Caste in Stone: Cultural Distance and the Education Game

Working Paper Draft – 2025

Bishmay Barik

Research Associate
Centre for Advanced Financial Research and Learning (CAFRAL)
Reserve Bank of India

bishmaykbarik@gmail.com
bishmay.barik@cafral.org.in

This paper represents original work by the author. It has not been submitted to any journal earlier. It was written as part of the Master's Thesis at Shiv Nadar University, India, under the supervision of Dr. Pubali Chakraborty (pchakraborty@bates.edu) and Dr. Partha Chatterjee (partha.chatterjee@snu.edu.in).

Acknowledgements

I would like to express my heartfelt gratitude to my supervisors, Dr. Pubali Chakraborty and Dr. Partha Chatterjee, for their guidance, support and encouragement over the past year. Their mentorship and motivation played a crucial role in shaping this thesis. I am especially thankful for the knowledge and insights they shared throughout the coursework and research process.

I am also deeply grateful to my faculty members at Shiv Nadar Institution of Eminence, Dr. Kurt Horner, Dr. Aditi Singh, Dr. Trishita Ray Barman and to all the members of the Macroeconomics Workshop for their valuable feedback and suggestions. Their input significantly strengthened the quality of this thesis.

A special thanks to Dr. Pintu Parui for reviewing my model and offering thoughtful suggestions that helped refine my work.

Lastly, and most importantly, I would like to thank my parents, Mummy and Papa, for their constant motivation and unwavering support throughout this journey.

Best regards,

Bishmay Barik

Contents

Abstract

This paper studies the disparities in human capital attainment and income distributions between disadvantaged minority groups and non-minority groups in India, and estimates the extent to which these gaps can be explained by cultural distance. We develop a three-period macroeconomic model to project human capital and income distributions for the next generation, based on observed parental distributions. Cultural distances between social groups are measured using total variation in cultural norms, drawing on data from the Notions of Identity Module 2.1 of the Samaj Survey Project. Human capital and income data are sourced from the Consumer Pyramids Household Survey. Our findings reveal persistent gaps in both human capital and income between the non-minorities and the disadvantaged minorities which includes Scheduled Castes, Scheduled Tribes, and Muslims. We estimate that approximately 22.15% of the gap in human capital attainment can be attributed to cultural distance, whereas cultural distance accounts for only about 7.7% of the income gap of next generations.

1 Introduction

Why do marginalized and minority communities remain trapped in educational poverty across generations? India, being a vast land that accommodates people of different castes, religions, and social groups, presents a challenge in identifying a singular reason to justify why such gaps in educational levels and educational expenditure persist. This is the case even though India has progressed towards a more integrative society compared to the one we witnessed four to five decades ago (Asher, Bhowmick, Bussolo, and Novosad 2023).

Although the concept of cultural distance has existed in the literature since it was introduced by Kogut and Singh (1988), it has gained limited traction in the field. This is mainly due to its broad definitions and nature, and particularly because of the lack of available data on cultural norms and behavior, which are essential to calculate cultural distance in India. However, with the recently released Samaj Survey Project, a sub-sample of households from the Consumer Pyramids Household Survey (CPHS) by the Centre for Monitoring Indian Economy (CMIE), we have identified 21 cultural norms through which we have attempted to measure cultural distances across social groups, including minorities and non-minorities in India.

To explore the explanatory power of cultural distances and assess how well they can explain gaps in educational levels and educational investments across social groups, we have developed a three-period overlapping generations model. In this model, we introduce cultural distances in a manner that affects parents' borrowing constraints. Our hypothesis is that as the cultural distance between two social groups increases, the disadvantaged group faces tighter borrowing constraints. This makes it more difficult for parents to invest in their children's education, and consequently, lower levels of educational attainment persist across both parental and child generations.

We have taken this model to the data, specifically from the Income Pyramids of the CPHS. Using this data, we estimate constrained borrowing and saving amounts of households, educational investments, and the resulting human capital and income distributions of children belonging to both minority and non-minority groups. In our model, parents who face borrowing constraints dependent on cultural distance invest in their children's education accordingly. We then conduct counterfactual exercises to understand the explanatory power of cultural distance in explaining the gaps in educational investment between parents of minority and non-minority groups. The results from these counterfactuals suggest that approximately 22.15% of the gap in human capital achievement distributions, or education level distributions, among children of different social groups is explained by cultural distance. However, only about 7% of the gap in income distributions across future generations, that is, children's income, is explained by cultural distances.

The next subsection provides a review of literature that has been done by other researchers in this domain. Section 2 gives a brief introduction to the caste system in India, with Section 3 focusing on an empirical exercise that examines the relative positions and rankings of social groups in terms of monthly educational investment by households. We implement a Two-Way Fixed Effects regression specification and find that it is the Other Religions (Jains, Buddhists, Sikhs, Parsis) who make the highest educational investments from 2014 to 2019, followed by the Upper and Intermediate Caste Hindus, and then by the Other Backward Classes. Among the minority groups specifically, Muslims show higher educational investments, though not as much as OBCs. They are followed by the Scheduled Tribes, and finally, the Scheduled Castes or Dalits, who exhibit the lowest educational investment among all minority groups.

Section 4 introduces the three-period model, where we set up the macroeconomic optimization problem. In Section 5, we take the model to the data, introduce the cultural distance calculation, and estimate it across social groups. Section 6 presents the counterfactual exercises, where in one case we eliminate the cultural distance between minorities and non-minorities, and in another, we equate wage rates of minorities to those of non-minorities. A subsection in Section 6 offers a brief extension to the previously done exercises, involving some changes in model parameters. Finally, Section 7 concludes the paper.

1.1 Literature Review

Caste and social inequality are longstanding concerns in the Indian context. It is the caste of an individual that largely determines the extent of access they receive. Caste also influences the kind of school an individual can attend, how they are treated by teachers, and the nature of relationships they form with classmates (Munshi 2019). A significant body of literature exists that documents the persistence of caste-based inequality in India. While some studies analyze wealth disparities using caste stratification and ANOGI decomposition techniques (Zacharias and Vakulabharanam 2011), others have examined educational disparities through an empirical lens.

Borooah and Iyer (2005) investigate the interplay between *vidya* (education), *veda* (religion), and *varna* (caste), and ask whether, and to what extent, school enrolment in India is shaped by community norms, such as those based on religion (Hindu or Muslim) or caste (Scheduled or non-Scheduled). Their findings suggest that under favorable circumstances such as when parents are literate, the influence of community norms becomes negligible. However, in less desirable conditions, these norms significantly influence enrolment rates, leading to considerable disparities.

Similarly, Borooah (2012) analyze inequalities in children's test scores within social groups. They argue that inter-group comparisons of educational attainment should account not only for average achievement levels but also for the distribution of achievement within each group. This approach helps explain why educational outcomes differ across social categories. Their findings show that Dalit, Adivasi, and Muslim children are severely disadvantaged in reading, writing, and arithmetic compared to Brahmin children.

Our study contributes to this literature by analyzing household educational expenditure patterns using recent data from 2014 to 2019. We explore, both empirically and theoretically from a macroeconomic perspective, how investment in education varies by social group. To the best of our knowledge, there has been limited research focused on educational investments among minority social groups during this time period, especially through a macroeconomic framework, and this is the gap our work seeks to fill.

Another strand of literature investigates cultural distance between social groups in India. Asher, Bhowmick, Bussolo, Mehta, and Novosad (2024) empirically study cultural capital at a granular level within Indian villages. They find that each unit increase in cultural distance from the dominant social group is associated with 2.5 to 3.7 fewer years of schooling and a 25 percentage point decrease in the likelihood of completing primary school. As for measuring cultural distance, Raza, Singh, and Dutt (2002) proposed a methodology for its calculation, though it has only been applied to the district of Prayagraj in Uttar Pradesh.

Our contribution here lies in extending the measurement of cultural distance to a broader, national level. We compute cultural distances across social groups throughout India and assess their explanatory power in understanding educational investment disparities. We empirically examine how well these distances correlate with observed gaps in educational spending between social groups.

From a theoretical standpoint, numerous macroeconomic models have studied the intergenerational transmission of human capital using Overlapping Generations (OLG) frameworks. Lee and Seshadri (2019), for example, present a model where human capital investments occur both over the life cycle and across generations. Their results show that much of life-cycle inequality is determined early in life, and is largely driven by differences in parental background. Other foundational studies in this area include Doepke and Tertilt (2016) and De La Croix and Michel (2002), both of which employ OLG models to illustrate how initial differences in parental human capital result in persistent disparities in subsequent generations.

In the Indian context, Goraya (2023) explore caste-based disparities in entrepreneurship, emphasizing the role of credit market constraints. However, to our knowledge, we are among the first to incorporate cultural distance directly into borrowing constraints. By allowing borrowing ability to vary across social groups, we examine how these constraints shape parental investments in education, particularly for historically disadvantaged communities.

2 Context and Background

2.1 Caste in India

The caste system in India is a centuries-old hierarchical social structure that stratifies people by birth, historically determining their occupation, status, and access to resources. Beteille (1965) explains that caste is *"characterised by endogamy, hereditary membership, and a specific style of life which sometimes includes the tradition of a particular occupation and is usually associated with a more or less distinct ritual status in a hierarchical system."* There has been a substantial body of work on caste in India, since it plays a role at every stage of an individual's economic life - in school, university, the labor market, and even into old age (Munshi 2019). Essentially, caste operates on two levels: *varna* and *jati*. The *varna* system categorizes people into four broad groups: *Brahmins* (priests), *Kshatriyas* (warriors and rulers), *Vaishyas* (merchants, traders, and farmers), and *Shudras* (manual and labor services, lower-level work), with its origins in ancient Hindu scriptures. In contrast, *jati* refers to one's endogamous community, often translated as caste, tribe, or community (Asher, Bhowmick, Bussolo, Mehta, and Novosad 2024). Among these, the Scheduled Castes (SCs), or also referred to as Dalits, were historically subjected to untouchability, and till date they face particular stigmatization (Krishnamurthi and Krishnaswami 2020). Another major minority group in India is the Scheduled Tribes (STs), or also referred to as *Adivasis*, with the term 'tribe' adopted by the communities themselves to represent the dispossessed and deprived people of a region (Xaxa 1999).

