\exp\left(\frac{-1}{\text{age}_{ijd}^2}\right)}_{\text{represents the secondary schooling dropout}}$$

$$+\ u_{jd} + \epsilon_{ijd}$$

This is our regression model for the d^{th} district. We perform 707 district-level regressions for 707 districts of India: $d = 1, 2, \dots, 707$

Here,

$u_{jd} \sim N(0, \sigma_u^2)$, Household level random effect for j^{th} Household at the d^{th} district.

Number of Years Education_{ijd} , Number of years of education for i^{th} individual from j^{th} Household at the d^{th} district.

(We focused on an age group ranging from 5 to 25 years old)

$\epsilon_{ijd} \sim N(0, \sigma^2)$ Residuals

$k \in \{\text{Muslim}, \text{SC/ST}, \text{Female}\}$, $p \in \{\text{Muslim}, \text{SC/ST}\}$

Model with Wealth Index (quintile) as a Covariate

$\text{Number of Years Education}_{ijd} =$

$$\beta_{0d} + \sum_k \beta_{kd} \cdot (\text{demographic\_factor}_k) + \sum_p \beta_{pd} \cdot \left(\text{Gender} \times \text{demographic\_factor}_p\right)$$
$$+ \quad \beta_{6d} \cdot \underbrace{\exp\left(\frac{-1}{\text{age}_{ijd}^2}\right)}_{\text{represents the secondary schooling dropout}} + \beta_{7d} \cdot \text{Wealth\_Index}_{ijd} + u_{jd} + \epsilon_{ijd}$$

$\text{Wealth\_Index}_{ijd}$ is the wealth quintile for i^{th} individual from j^{th} Household at the d^{th} district.

As it is quintile-based indicator - it can have 5 categories where top Wealth index = 5 indicates ‘richest’ and 1 indicates ‘poorest’. The preparation of this indicator for each household was meticulously done by the NFHS survey and it is adjusted for the respective urban or rural wealth distribution.

Spatial Spillover : Bayesian Approximation

Spillover : We performed 707 district level missed-effect regressions. But neighbouring districts may have spillover effects.

Multistage effect : The districts are part of other broader entities such as the state/country.

Most common solution : Geographically Weighted Regression (GWR) is a widely utilized. GWR functions as a localized fitting method where regression coefficients are influenced using a distance matrix . The neighborhood spillover is captured using a weight matrix $(X^{\top}W_iX)^{\{-1\}}X^{\top}W_iy$. But this approach is impractical for moderately large datasets. (For example, using Spgwr package in R, an estimated two weeks or more time is needed for a dataset comprising one hundred thousand points and five predictor variables (Harrish et al. 2010))

Bayesian Approximation:

Step 1: We hypothesize that the spatial clusters used, being based on spatial autocorrelation, will contain any spatial influence within their boundaries.