Discrimination against these caste groups, especially the backward and socially disadvantaged castes, is deeply rooted and well-documented in Indian literature. There is significant inequality in educational institutions (Hanna and Linden 2009), job markets (Banerjee and Knight 1985), and even in marriage markets (Ahuja and Ostermann 2016). However, this paper specifically focuses on education and educational expenditures, comparing the spending patterns of SCs, STs, and Muslims with those of non-minorities. It aims to identify the possible reasons behind the observed disparities in educational investment among these groups. Understanding these gaps is crucial, as education remains one of the most important vehicles for upward mobility in Indian society.

3 Empirical Motivation

The minority groups in India, particularly SCs, STs, and Muslims, have been in a disadvantaged position for a very long time in history. The Scheduled Castes have been severely oppressed because of caste categorization, mainly due to untouchability and the historical oppression of Dalits (then referred to as "Depressed Classes"), which are deeply embedded in Hindu religious and social structures (Ambedkar 1936). It is primarily the employment patterns of these untouchables that led to their constitutional recognition as Scheduled Castes. Due to this untouchability, SCs have historically lacked access to education in India (Thorat and Joshi 2020). Similarly, the marginalization of certain tribal communities, and their exclusion from mainstream society, led to the formation of the Scheduled Tribes category. Being isolated from mainstream society, they have always remained distant from educational institutions and have had limited exposure to the mainstream educated society (Baviskar 1999).

On the other hand, Muslims as a religious group have faced backlashes not only in terms of education but also across various other dimensions. After the 1857 revolt, Muslims were particularly targeted by the British due to their perceived role in the uprising. Their educational and economic status declined significantly during the colonial period. Post-independence, the partition led to mass displacement and violence, further marginalizing many Muslim communities. Since then, Muslims have remained underrepresented in education, government jobs, and across economic indicators (Sachar Committee Report 2006). Altogether, these three groups - SCs, STs, and Muslims - constitute the primary minority groups in India that have faced severe historical and contemporary disadvantages, especially in the field of education.

However, the problems stated above are more on the supply side—barriers and disadvantages faced by the minorities in accessing education. As this study extends its focus to understand the patterns of educational expenditure across different caste and religious households, we find persistent gaps and level differences in educational spending among social groups in India. Figure 1 illustrates how various social groups and households in India have spent on children's education between 2014 and 2019. The data used to create this figure comes from the Consumption Pyramids Household Survey by the Centre for Monitoring Indian Economy. The figure shows the share of household income spent on education across Other Hindus (which includes Upper and Intermediate Caste Hindus), OBCs (as non-minority representatives), and SCs, STs, and Muslims (as minority groups). It is evident that Other Hindus have consistently been at the top in terms of income share spent on education. This implies that historically, Other Hindu households have allocated the highest share of their income towards children's education. They are followed by OBCs, who belong to the non-minority category. Among the minority households, persistent gaps existed between SCs, STs, and Muslims until about the third quarter of 2018. Among these, Muslims have performed better than other minority groups, followed by SCs, while STs remain the most disadvantaged.

To strengthen these empirical observations, we run a series of regressions using the same data from 2014 to 2019, to investigate differences in educational expenditure across social groups in India. The empirical specification is as follows:

$$y_{hgdt} = \alpha + \sum_{g \in \mathcal{G} \setminus \{g_0\}} \beta_g \mathcal{D}_h^{(g)} + \gamma_1 \,\text{HH Income}_{hdt} + \gamma_2 \,\text{HH Size}_{hdt} + \gamma_3 \,\text{Num. Children}_{hdt} + \lambda_t + \mu_d + \varepsilon_{hgdt} \tag{1}$$

where the dependent variable is the educational expenditure by household h on children belonging to a social group g, in district d, during month t. The specification aims to identify differences in educational expenditure across caste groups. On the right-hand side, $\mathcal{D}_h^{(g)}$ represents a dummy indicator for each social group. Since the CPHS dataset lacks consistent information on the social group of all individuals in a household, the variable $\mathcal{D}_h^{(g)}$ is based on the social group of the household head only. For each social group $g \in \mathcal{G}$, if the household belongs to group g, then $\mathcal{D}_h^{(g)} = 1$, otherwise it is 0. The terms λ_t and μ_d represent time fixed effects and district fixed effects, respectively. The reference group g_0 in this specification is the 'Other Hindus' group, and all comparisons are made relative to this group. Additional controls include HH Income$_{hdt}$, which is the total household income; HH Size$_{hdt}$, the number of household members excluding children below 18 years; and Num. Children$_{hdt}$, the number of children under 18 in the household.

The regression results are presented in Table 1. Column (1) shows the regression of total educational expenditure on only the social group dummies. With Other Hindus as the reference, households categorized under Other Religions (which include Christians, Upper Caste Buddhists, Jains, Sikhs, etc.) spend more on children's education. OBC households spend slightly less than Other Hindus but remain the lowest among non-minority groups. Among minority groups, ST households are in the most disadvantaged position during 2014–2019, with the lowest educational expenditure per month on children. Column (2) adds household income as a control; while the absolute values of the coefficients decrease, the relative ranking of social groups remains unchanged. Column (3) further includes household size as a control, again preserving the ranking. When the number of children is added in Column (4), there is a switch in the relative positions of Muslims and SCs. The updated ranking now shows Other Religions at the top, followed by Other Hindus, then OBCs, SCs, Muslims, and finally STs.

Initially, SC households appeared to spend more on education than Muslim households. However, after accounting for the number of children, Muslim households show higher educational spending. This shift suggests that Muslim households, on average, have more children than SC households. In earlier specifications, the total educational expenditure was diluted across more children in Muslim households, which masked their higher spending tendency. By controlling for the number of children, we standardize for family size, revealing that Muslim households allocate more resources per child to education than SC households, given similar income and household composition. Column (5) includes time fixed effects, where the relative ranking of social groups remains consistent with Column (4). However, when district fixed effects are added in Column (6), the most disadvantaged group is the SCs, which is the Dalit category. Therefore, the overall ranking mostly remains the same, with SC being the group with least educational investments in children. There is no single universal reason why Scheduled Castes (SCs) invest the least in their children's education. However, one possible explanation is the persistent social stigma associated with SCs, rooted in their historical classification as "untouchables." This stigma may have led to limited exposure to educational institutions due to supply-side discrimination (Kumar 2011). Nevertheless, this paper does not explore supply-side discrimination, as it falls outside the scope of our analysis.

The broader picture we aim to capture is the persistent gap in educational expenditure between minority and non-minority social groups. Our primary interest lies in investigating the underlying reasons for this gap and quantifying how much of it can be explained by the cultural norms and parameters specific to each social group. The next section introduces a theoretical model to explore the intergenerational dynamics and how cultural norms may influence educational levels and spending patterns across different social groups.

4 Model

This section consists of the lifecycle model comprising of households with one parent and one child. The model consists of three stages, and the stages are: the Childhood Period (Stage I), during which individuals acquire education and other foundational skills; the Working Life (Stage II), when individuals participate in the labor market, earn income, and make investment decisions; and the Retirement Period (Stage III), wherein individuals exit the labor force and rely on pensions or any retirement stream of income.

In this model, it is the individual in the working stage who makes the educational investment decisions for the child. To keep the model simple, we assume that the child's consumption is included in the working stage parent's overall consumption. During the working stage, the individual receives a positive stream of income. In the retirement stage, the individual is no longer part of the active labour force. However, they continue to receive a stream of income during this stage — for example, through pensions.

4.1 Preferences

The individuals follow a Constant Relative Risk Aversion (CRRA) utility function, which is given by:

$$U(c_0, c_1, h_1) \quad = \quad \frac{c_0^{1-\sigma}}{1-\sigma} + \beta \frac{c_1^{1-\sigma}}{1-\sigma} + \gamma \frac{h_1^{1-\rho}}{1-\rho} \tag{2}$$

where c_0 denotes the consumption level of the working-period individual (parent), c_1 denotes the consumption level of the parent when retired, and h_1 denotes the human capital level of the child of the working-period parent. Each of these variables is assumed to vary with the group the specific individual, or the parent is affiliated to. The parameter β is the discount factor, where $\beta \in (0,1)$, and a higher value of β implies a higher weight given to future consumption. The parameter γ represents the weight on the utility derived from the child's human capital, capturing altruism or parental preference for child outcomes.

The human capital accumulation of the child follows the process given below:

$$h_1 = (Ae_0)^{\theta} h_0 \tag{3}$$

Here, the human capital accumulation process takes a Cobb-Douglas form. The term e_0 denotes the educational expenditure made by the parent towards the child, and h_0 denotes the exogenously given human capital endowment of the parent. It is assumed that each individual belonging to social group i is endowed with human capital $h_0 > 0$. This initial human capital may vary across social groups and is distributed according to a probability distribution function $\Gamma_0^{i,h}(\cdot)$, where the distribution $\Gamma_0^{i,h}$ differs for each group i.

4.2 Working Stage

Individuals in the working period face a linear budget constraint, given by:

$$y_0 + B_0 \geq c_0 + e_0 \tag{4}$$

where y_0 is the stream of income received during the working period by an individual. Additionally, the working period income depends on the individual's level of human capital, and hence can be expressed as:

$$y_0 = w^i h_0 \tag{5}$$

where w^i is the group-specific wage rate, given exogenously. Given this setup, even y_0 follows a probability distribution function, which differs by social group i. This probability distribution function can then be considered as a scaled distribution, given by $\Gamma_0^{i,y}(\cdot)$. In this model, it is assumed that the individual borrows to fund the child's education; therefore, the borrowing amount is denoted by the term B_0.