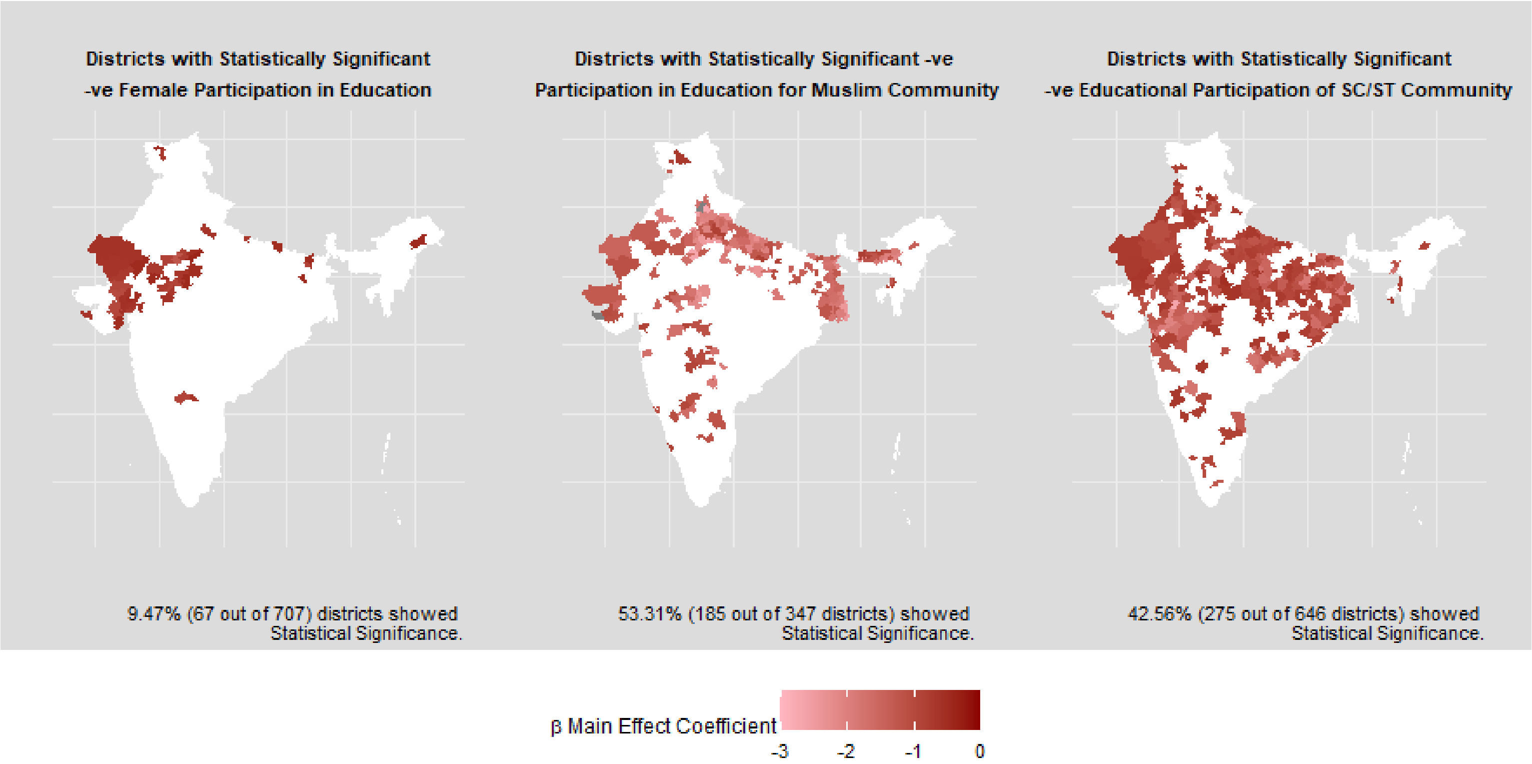
Step 2: We apply empirical Bayesian approach to the district-level estimates (β's), following the James-Stein estimator (Stein, 1961), for district d at a cluster K (K = 1,2,…,12) we derive the shrunk estimate for the m^{th} parameter effect (such as gender caste, wealth etc.) . If n_K is the number of districts in cluster K :

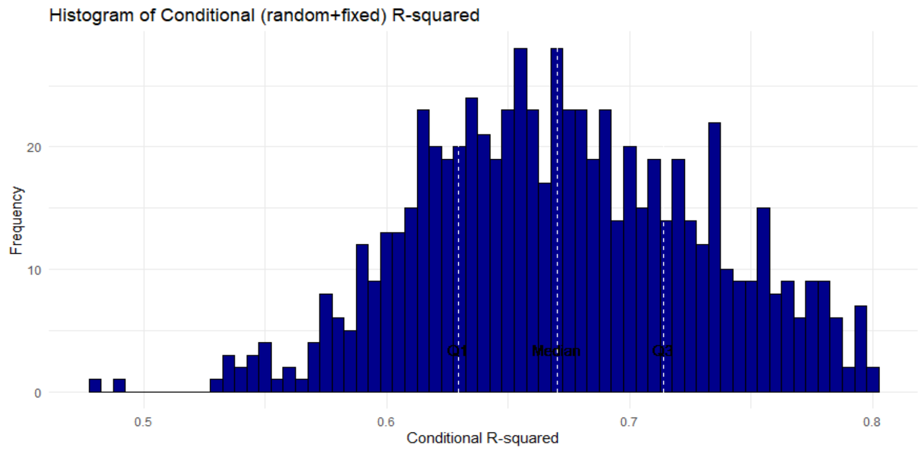
$$\text{the updated estimates :} \qquad \beta_{m,d}^{\widehat{\text{Stein}}} = \lambda_d \, \mu_{\beta_m} + (1-\lambda_d)\beta_{m,d}$$