Individuals belonging to specific groups face a borrowing constraint, which is given by:

$$B_0 \leq (1 - D_i)y_0 \tag{6}$$

where D_i denotes the cultural distance of the social group from the culturally dominant group in the economy (which also serves as the reference group), and $D_i \in (0,1)$. A value of $D_i = 0$ implies no cultural distance between the disadvantaged and the dominant group, whereas an increase in D_i reflects a greater distance between the two.

The borrowing constraint is modeled in a way that intuitively captures the limitations faced by individuals from disadvantaged caste groups. The idea is that if the locally dominant social group can borrow an amount x, then the disadvantaged group, due to their cultural distance, can only borrow $(1 - D_i)x$ units. This setup is inspired from the empirical observation, as done by Anderson (2005). The constraint also depends on the individual's income. Individuals employed in high-skilled jobs, who typically earn higher incomes, are able to borrow more than those employed in low-skilled sectors. This implies, as individuals in low-skilled occupations tend to face a higher risk of default, and hence their credit limits are lower compared to those in high-skilled employment.

Another important implication of this formulation is that caste groups that are culturally the furthest from the locally dominant group may face severe discrimination. As a result, their access to formal credit markets may be restricted, further limiting their borrowing capacity relative to non-disadvantaged social groups in the economy (Kumar 2011).

4.3 Retirement Stage

In this model, the retirement period constraint is given by the following equation:

$$R_1 \geq c_1 + (1+r)B_0 \tag{7}$$

It is assumed that the individual receives a retirement stream of income (e.g., pensions), denoted by R_1. To make the model realistic, the retirement income depends on the type of job the individual held during their working life. For instance, if the individual was employed in a skill-intensive job, they receive a higher pension; on the other hand, if they were not employed in such a job, the retirement income is comparatively lower. Keeping this in mind, we can define the retirement income as:

$$R_1 = \psi y_0 \quad \text{where} \quad \psi \in (0,1) \tag{8}$$

The parameter ψ indicates that the individual receives a fraction of their working life income as a retirement stream, which is less than what they earned during the working period. With this setup of the retirement stream of income, and that being dependent on the income from working stage, even R_1 follows a scaled distribution which differs with social group i, given by $\Gamma_0^{i,R}(\cdot)$. Additionally, the amount borrowed by the individual to fund education is now repaid during retirement, at an exogenously determined market interest rate r.

4.4 The Optimization Problem

With all the components established above, we can now represent the problem faced by the working-age individual using a value function formulation.

Let $V^i(h_0)$ denote the value function of an individual belonging to group i, with a human capital endowment h_0. The individual's problem is to choose consumption in both periods, educational expenditure on the child, and the borrowing amount to maximize lifetime utility:

$$V^i(h_0) = \max_{c_0,c_1,e_0,B_0} \left\{ \frac{c_0^{1-\sigma}}{1-\sigma} + \beta \frac{c_1^{1-\sigma}}{1-\sigma} + \gamma \frac{h_1^{1-\rho}}{1-\rho} \right\} \tag{9}$$

subject to the following constraints:

$$c_0 + e_0 \le y_0 + B_0 \quad \text{(Working period budget constraint)} \tag{10}$$

$$c_1 + (1+r)B_0 \le R_1 \quad \text{(Retirement period budget constraint)} \tag{11}$$

$$y_0 = w^i h_0 \quad \text{(Working period income)} \tag{12}$$

$$R_1^i = \psi y_0 \quad \text{(Retirement income)} \tag{13}$$

$$h_1 = (Ae_0)^\theta h_0 \quad \text{(Child's human capital accumulation)} \tag{14}$$

$$B_0 \le (1 - D_i) y_0 \quad \text{(Borrowing constraint)} \tag{15}$$

Here, the individual chooses how much to consume during the working period (c_0), how much to consume during retirement (c_1), how much to invest in the child's education (e_0), and how much to borrow (B_0), given their human capital endowment h_0, group-specific wage rate w^i, interest rate r, and the retirement return parameter ψ. Each of the variables in this model may vary with every social group i.

4.5 Solving the Model

This section presents the theoretical results derived from solving the model. The detailed mathematical derivations are provided in Appendix [12]. We analyze two distinct cases based on whether the borrowing constraint binds: the non-binding (unconstrained) case and the binding (constrained) case, where cultural distance plays a critical role.

4.5.1 Non-Binding Case of Borrowing Constraint

In the non-binding scenario, the borrowing amount is strictly less than the upper borrowing limit, allowing households to optimize intertemporally without restriction. This represents an unconstrained optimization problem. The first-order conditions yield the following key equations:

$$(w^i h_0 + B_0 - e_0)^{-\sigma} = \gamma\theta A^{(1-\rho)\theta}(e_0)^{(1-\rho)\theta-1}(h_0)^{-\rho} \tag{16}$$

$$(w^i h_0 + B_0 - e_0)^{-\sigma} = \beta(1+r)(\psi y_0 - (1+r)B_0)^{-\sigma} \tag{17}$$

Equation (16) corresponds to the optimality condition for educational investment, equating marginal utility loss from current consumption to marginal gain from future human capital. Equation (17) is the standard Euler equation, capturing intertemporal consumption trade-offs. Together, these two non-linear equations characterize the interior solution and can be solved numerically for the two unknowns (B_0 and e_0) using the `fsolve` function from Python's `scipy.optimize` library.

4.5.2 Binding Case of Borrowing Constraint

When the borrowing constraint binds, the household is forced to borrow exactly at the limit, given by $B_0 = (1 - D_i)y_0$, where D_i captures the cultural distance of group i from the dominant group. This transforms the problem into one of constrained optimization. Substituting this borrowing limit into Equation (17), we obtain a closed-form solution for optimal educational investment:

$$e_0^i = w^i h_0 + (1 - D_i)w^i h_0 - \frac{\left[\psi w^i h_0 - (1+r)(1-D_i)w^i h_0\right]}{[\beta(1+r)]^{1/\sigma}} \tag{18}$$

Equation (18) shows that parental investment in children's education is a linear function of parental human capital and is inversely related to cultural distance. Differentiating with respect to D_i yields:

$$\frac{\partial e_0}{\partial D_i} = -\left[w^i h_0 + \frac{(1+r)w^i h_0}{[\beta(1+r)]^{1/\sigma}} \right] \quad < \quad 0 \tag{19}$$

The negative sign in Equation (19) shows that as cultural distance increases, parents tend to invest less in their children's education. This happens because households that are culturally more distant from the dominant group often face tighter credit constraints. Cultural distance in this case captures deeper barriers like caste-based discrimination, weaker community networks, or limited access to formal credit and educational institutions. These challenges make it harder for such households to borrow or see the full benefits of investing in education. As a result, even if parents want to invest more in their child's future, they might be unable to do so. This can lead to lower educational investments and may contribute to the persistence of poverty and inequality across generations.

5 Parameterization

In this section, we discuss our parameter choices for the model. We begin by describing the datasets used, take initial distributions from the data, estimate the wage rates in the model, and solve the model to determine educational investments and borrowing levels under both the non-binding and binding cases. We also interpret the resulting outcomes.

5.1 Data

Two datasets have been used from which parameter estimates are taken. The first dataset is the Consumption Pyramids Household Survey (CPHS) by the Centre for Monitoring Indian Economy, which provides data from 2014 onwards. The data is collected in waves, where each wave corresponds to one-third of a year: Wave 1 includes January to April, Wave 2 includes May to August, and Wave 3 covers September to December. Data is collected from households in each wave. We have used the Income Pyramids module, specifically the member-level version, which provides information on individual income from wages, the number of years of education received, caste category, and religion.

In this dataset, we classify individuals into two broad comparison groups, which align with our theoretical framework: minorities and non-minorities. Our classification considers Scheduled Castes (SCs), Scheduled Tribes (STs), and Muslims as minority groups, while Upper Caste Hindus and Other Backward Classes (OBCs) are considered non-minorities. To remain constitutionally consistent within our model, we account for the fact that, as per the Indian Constitution, only individuals from Hindu, Buddhist, and Sikh religions can be classified under the SC category. Therefore, any individual identified as both Muslim and SC in the data is categorized solely as Muslim. STs and OBCs include individuals from all religious groups.

The second dataset used is the Samaj Survey Project, specifically the Notions of Identity 2.1 module, which is a sub-sample of the CPHS data. This survey was conducted among the same households in 2014 and includes questions on cultural norms and personal opinions, allowing for an exploration of cultural attitudes. A detailed list of the questions asked is provided in Table 2. Based on these questions, we inferred the cultural attitudes of individuals and computed cultural distances across social groups, using the methodology described in Section 6.5.

5.2 Solving the Non-Linear Equations in the Non-Binding Case

As shown in Equation (16) and Equation (17) in the previous section, we solve the model using non-linear solvers in Python, specifically employing the `fsolve` function from the `scipy.optimize` library. The estimates required from the dataset include the wage rates w^i and the initial human capital distribution h_0. The initial human capital distribution is proxied by the distribution of years of education received by each individual in the survey. Accordingly, each grid point in the dataset can be assumed to correspond to a specific individual, providing us with the complete distribution of educational attainment. Hence, assigning values to each h_0 for solving the non-linear equations is straightforward.

The next step involves estimating the wage rates for the minority group (SCs, STs, and Muslims) and the non-minority group (Upper Caste Hindus and OBCs). To do this, we run a regression with the following specification:

$$Y_{it} = \beta_0 + \beta_1(\text{Education Years}_{it}) + \beta_2(\text{Education Years}_{it} \times \mathcal{D}_{it}^{\text{Min}}) + \lambda_i + \epsilon_{it} \tag{20}$$

Here, the dependent variable is the income of individual i at time t. The term $\mathcal{D}_i^{\text{Min}}$ is a dummy variable that takes the value 1 if the individual belongs to any of the minority groups. In the dataset, some individuals change their social status over time; for instance, someone may be classified as an OBC Hindu in 2014 and as a Muslim OBC in 2016. Since some individuals in the dataset changed their religion over time, we tracked these individuals and consistently assigned them the caste category they belonged to in January 2014 for all subsequent periods. Human capital accumulation is proxied by the years of education received, represented by the variable Education Years_{it}. Lastly, λ_i represents individual-level fixed effects.

This regression setup is motivated by the relationship between an individual's income and their human capital accumulation, as described in Equation (12) in the model. In the context of the regression in Equation (20), the estimate $\hat{\beta_1}$ represents the wage rate for non-minorities, while the estimate $\hat{\beta_1} + \hat{\beta_2}$ corresponds to the wage rate for minorities.

The regression results for the specification above are presented in Table 3. From Table 3, it is evident that the estimated wage rate for non-minorities is 145.63 units, while that for minorities is 117.14 units. Using these wage estimates, the parameter values listed in Table 4, and the initial human capital distribution (proxied by the years of education), we solve the non-linear equations (Equations (16) and (17)) to determine the optimal distribution of educational expenditure made by parents on children in both the minority and non-minority communities.

5.3 Case of the Non-Binding Borrowing Constraint

Once the non-linear equations were solved, we obtained the distribution of educational investments made by parents on their children, corresponding to the given initial distribution of human capital. In addition to this, we estimate the borrowing amounts incurred by parents to finance their children's education, using the model, the parameter values, and the wage estimates derived earlier.

A tabular representation and summary statistics of the results from the CPHS dataset for minority and non-minority groups are presented in Table 5. It can be observed that the two groups follow distinct distributions of educational expenditure (e_0) and borrowing amounts (B_0). The results further indicate that across all key parameters, whether income distribution, years of education, educational investment, or borrowing amounts, it is the non-minorities that exhibit higher average values compared to minority groups.

However, these categories are relatively broad, and a more detailed categorical summary is provided in Table 6. Still, the summary statistics presented here are simple averages and may not fully capture the nuances of the underlying distributions. For a more detailed and granular visual understanding, the following subsections present distributional analysis for the non-binding borrowing constraint case.

5.3.1 Results for the Educational Investments Distributions in Non-Binding Case

The results for the values of educational investments across the minority and non-minority groups are presented in Figure 2, covering all individuals included in the CPHS dataset. We observe that, given the varying human capital distributions among parents, there appears to be a persistent gap in educational investments made by the two groups. For each additional year of education attained by a parent, both groups tend to invest more in their children's education.

However, several trends emerge from Figure 2. First, for lower initial levels of human capital, particularly around 1 to 2 years of education, we notice that as the educational attainment of parents increases, it is the non-minority parents who seem to invest more aggressively in their children's education than minority parents. A plausible explanation for this could be that non-minority parents, historically having greater access to quality educational institutions, might perceive educational investment as a more valuable and viable option. This might indicate that non-minority parents have, over time, learned to prioritize and internalize the value of education to a greater extent than their minority counterparts.

This could be linked to differences in exposure: non-minority parents may have had more consistent and positive interactions with educational systems, which could shape their attitudes toward investing in their children's education. Minority parents, on the other hand, may have historically faced barriers, ranging from discrimination to structural exclusion, that have limited their access to the benefits of education. Thus, one could argue that it is this asymmetry in exposure that drives the sharper initial increase in educational investment among non-minority parents.

A second observation is that educational investment by minority parents appears to plateau earlier than that of non-minority parents. This might suggest that minority parents either perceive lower marginal returns to investing in education. Finally, across the entire distribution of parental educational attainment, the educational investments made by non-minority parents consistently remain higher than those made by minority parents. This persistent gap supports the notion that a gap exists in how different social groups are able to choose to invest in education.

5.3.2 Results for the Borrowing Distributions in Non-Binding Case

The second variable for which we have derived values from the data, along with the available parameter estimates, is the borrowing levels of minority and non-minority group parents. Figure 3 provides a clear depiction of the distribution of borrowing levels across individuals from these two distinct social categories in India.

The observations are that for parents with less than four years of education, borrowing levels are higher among non-minority parents than among those from minority groups. One possible explanation for this pattern is that non-minority parents with low educational attainment which corresponds to lower human capital, may not be able to afford educational expenses upfront (given that income y_0 is determined by $y_0 = wh_0$, so lower h_0 implies lower y_0). However, these parents may still choose to borrow more because their prior exposure to better educational opportunities may have led them to recognize the long-term benefits of investing in education. Having internalized the value of education, non-minority parents are more willing to incur debt to finance their children's schooling, which is reflected in the observed borrowing patterns.

As parental human capital increases along the distribution, we begin to observe a shift from borrowing to negative borrowing, which intuitively represents savings. This suggests that once non-minority parents reach a certain threshold of human capital, they are no longer required to borrow to fund education. Instead, owing to their improved earning capacity (facilitated by better educational attainment), they begin to save. Notably, this transition to savings is more pronounced among non-minority parents compared to minority group parents. This implies that non-minority households not only prioritize education earlier but also accumulate enough resources over time to begin saving, whereas minority households may take longer to reach that stage due to relatively constrained opportunities and access.

5.4 Case of the Binding Borrowing Constraint

From the available CPHS dataset, as discussed in the previous section, approximately 26.77% of individuals were found to have a binding borrowing constraint. This implies that:

$$B_0 > (1 - D_i)y_0$$

Among those with a binding constraint, 50.52% were non-minorities, while 49.48% belonged to the minority category. Within the subset of minorities facing a binding constraint, 26.94% were Muslims, 55.63% were Scheduled Castes (SCs), 16.79% were Scheduled Tribes (STs), and 0.64% belonged to other religions such as Khasi, Parsi, etc.

Before delving into the results under the binding constraint scenario, and subsequently quantifying the extent to which our results are driven by the cultural distance parameter, it is essential to first understand the methodology for calculating cultural distance. The following sections present the relevant framework and explore cultural distances across different social groups in the Indian context.

5.5 Cultural Distance

While there is limited literature in economics that measures cultural distance across social groups, this paper makes an attempt to estimate it systematically. Our approach is inspired by the method proposed by (Raza, Singh, and Dutt 2002) for calculating cultural distance. In addition, we draw on the methodology used in (Asher, Bhowmick, Bussolo, Mehta, and Novosad 2024). The steps we follow in this paper are outlined below.

5.5.1 Calculating Cultural Distance

To measure cultural distance between social groups, we use responses to a set of questions related to cultural norms. For each question, we calculate how different the answers are between two groups. This allows us to see how far apart the groups are in terms of their cultural beliefs and practices.

Step 1. *Calculate Total Variation Distance (TVD) for Each Question*

For each cultural norm question, we calculate the Total Variation Distance (TVD) between the response distributions of two groups, say Group A and Group B:

$$TVD_j(A,B) = \frac{1}{2}\sum_x \Big|\Pr(x \mid \text{Group} = A) - \Pr(x \mid \text{Group} = B)\Big| \tag{21}$$

Here, x represents a possible answer to the question. $\Pr(x \mid \text{Group} = A)$ is the proportion of people in Group A who chose answer x, and similarly for Group B. The TVD measures how different the two groups' answers are for each question. A higher TVD means bigger differences.

Step 2. *Average the TVDs Across All Questions*

Once we calculate the TVD for each question, we take a simple average across all questions to get the overall cultural distance:

$$\text{Cultural Distance}(A,B) = \frac{1}{J}\sum_{j=1}^{J} TVD_j(A,B) \tag{22}$$

where J is the total number of questions. We treat all questions as equally important, so each one gets the same weight in the average.

In this way, cultural distance shows how different two groups are in their views and norms. A larger value means the groups are culturally more distant. The data, list of questions, and results will be discussed in the next subsections.

5.5.2 Cultural Distances between Social Groups

As previously discussed, the data used to compute cultural distances is sourced from the Samaj Survey Project, specifically the 'Notions of Identity 2.1' module. Following the methodology outlined earlier, we focus primarily on questions related to cultural norms. These questions, and the responses gathered from individuals, form the basis for the computation. Summary statistics of these selected cultural norm questions are presented in Table 2. The questions included in our analysis have been filtered from the larger set available in the dataset, as they directly capture aspects of cultural identity and social behavior. Examples include questions on the degree of endogamy within caste groups, dietary preferences, levels of religiosity, attitudes towards discrimination, and whether respondents would personally engage in discriminatory behavior against members of other social groups.

Using the methodology discussed in the preceding subsection, we calculate the cultural distances between different social groups. The resulting cultural distance matrix is presented in Table 7. From the results, we can see that the highest cultural distance exists between the Muslims and the STs (0.13 units). This implies that it is the Muslims and the STs whose cultural norms and behaviour are the most distinct, and hence, their cultural traditions do not overlap as much as they do across other social groups. On the other hand, the distance between the SCs and the OBCs is the smallest in terms of culture (0.06 units). This means that most of the cultural norms of SCs and OBCs overlap with each other. However, for our purpose, since we are looking at the minority and non-minority groups at a broader level, the calculated cultural distance between the minorities and non-minorities is found to be 0.0905 units.

5.5.3 Results for the Educational Investments Distributions in Binding Case

Once the individuals whose borrowing constraint was binding were identified and filtered out from the dataset, their distributions of educational investments against their human capital levels are presented in Figure 4. As already established in Equation (18), e_0 is linear in h_0, which is clearly reflected in Figure 4. That is, the educational investments made by parents with a binding borrowing constraint increase linearly with their human capital levels. The results align with expectations that at zero years of education, parents do not earn any income and thus do not spend on their child's education.

As parents attain marginal increases in education, they begin to invest in their children's education. Similar to the non-binding case, we observe that although both investment curves are linear, the slope for non-minorities is steeper than that of minorities. This implies that non-minority parents tend to value educational returns more and, as a result, allocate more resources towards their children's education compared to minority parents. While minority parents also begin to invest once they gain some education, the rate at which they invest is slower than that of non-minorities. This results in an increasing gap between the two linear curves, representing educational investments as a function of parental human capital among those with binding borrowing constraints.