Where, $$\lambda_d = \max\left\{0, 1 - \frac{(n_K-2)\widehat{\sigma^2}}{\sum_{i=1}^{n_K}\left(\beta_{m,i}-\mu_{\beta_m}\right)^2}\right\} \text{ with } \mu_{\beta_m} = \frac{1}{n_K}\sum_{i=1}^{n_K}\beta_{m,i} \text{ and } \widehat{\sigma^2} = \frac{1}{n_K-1}\sum_{i=1}^{n_K}\left(\beta_{m,i}-\mu_{\beta_m}\right)^2$$

The λ_d is called the borrowing factor/shrinking factor – that is how much a district borrows from its neighbour in the same cluster.

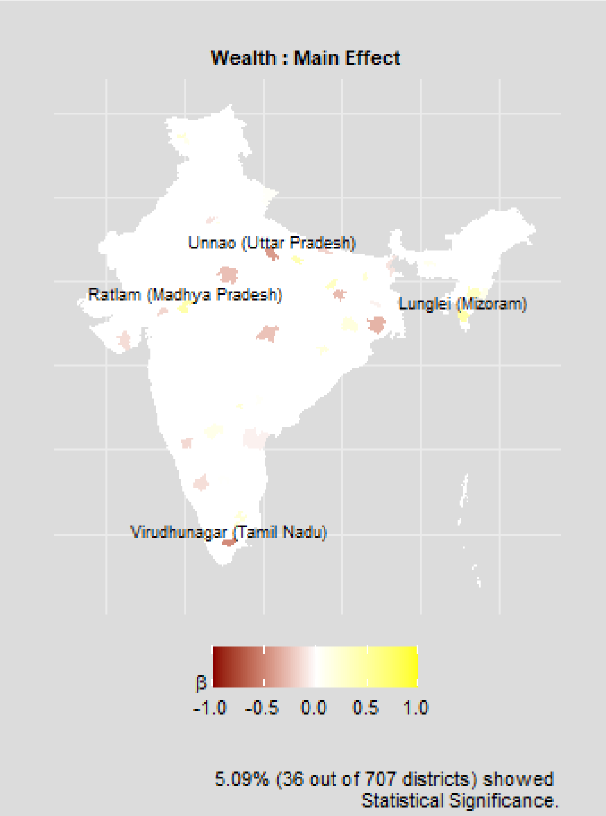
Baseline Regressions : Main Effects (interaction effects were mostly not significant)