5.5.4 Results for Future Human Capital and Income Distributions in Binding Case

Having examined the educational investments by parents with binding borrowing constraints, we now turn to the resulting human capital distribution of their children. Figure 5 visually represents how children's human capital varies with the parental human capital distribution in such cases. As shown in Figure 4, educational investment increases with parental human capital. Given the human capital accumulation process defined in Equation (14), where children's human capital is a positive function of parental investment, it follows that children's human capital rises as educational investments increase.

The non-linearity in this model arises from the variation in parental human capital levels. Nevertheless, a clear upward trend is evident—more educated parents are more likely to invest in their children's education.

Since non-minority parents have better recognized the value of education, their higher investments lead to higher human capital accumulation in their children compared to minority parents. A similar pattern is observed in children's future income distributions, depicted in Figure 6. As expected, income also exhibits an upward trend and a widening gap between minority and non-minority groups. This is consistent with prior explanations, that non-minority parents recognize the returns to education and hence aim to equip their children with higher human capital. Consequently, these children are more likely to acquire skills suited for skill-intensive sectors, obtain formal employment, and earn higher incomes than those with lower levels of education or human capital.

Moving forward, it becomes important to examine the extent to which different parameters contribute to this observed gap. In particular, we assess the role of cultural distance in explaining disparities between minority and non-minority groups. The next section presents counterfactual exercises to explore this in greater detail.

6 Counterfactual Analysis

In this section, we explore and understand how much of the gaps seen in Figure 5 and Figure 6 are explained by cultural distances. If not cultural distance, then what is it that makes a gap exist between the minorities and the non-minorities? For this, we perform a counterfactual exercise. We take different cases and modify some parameter or the other to check: if characteristics between the minorities and the non-minorities are the same, then what differences would we observe in the distributions of future generation human capital and future generation income? Throughout this counterfactual exercise, we assume that it is the minorities who are becoming like the non-minorities, whether that be in terms of cultural distances, or even wage rates. Table 8 gives a clear description and an idea of what the counterfactual exercises are going to be. In the first counterfactual exercise, we eliminate the cultural distance between the minorities and the non-minorities.

In the second counterfactual exercise, we make the wage rates of the minorities equal to the wage rates of the non-minorities and report the results for both the counterfactual exercises.

6.1 Counterfactual #1: Eliminate Cultural Distance between Minorities and the Non-Minorities

In the first counterfactual exercise, we eliminate the cultural distance between the non-minorities and the minorities. Throughout our analysis till now, we have considered the non-minorities as the reference group, and hence their cultural distance was by default zero, while the initial cultural distance for the minorities was considered to be 0.0905. Now, since this is a counterfactual exercise, we eliminate the cultural distance between the minorities and the non-minorities and set the cultural distance for minorities to be the same as that for non-minorities, i.e., 0. The results for this exercise can be seen in Figure 7, with average values of h_1 and y_1 for each education year is available in Table 9. The counterfactual results suggests that the trends for the minority group in the counterfactual are similar to those in the benchmark case. One clear observation is that, as the cultural distance between minorities and non-minorities is eliminated, the distribution of future generation human capital for the minorities moves closer to that of the non-minorities. This implies that the gap between the minorities and the non-minorities shrinks. However, out of the entire difference, on average, for binding parents with only one year of education exposure, cultural distance explains only about 22.14% of the distance between the minorities and the non-minorities in terms of future generation human capital. As we proceed further, and even for two and three years of education for binding parents, the explanatory power of cultural distance in the gap between minorities and non-minorities remains around 22.15% in terms of future generation human capital h_1. For future generation income, as seen in Figure 8, cultural distance explains even less, accounting for only about 7.79% of the entire gap in future generation income y_1, across all years of education of the parents.

6.2 Counterfactual #2: Wage Rate of Minorities is Equal to the Wage Rate of Non-Minorities

In this subsection too, the comparisons are made with the benchmark case, as described in Table 8. In this counterfactual, the wage rates of the minorities are made equal to the wage rates of the non-minorities. We had obtained the initial wage rates from the regression results of Equation (20); however, for this counterfactual exercise, the wage rates for minorities are set equal to those of the non-minorities, i.e., 145.63 units for both groups, as seen in Table 8. The results for this counterfactual are available in Table 9, and a visual representation can be seen in Figure 9. The results suggest that, given the initial distribution of parental human capital, around 74.30% of the gap in future human capital distributions can be explained by differences in wage rates across all education levels. However, the results for the future generation income in this counterfactual exercise are significantly different from what we observed in the first counterfactual exercise. As shown in Figure 10, the results suggest that, on average, around 88.56% of the gap in future generation income, across all education levels, can be explained by the differences in wage rates between minorities and non-minorities. One way to interpret this is: if the wage rates of minorities had been the same as those of the non-minorities, they would have been 88.56% closer to the non-minority group in terms of future generation income.

From these two counterfactuals, it is clear that it is mostly the differences in the initial wages for the two distinct groups that create the gap in future generation human capital (h_1) and future generation income (y_1). Although cultural distance does contribute to the creation of gaps in these variables, its explanatory power is relatively limited compared to that of wage differences.

6.3 Reducing the Discount Factor to $\beta = 0.92$

The next part of this paper deals with conducting an exercise with slight changes to the parameter values of the model. As per Table 4, the value of the discounting factor, or the weight assigned to future consumption, is now reduced. Intuitively, this means that the individual becomes more impatient, valuing current consumption more than future consumption. Since this is an intertemporal choice model, reducing β simply means that we are making the individual more short-termist, preferring immediate consumption happiness over long-term benefits. Hence, from now on, in this section, we consider the value of $\beta = 0.92$ instead of 0.96, and proceed with the analysis.

A preliminary results for the distribution of future generation human capital, and the future generation income is present in Table 10. Comparing it to the previous case where the value of $\beta = 0.96$, we see that as the parent becomes more impatient, then the value of h_1, which is the future generation human capital distribution at each of the education years of parents increases. To be more specific, for a 4.17% fall in the discount factor β, the future generation human capital has increased by almost 0.32% for all education years of the non-minority parents with binding constraint. On the other hand, with this fall in discount factor β by 4.17%, the future generation human capital has only increased by 0.28%, which is slightly lower than how much it has increased for the non-minority future generation.

However in terms of income of the future generation, the income for the non-minority future generation has increased by almost 0.31% for all education years h_0, but again, for the minorities, it has not increased much, because the income for future generation has increased only by 0.28%, which is again not as much as for the non-minority groups.

7 Conclusion

To summarize, this paper develops a three-period overlapping generations model that incorporates cultural distance into the borrowing constraints faced by individuals during their working years. Our findings indicate that greater cultural distance significantly hinders the borrowing capacity of culturally marginalized or disadvantaged social groups, thereby limiting their ability to invest in their children's education. Through simulations estimating the distribution of human capital and income among the children of these individuals, we find that approximately 22.14% of the gap in human capital between minorities and non-minorities can be attributed to cultural distance. In effect, if cultural distance were absent, minority groups would be 22.14% closer to non-minorities in terms of human capital attainment and 7.7% closer in terms of the next generation's income distribution.We also find that wage gaps between minorities and non-minorities play an even more substantial role. Holding the initial distribution of parental human capital constant, approximately 74.30% of the future gap in human capital can be attributed to differences in wage rates across education levels. In contrast to the effects of cultural distance, wage disparities account for an even larger share of about 88.5% of the gap in income distribution across generations. Thus, while cultural distance contributes meaningfully to disparities in human capital and income outcomes, wage inequality emerges as the primary driver of intergenerational differences in distributions of human capital and income.

A relevant policy implication of these findings is the continued and enhanced use of affirmative action, such as reservations for disadvantaged minority groups. While India has a long-standing history of implementing reservations, the persistent stigmatization of certain caste groups establishes cultural separation. Bridging these cultural divides will be a gradual and long-term endeavor. In the short to medium term, policy efforts should focus on reducing wage disparities between minorities and non-minorities. This can be achieved through educational programs aimed at empowering culturally disadvantaged individuals and integrating them into skill-intensive sectors. Such efforts would help reduce gaps in both human capital and wages, which together account for nearly 77% of the intergenerational disparities in human capital and income. Nonetheless, addressing the deeper challenge of cultural distance rooted in historical and social stigmatization remains a critical and enduring task for Indian society.

Ultimately, this paper addresses a stark truth: we cannot fully understand economic inequality in India without looking at the role of cultural distance. As long as social stigma and cultural exclusion exist, they will continue to weaken the power of education and pass inequality from one generation to the next. Closing this cultural gap is not just the right thing to do but it is also needed for economic progress. Only by breaking down these deep-rooted social barriers can we help marginalized communities reach their full potential and build a fairer and more equal society.

8 Tables

Table 1: Determinants of Household Education Expenditure

Dependent Variable: Total Educational Expenditure						
	(1) No Controls	(2) + HH Income	(3) + HH Size	(4) + Children	(5) + Time FE	(6) + Time & District FE
Social Groups (Omitted: Other Hindus)						
SC	-359.298***	-178.188***	-195.461***	-247.125***	-249.444***	-287.405***
	(2.728)	(2.750)	(2.773)	(2.768)	(2.767)	(2.876)
ST	-487.510***	-287.681***	-302.717***	-359.027***	-361.585***	-268.641***
	(4.482)	(4.478)	(4.513)	(4.498)	(4.496)	(4.948)
OBC	-230.908***	-93.499***	-102.434***	-134.930***	-138.418***	-126.559***
	(2.446)	(2.454)	(2.471)	(2.463)	(2.463)	(2.656)
Muslims	-324.720***	-144.227***	-178.601***	-258.403***	-259.348***	-249.219***
	(3.618)	(3.621)	(3.670)	(3.666)	(3.663)	(3.833)
Other Religions	342.839***	201.778***	211.756***	232.190***	232.357***	124.468***
	(8.298)	(8.238)	(8.276)	(8.240)	(8.230)	(8.731)
Other Controls						
HH Income		0.022***	0.0216***	0.0234***	0.0232***	0.0204***
		(0.00006)	(0.00006)	(0.00006)	(0.00006)	(0.00006)
HH Size			23.191***	-0.381	-2.709***	0.242
			(0.318)	(0.329)	(0.333)	(0.344)
Num. Children				237.183***	240.868***	236.679***
				(0.887)	(0.890)	(0.907)
Constant	905.925***	402.620***	296.575***	202.479***	217.820***	254.877***
	(1.845)	(2.280)	(2.731)	(2.742)	(2.781)	(2.883)
Fixed Effects	None	None	None	None	Time	Time, District
Observations	8,204,539	8,204,539	8,119,071	8,119,071	8,119,071	8,119,071
R-squared	0.0036	0.0199	0.0205	0.0290	0.0315	0.0431

Standard errors in parentheses
$^{***}p < 0.01$

Table 2: Descriptive Statistics on Cultural Norms and Practices

Panel A: Cultural Norms and Beliefs (Binary/Trichotomous Responses)

Cultural Norm / Question	Don't Know / No Response (%)	No (%)	Yes (%)
Would you eat with a Scheduled Caste person at home?	2.23	60.76	30.35
Would you eat with a Scheduled Caste person at an eatery?	3.34	59.35	32.55
Have you ever eaten with a non-relative at home?	8.06	61.15	30.79
Have you ever eaten with a non-relative at an eatery?	9.43	59.28	31.29
Have you married outside your caste?	4.29	82.59	13.11
Have you married outside your caste category (e.g., SC/ST/OBC/General)?	4.72	84.80	10.49
Have you married outside your religion?	6.70	86.81	6.49
Are you a member of any caste association?	5.09	83.43	11.48
Are you a member of any religious organization?	6.46	81.94	11.60
Would you accept your child marrying outside caste?	13.13	72.04	14.83
Is marriage between different castes acceptable?	6.56	82.10	11.34
Is marriage between different religions acceptable?	6.50	85.57	7.93
Is marriage between different language groups acceptable?	6.48	84.11	9.41
Does the religion or caste of the cook matter to you?	20.49	37.12	31.51
Do you think there is discrimination against people based on religion?	19.39	37.39	32.91

Panel B: Cultural Practices and Self-Perception (Detailed Categories)

Cultural Practice / Description	Distribution of Responses (%)
How frequently do you eat out or buy food?	DK/NR: 12.03; Weekly: 26.68; Fortnightly: 14.09; Multiple Times: 19.64; Never: 27.57
How frequently do you visit your religious institution?	DK/NR: 3.25; Weekly: 46.21; Fortnightly: 9.01; Never: 2.06
Do you eat meat?	DK: 2.50; No: 22.26; Inside Home: 35.56; Outside Home: 3.78; Both: 35.90
How would you describe yourself culturally?	DK: 12.25; Very Modern: 1.28; Modern: 17.98; Traditional: 57.20; Very Traditional: 11.30
How would you describe your family culturally?	DK: 12.31; Very Modern: 1.46; Modern: 19.25; Traditional: 54.44; Very Traditional: 12.54

Note: All values are percentages and based on valid responses. Panel A includes trichotomous responses; Panel B includes frequency-based or ordinal measures.

Table 3: Effect of Education and Minority Status on Income from Wages

	Dependent Variable: Income from Wages	
Variable	**Coefficient**	**Std. Error**
Years of Education	145.6374***	2.7934
Minority × Education	-28.4981***	4.0034
Constant	3220.7300***	25.2388
Model Statistics		
Observations	11,006,644	
Individual Fixed Effects	Yes	
Number of Individuals	512,896	
R-squared (Within)	0.0003	
R-squared (Between)	0.1011	
R-squared (Overall)	0.0776	
F-statistic	1165.63	
Prob > F	0.0000	
Intra-class Correlation (ρ)	0.7106	

Notes: $^{***}p < 0.01$. The model include individual fixed effects.

Table 4: Model parameter values with brief descriptions and sources.

Parameter	Value	Description	Source
θ	0.50	Elasticity of human capital w.r.t. education	(Doepke and Tertilt 2016)
γ	3.11	Weight on child's human capital	(Lee and Seshadri 2019)
A	1.00	Education productivity	(Doepke and Tertilt 2016)
σ	1.50	Risk aversion parameter	(Chetty 2006)
ρ	1.50	Utility weight on human capital	(Cunha, Heckman, and Schennach 2010)
r	0.065	Interest rate	Data
ψ	0.50	Pension income share	Data
β	0.96	Discount factor	(Gourinchas and Parker 2002)

Table 5: Summary Statistics by Minority Status

Group / Variable	Obs. (1)	Mean (2)	Std. Dev. (3)	Min (4)	Max (5)
Non-Minority					
Income from Wages	7357211	4559.69	8512.42	0	800000
Education Years	7357211	8.33	5.45	0	20
e_0	6009726	326.91	41.80	133	372
B_0	6009726	−200.76	128.79	−523	29
Minority					
Income from Wages	3649433	3926.80	6059.45	0	400000
Education Years	3649433	5.20	5.10	0	20
e_0	2280638	236.47	36.64	103	282
B_0	2280638	−117.36	94.89	−419	22

Table 6: Summary Statistics by Social Group

Social Group / Variable	Obs.	Mean	Std. Dev.	Min	Max
	(1)	(2)	(3)	(4)	(5)
Intermediate Caste					
Income from Wages	1 085 602	4783.85	8751.03	0	650 000
Education Years	1 085 602	8.06	5.30	0	20
e_0	885 596	324.32	44.21	133	372
b_0	885 596	−190.73	124.52	−523	29
Muslim					
Income from Wages	1 075 074	3719.50	6139.21	0	400 000
Education Years	1 075 074	5.62	5.16	0	20
e_0	710 280	238.26	34.79	103	282
b_0	710 280	−121.43	95.82	−419	22
OBC					
Income from Wages	3 643 758	4216.29	7268.60	0	800 000
Education Years	3 643 758	6.94	5.32	0	20
e_0	2 731 672	319.48	44.58	133	372
b_0	2 731 672	−171.22	123.44	−523	29
Other Religion					
Income from Wages	361 561	4973.64	10 164.92	0	600 000
Education Years	361 561	9.56	5.14	0	20
e_0	319 109	327.19	43.04	103	372
b_0	319 109	−218.10	125.56	−523	29
SC					
Income from Wages	2 042 895	4060.63	5983.90	0	330 000
Education Years	2 042 895	5.21	5.08	0	20
e_0	1 276 977	236.84	35.93	103	282
b_0	1 276 977	−117.29	93.99	−419	22
ST					
Income from Wages	500 129	3824.11	6228.59	0	300 000
Education Years	500 129	4.21	4.90	0	20
e_0	270 949	229.40	43.73	103	282
b_0	270 949	−105.52	95.74	−419	22
Upper Caste					
Income from Wages	2 297 625	4924.95	9787.58	0	750 000
Education Years	2 297 625	10.45	5.03	0	20
e_0	2 095 781	336.75	34.95	133	372
b_0	2 095 781	−240.17	126.92	−523	29

Table 7: Weighted Cultural Distance Matrix (Lower Triangular)

Group	Muslim	Upper Caste	OBC	SC	ST
Muslim	**0.00**				
Upper Caste	0.08	**0.00**			
OBC	0.08	0.08	**0.00**		
SC	0.07	0.07	0.06	**0.00**	
ST	0.13	0.11	0.08	0.09	**0.00**

Table 8: Description of Counterfactual Exercises

Scenario	Minority Group	Non-Minority Group	Description
Initial (Benchmark)	$D_i = 0.0905$	$D_i = 0$	Cultural distance differs
	$w_i = 117.14$	$w_i = 145.63$	Wage rates differ
	h_0: Minority distribution	h_0: Non-minority distribution	Human capital differs
Counterfactual #1	$D_i = 0$	$D_i = 0$	Minorities have same cultural distance as non-minorities
Counterfactual #2	$w_i = 145.63$	$w_i = 145.63$	Minorities have same wage rate as non-minorities

Table 9: Summary of h_1 and y_1 Variables Across Counterfactuals and Actual Outcomes

Group	**Edu Years**	**Counterfactual #1**		**Counterfactual #2**		**Actual**	
		h_1	y_1	h_1	y_1	h_1	y_1
Non-minority	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Non-minority	1	19.4524	2878.949	19.4524	2878.949	19.4524	2878.949
Non-minority	2	55.0196	8142.896	55.0196	8142.896	55.0196	8142.896
Non-minority	3	101.0774	14959.460	101.0774	14959.460	101.0774	14959.460
Minority	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Minority	1	17.2759	2021.284	18.7335	2772.561	16.6564	1948.800
Minority	2	48.8637	5717.054	52.9864	7841.985	47.1114	5512.039
Minority	3	89.7684	10502.900	97.3422	14406.650	86.5492	10126.260

Table 10: Summary of h_1 and y_1 Variables for Actual Outcomes when $\beta = 0.92$

Group	**Edu Years**	**Actual**	
		h_1	y_1
Non-minority	0	0.0000	0.0000
Non-minority	1	19.5132	2887.955
Non-minority	2	55.1917	8168.371
Non-minority	3	101.3936	15006.250
Minority	0	0.0000	0.0000
Minority	1	16.7030	1954.253
Minority	2	47.2433	5527.462
Minority	3	86.7914	10154.600

9 Figures

Figure 1: Quarterly share of household income spent on education over time.