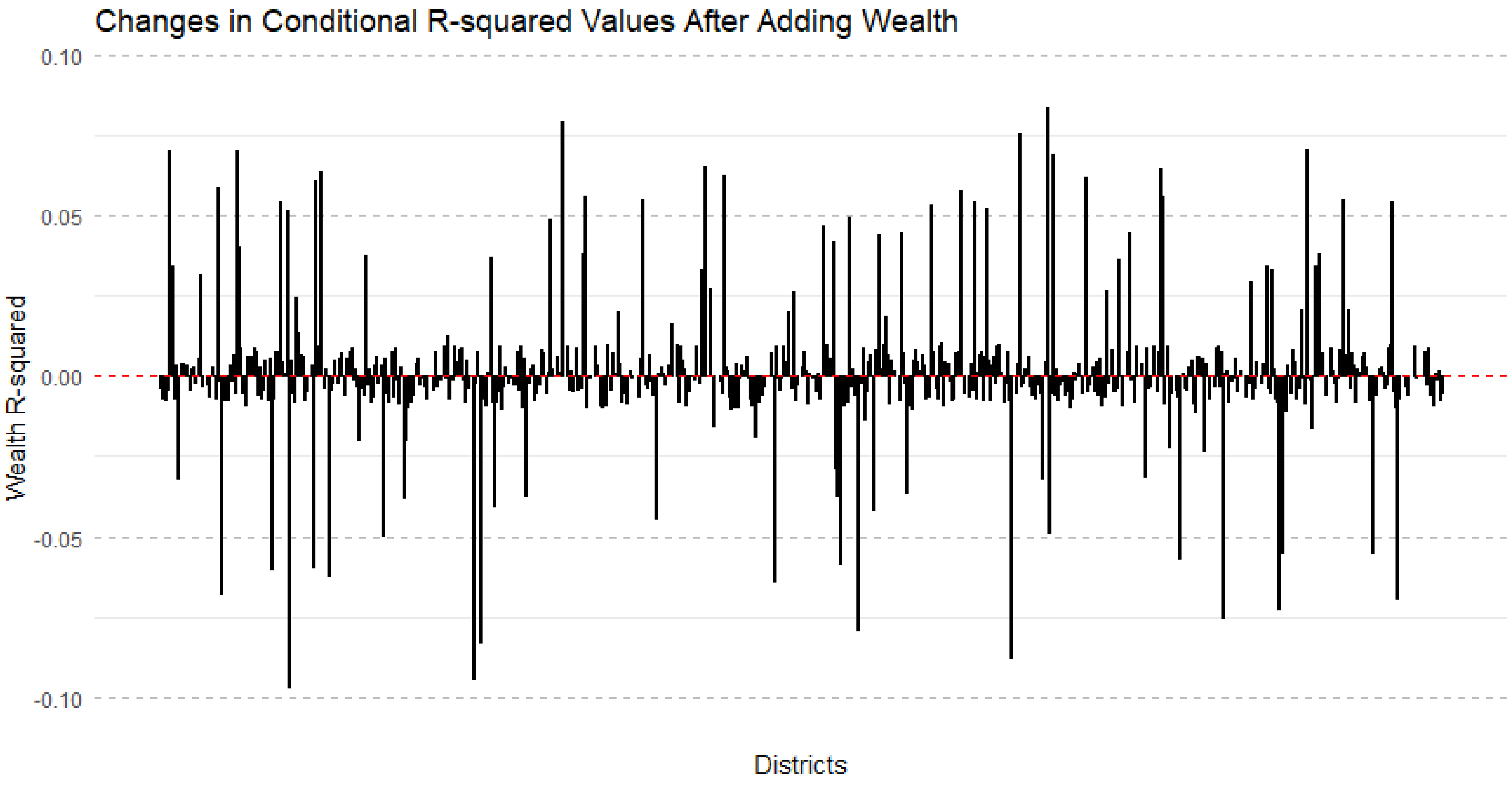




Data Source : NFHS 2019-21

Regression Results After Wealth Included: Main Effects and Conditional R-Squared





Data Source : NFHS 2019-21

What does it imply if wealth is not a significant or key factor in influencing educational attainment?

➔ First, educational participation is inherently an action or behavior.

➔ In our model we controlled for Household level random effect, the district level control account for the access/administration aspects to some extent, the Bayesian approximation accounts for the possible spillover from neighboring districts.

➔ Despite the considerable effect size and the NFHS dataset containing at least 3000 observations for any district, which suggests strong statistical power to detect true effect, wealth did not emerge as significant.

➔ Therefore, if demographic characteristics (caste, religion, gender) account for the variability in that participation behaviour, we can conclude that educational attainment is a community behaviour!

Community Behaviour

By saying community behaviour, we do NOT mean it is rooted in cultural or tradition !

There can be many explanations:

Within-District Residential Segregation: Even within districts, residential areas can be highly segregated.

Low Return to Education: Qualitative studies indicate that students often report facing different kinds of questions in interview settings, which discourages them from pursuing higher education.

Historical Professional Restrictions: While there may not always be direct discrimination, the historical association of castes or Jatis with specific professions can result in children experiencing prejudice or discouragement based on their parents' occupation.

Correlation Between Wealth and Demographics: Wealth segregation and mobility are closely linked to demographic segregation.

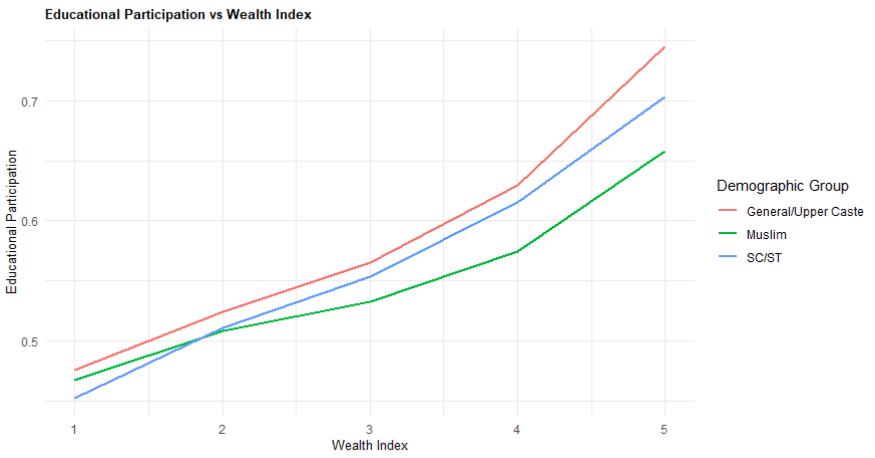
Modelling Community Class Transitions Perception:

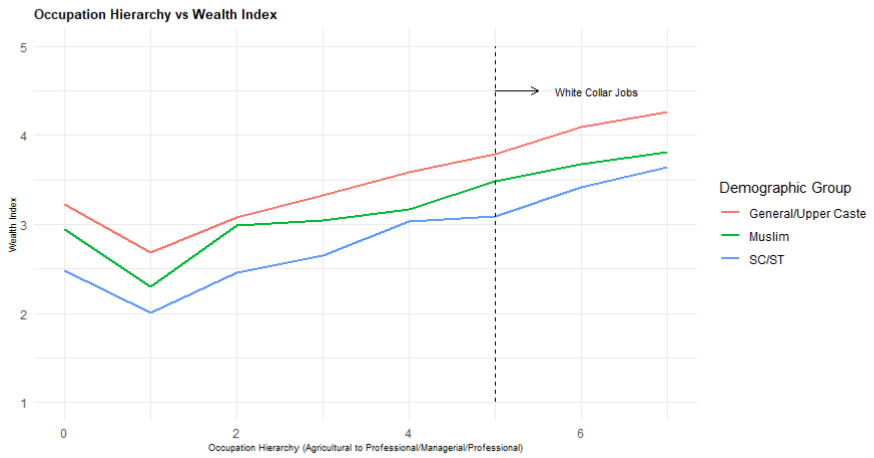
-

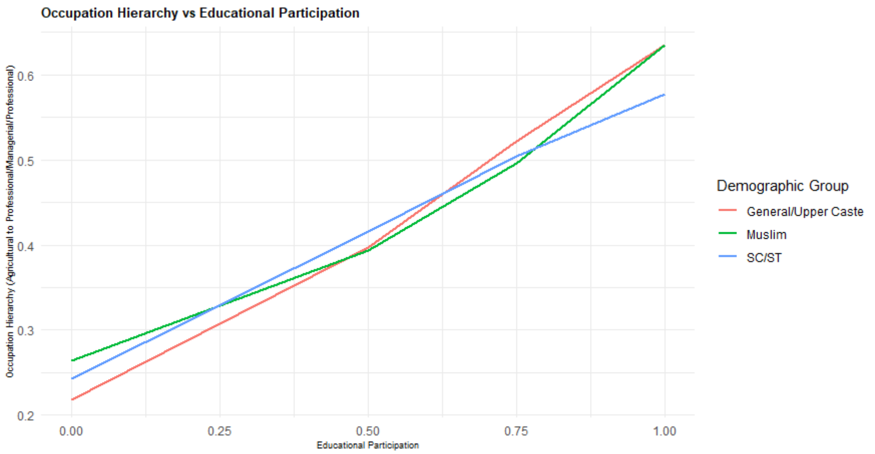
Markov Chain Analysis

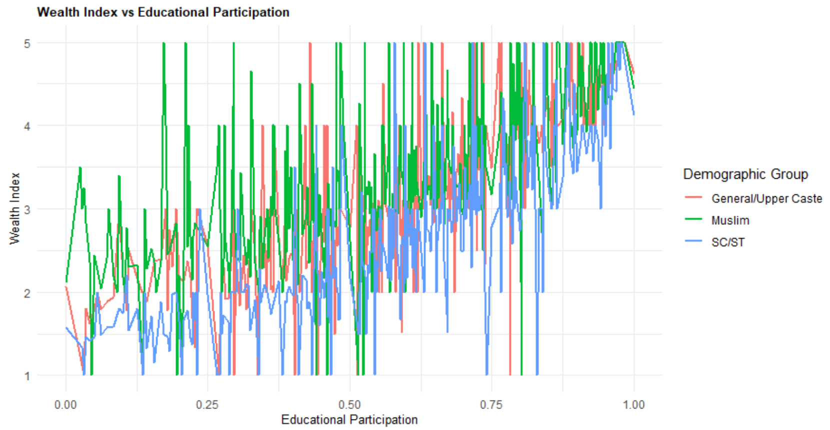
-

From the four graphs we can calculate a set of probabilities









Data Source : NFHS 2019-21

Markov Chain Analysis :

-

For each community what's the probability perception of starting in wealth state W_i and moving to wealth state W_j?

-

- Age group – 12 to 25

• What is the probability of an individual from Wealth Group W_i achieving Education Level E_i ?

- Age group – 25 or above

• What's the probability of an individual with Education Level E_i obtaining Occupation O_i?

• What's the probability of an individual in Occupation O_i accumulating Final Wealth W_j^f?

Markov Chain Analysis : Final Probabilities

P_{w_j,w_i} = P (Final wealth = W_j^f|initial wealth = W_i)

$$\sum_{k=1}^{4} P\left(\text{Final wealth} = W_j^f \,\middle|\, \text{Occupation} = O_k\right) \cdot \sum_{i=1}^{3} P\left(\text{Occupation} = O_k \,\middle|\, E_i\right) \cdot P\left(E_i \,\middle|\, W = W_i\right)$$

For example, For general/upper-caste group for whom W= ‘middle’ :

a) P (E_i|W = ‘middle’) = % of 12-25 years old with education E_i and Wealth index = ‘middle’.

b) P ($\text{Occupation} = O_k|E_i$) = % of 25+ years old (we change the subset to all employable age groups)

The combined probability P ($\text{Occupation} = O_k$|W = ‘middle’), for general/upper-caste group will thus be :

$\Sigma_{i=1}^{3} P\left(\text{Occupation} = O_k \,|\, E_i\right) \cdot P\left(E_i \,|\, W = \text{'middle'}\right)$