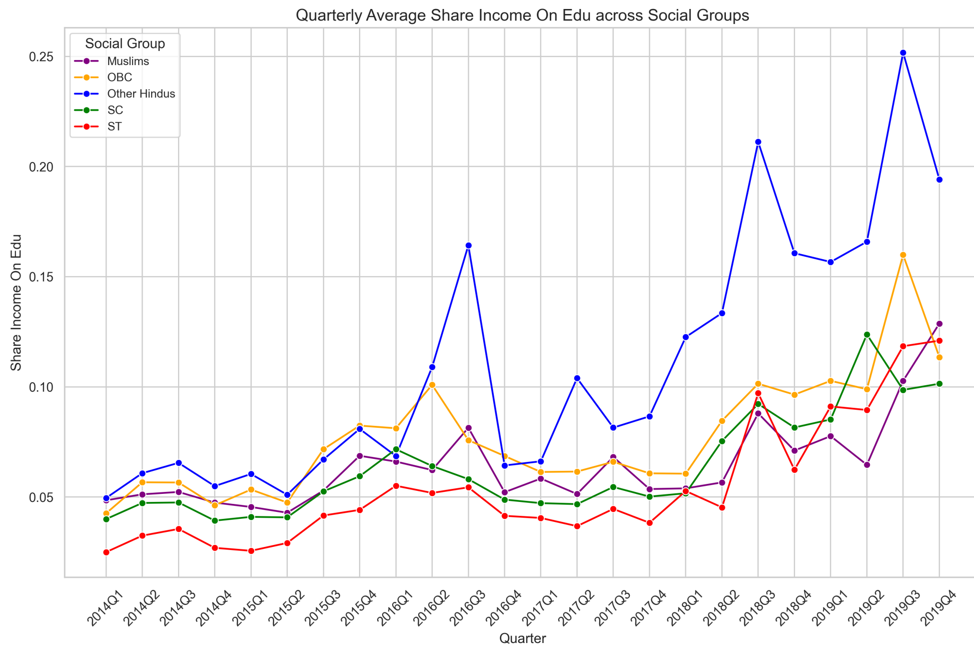


Figure 2: Initial Educational Investment by Years of Education

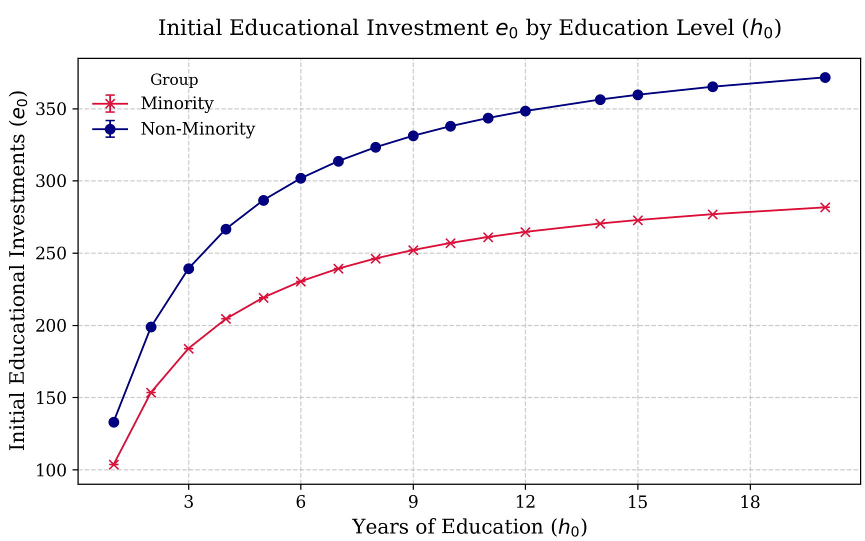


Figure 3: Initial Borrowings by Years of Education

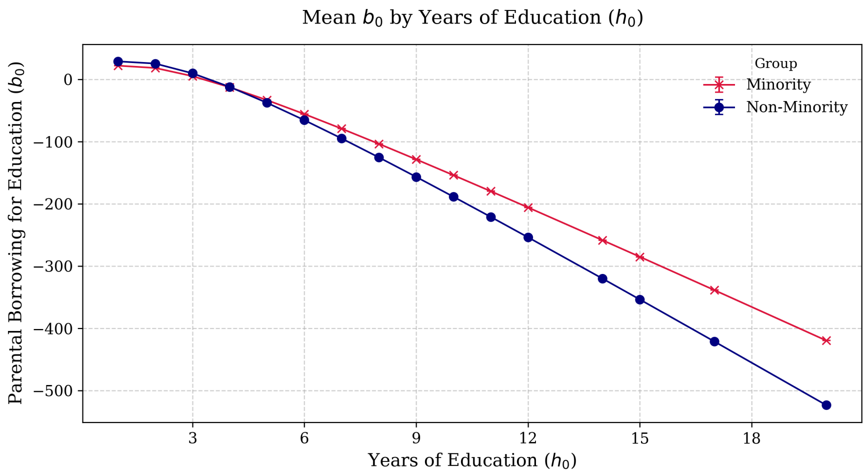


Figure 4: Educational Investment by Binding Parents by Years of Education

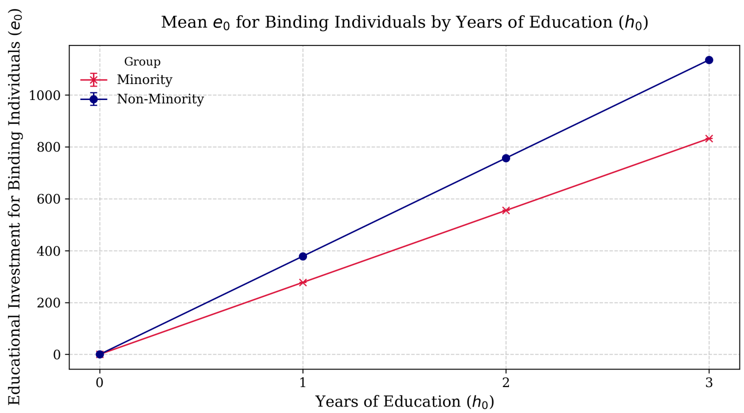


Figure 5: Children's Human Capital in next Stage by Years of Education

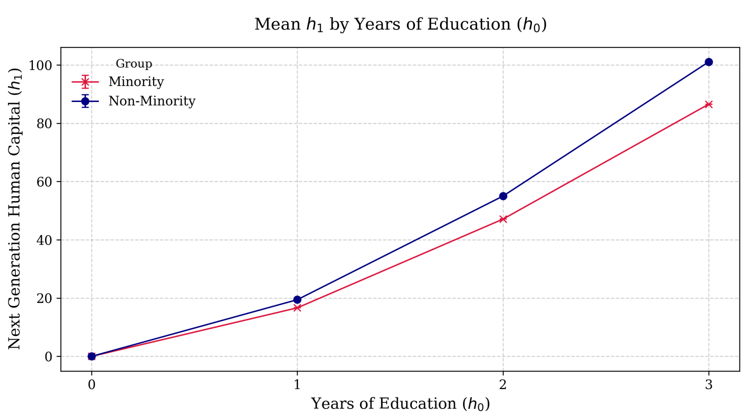


Figure 6: Children's Income in next Stage by Years of Education

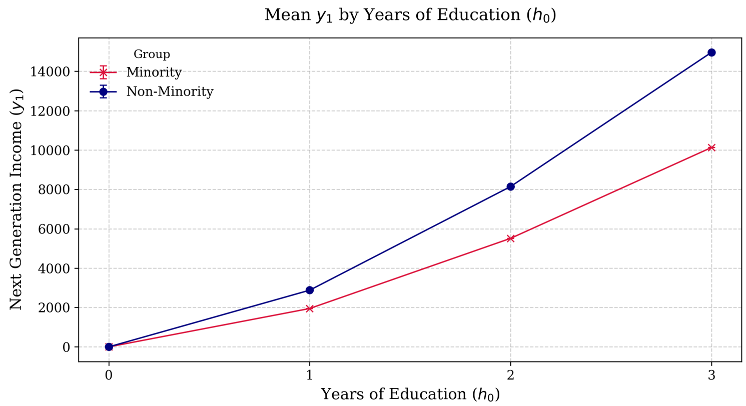


Figure 7: Comparison of h_1 under Counterfactual 1

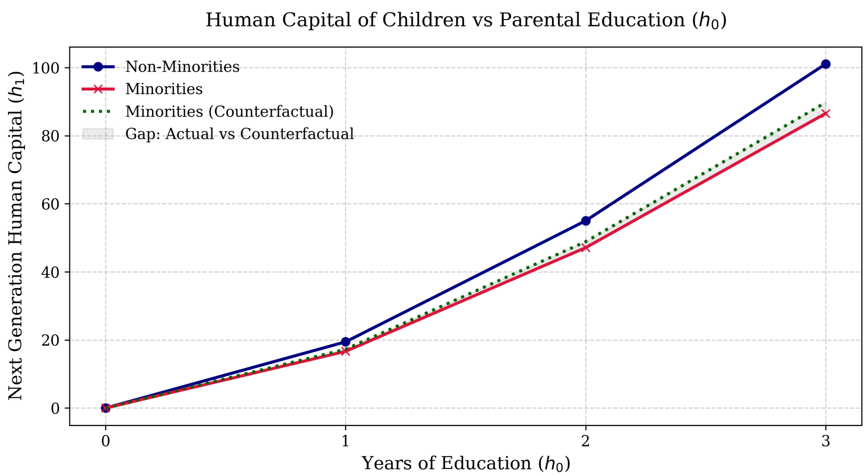


Figure 8: Comparison of y_1 under Counterfactual 1

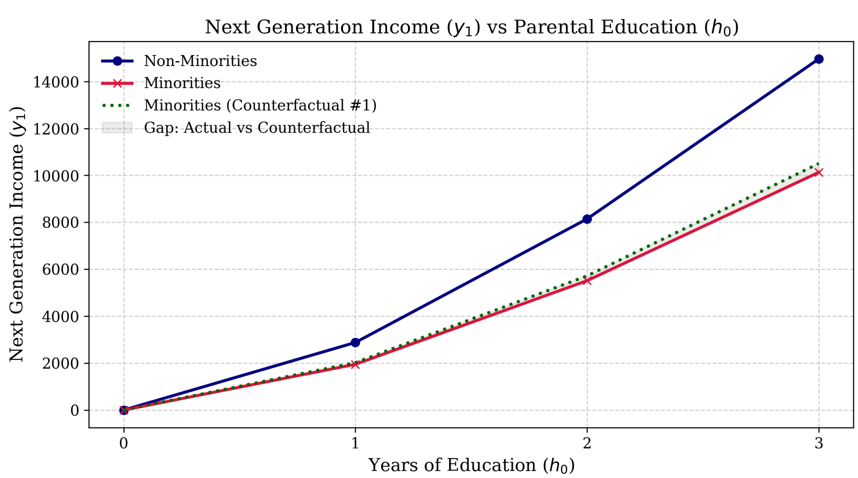


Figure 9: Comparison of h_1 under Counterfactual 2

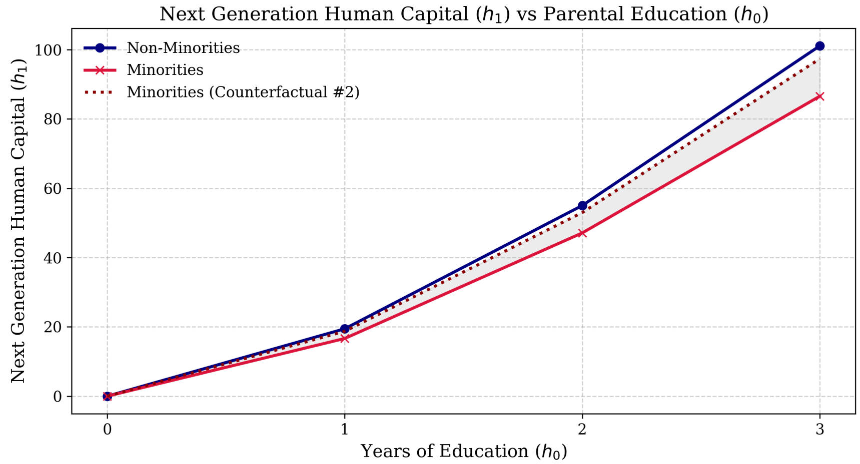
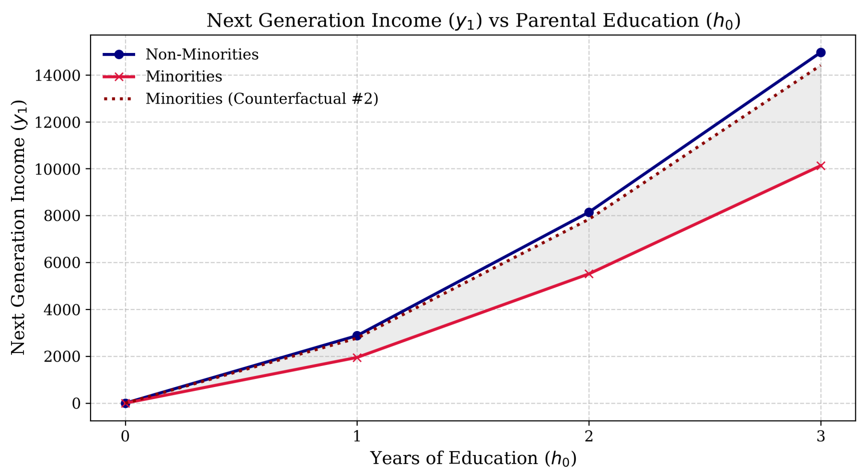


Figure 10: Comparison of y_1 under Counterfactual 2



References

AHUJA, A., AND S. L. OSTERMANN (2016): "Crossing caste boundaries in the modern Indian marriage market," *Studies in Comparative International Development*, 51, 365–387.

AMBEDKAR, B. R. (1936): *Annihilation of Caste*. Self-published, Reprinted in various editions, including 2014 annotated edition by Navayana.

ANDERSON, S. (2005): "Some economic implications of caste," .

ASHER, S., A. BHOWMICK, M. BUSSOLO, V. MEHTA, AND P. NOVOSAD (2024): "Cultural Capital and Economic Opportunity in Rural India," .

ASHER, S., A. BHOWMICK, M. BUSSOLO, AND P. NOVOSAD (2023): "Intergenerational Mobility in South Asia," Working Paper.

BANERJEE, B., AND J. B. KNIGHT (1985): "Caste discrimination in the Indian urban labour market," *Journal of development Economics*, 17(3), 277–307.

BAVISKAR, A. (1999): *In the belly of the river: tribal conflicts over development in the Narmada Valley*. Oxford University Press, New Delhi.

BETEILLE, A. (1965): "The future of the backward classes: the competing demands of status and power," *Indian Journal of Public Administration*, 11(1_suppl), 1S–39S.

BOROOAH, V. K. (2012): "Social identity and educational attainment: the role of caste and religion in explaining differences between children in India," *Journal of Development Studies*, 48(7), 887–903.

BOROOAH, V. K., AND S. IYER (2005): "Vidya, Veda, and Varna: The influence of religion and caste on education in rural India," *The Journal of Development Studies*, 41(8), 1369–1404.

CHETTY, R. (2006): "A new method of estimating risk aversion," *American Economic Review*, 96(5), 1821–1834.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78(3), 883–931.

De La Croix, D., and P. Michel (2002): *A theory of economic growth: dynamics and policy in overlapping generations*. Cambridge University Press.

Doepke, M., and M. Tertilt (2016): "Families in macroeconomics," in *Handbook of macroeconomics*, vol. 2, pp. 1789–1891. Elsevier.

Goraya, S. S. (2023): "How does caste affect entrepreneurship? Birth versus worth," *Journal of Monetary Economics*, 135, 116–133.

Gourinchas, P.-O., and J. A. Parker (2002): "Consumption over the life cycle," *Econometrica*, 70(1), 47–89.

Hanna, R., and L. Linden (2009): "Measuring discrimination in education," Discussion paper, National Bureau of Economic Research.

Kogut, B., and H. Singh (1988): "The effect of national culture on the choice of entry mode," *Journal of international business studies*, 19, 411–432.

Krishnamurthi, G., and C. Krishnaswami (2020): "Title VII and caste discrimination," *Harv. L. Rev. F.*, 134, 456.