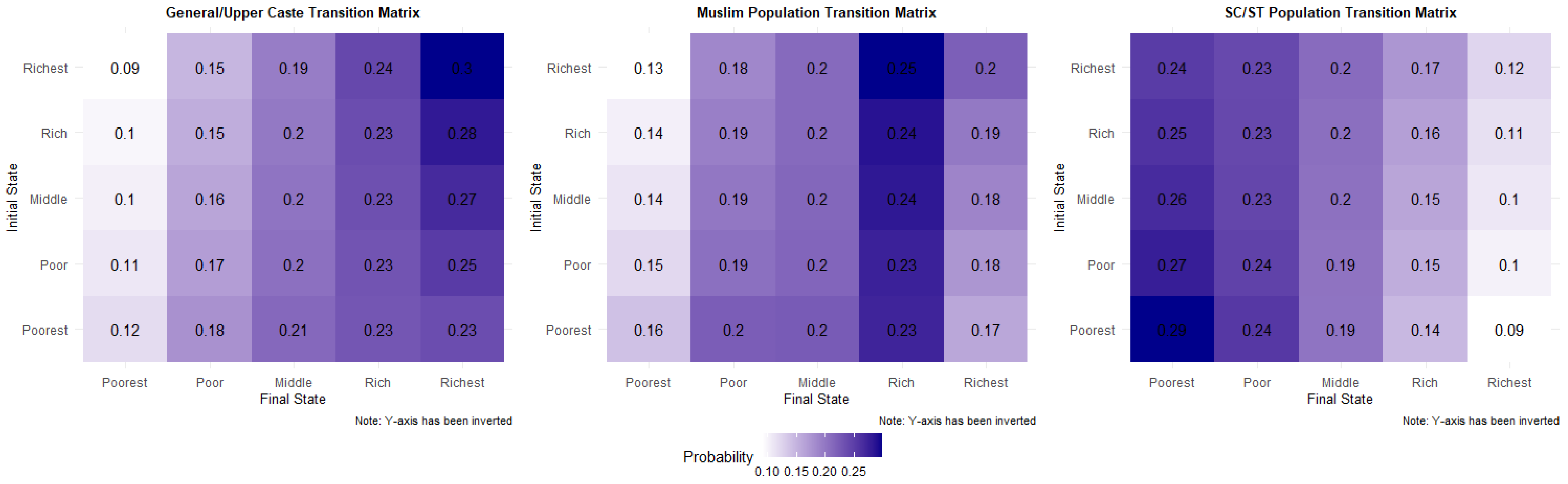
c) P (Final wealth = $W_j^f|\text{Occupation} = O_k$)

= $\Sigma_{k=1}^{4} P\left(\text{Final wealth} = W_j^f \,|\, \text{Occupation} = O_k\right) \cdot P\left(\text{Occupation} = O_k \,|\, W = \text{'middle'}\right)$

= $\Sigma_{k=1}^{4} P\left(\text{Final wealth} = W_j^f \,|\, \text{Occupation} = O_k\right) \cdot \Sigma_{i=1}^{3} P\left(\text{Occupation} = O_k \,|\, E_i\right) \cdot P\left(E_i \,|\, W = \text{'middle'}\right)$

If we replace W =‘middle’ by $W = W_i$ we will get the general expression for P_{w_j,w_i} as stated above.

For each group we will have a set of 5x5 = 25 such probability expressions which can be set as a 5X5 Markov Transition Matrix.



Data Source : NFHS 2019-21

Steady-State Distribution

-

The steady-state distribution of a Markov chain represents the long-term behavior of the system. It shows the proportion of time that the system will spend in each state if it is observed over a long period. In our case, we interpret it as the expectations different communities hold regarding their long-term wealth mobility.

-

If $\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$ is the steady-state distribution and P is the transition matrix, then :

$$\pi_1 = \pi_1 P_{w_1,w_1} + \pi_2 P_{w_1,w_2} + \pi_3 P_{w_1,w_3} + \pi_4 P_{w_1,w_4} + \pi_5 P_{w_1,w_5}$$

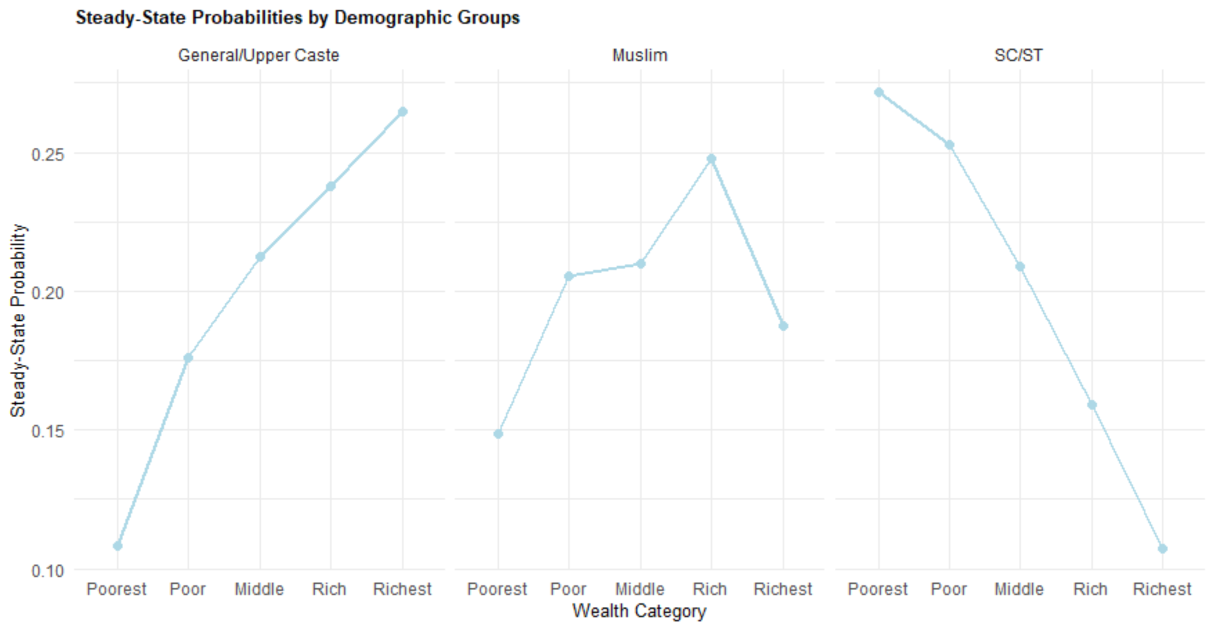
$$\pi_2 = \pi_1 P_{w_2,w_1} + \pi_2 P_{w_2,w_2} + \pi_3 P_{w_2,w_3} + \pi_4 P_{w_2,w_4} + \pi_5 P_{w_2,w_5}$$

$$\pi_3 = \pi_1 P_{w_3,w_1} + \pi_2 P_{w_3,w_2} + \pi_3 P_{w_3,w_3} + \pi_4 P_{w_3,w_4} + \pi_5 P_{w_3,w_5}$$

$$\pi_4 = \pi_1 P_{w_4,w_1} + \pi_2 P_{w_4,w_2} + \pi_3 P_{w_4,w_3} + \pi_4 P_{w_4,w_4} + \pi_5 P_{w_4,w_5}$$

$$\pi_5 = \pi_1 P_{w_5,w_1} + \pi_2 P_{w_5,w_2} + \pi_3 P_{w_5,w_3} + \pi_4 P_{w_5,w_4} + \pi_5 P_{w_5,w_5}$$

Where, $\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1$ and π_i is the long-run probability that the system (the specific demographic group) be in state i.



Data Source : NFHS 2019-21

In conclusion

- By analyzing individuals of school-going age and males in the labor market age group, the study aimed to develop a mathematical model for what we already knew from various qualitative studies regarding inter-community socioeconomic behaviour and social security, mobility.

- Demographic communities experience differing realities. Therefore, the observed patterns in class transitions make it unsurprising that communities exhibit distinct behaviour to educational participation.

- Although not a very direct causal link, this understanding sheds light on why wealth may not serve as a reliable predictor of educational attainment.

- It is reasonable to argue that policies should be aimed at improving the upward transition probabilities for SC/ST and Muslim communities to encourage higher educational participation.