Kumar, J. (2011): "Blocked by caste: economic discrimination in modern India, edited by Sukhdeo Thorat and Katherine S. Newman: New Delhi, Oxford University Press, 2010, 377 pp., Rs 750 (hardback), ISBN 13:978-0-19-806080-2, ISBN 10:0-19-806080-7," *South Asian History and Culture*, 2(3), 431–433.

Lee, S. Y., and A. Seshadri (2019): "On the intergenerational transmission of economic status," *Journal of Political Economy*, 127(2), 855–921.

Munshi, K. (2019): "Caste and the Indian economy," *Journal of Economic Literature*, 57(4), 781–834.

Raza, G., S. Singh, and B. Dutt (2002): "Public, Science, and Cultural Distance," *Science Communication*, 23(3), 293–309.

Sachar Committee Report (2006): "Social, Economic and Educational Status of the Muslim Community of India: A Report," Also known as the Sachar Committee Report, chaired by Justice Rajindar Sachar.

Thorat, A., and O. Joshi (2020): "The continuing practice of untouchability in India," *Economic & Political Weekly*, 55(2), 37.

Xaxa, V. (1999): "Tribes as indigenous people of India," *Economic and political weekly*, pp. 3589–3595.

Zacharias, A., and V. Vakulabharanam (2011): "Caste stratification and wealth inequality in India," *World Development*, 39(10), 1820–1833.

10 Appendix

10.1 Solving the Model

The optimization problem which we will have to solve is the following:

$$V^i(h_0) = \max_{c_0, c_1, e_0, B_0} \left\{ \frac{c_0^{1-\sigma}}{1-\sigma} + \beta \frac{c_1^{1-\sigma}}{1-\sigma} + \gamma \frac{h_1^{1-\rho}}{1-\rho} \right\}$$

subject to the following constraints:

$$\begin{aligned}
c_0 + e_0 \leq y_0 + B_0 \quad & \text{(Working period budget constraint)} \\
c_1 + (1+r)B_0 \leq R_1 \quad & \text{(Retirement period budget constraint)} \\
y_0 = w^i h_0 \quad & \text{(Working period income)} \\
R_1^i = \psi y_0 \quad & \text{(Retirement income)} \\
h_1 = (Ae_0)^\theta h_0 \quad & \text{(Child's human capital accumulation)} \\
B_0 \leq (1 - D_i) y_0 \quad & \text{(Borrowing constraint)}
\end{aligned}$$

To solve this, firstly, we consider the budget constraints of the working period and the retirement period, and rewrite them as:

$$\begin{aligned}
c_0 &= y_0 + B_0 - e_0 \\
c_1 &= R_1 - (1+r)B_0
\end{aligned}$$

Next, these rewritten budget constraints are to be replaced in the Lagrangian to maximize the value function:

$$\begin{aligned}
\mathcal{L} = \quad & \frac{(w^i h_0 + B_0 - e_0)^{1-\sigma}}{1-\sigma} \quad + \beta \frac{(\psi w^i h_0 - (1+r)B_0)^{1-\sigma}}{1-\sigma} + \gamma \frac{((Ae_0)^\theta h_0)^{1-\rho}}{1-\rho} \\
& + \mu[(1 - D_i) w^i h_0 - B_0]
\end{aligned}$$

Now, since the borrowing constraint is an inequality, because of the Kuhn-Tucker conditions, followed by the Complementary Slackness Conditions, we can say that

$$\mu[(1-D_i)w^i h_0 - B_0] = 0$$

And therefore, this can be true if either $\mu = 0$ or $[(1-D_i)w^i h_0 - B_0] = 0$. From these two cases, we can proceed for solving the unconstrained/Non-Binding case where $\mu = 0$, and the other being the constrained/Binding case where $\mu \neq 0$ but $[(1-D_i)w^i h_0 - B_0] = 0$

10.1.1 The Non-Binding Case

The Non-Binding Case, also referred to be the unconstrained case considers the total borrowing amount B_0 to exist in it's absolute value. Then, taking the first order conditions of the Lagrangian of this optimisation problem, we get the following conditions for the unconstrained case:

$$\frac{\partial \mathcal{L}}{\partial e_0} = (w^i h_0 + B_0 - e_0)^{-\sigma}(-1) + \gamma[(Ae_0)^\theta h_0]^{-\rho}\theta A^\theta e_0^{\theta-1} = 0$$
$$\frac{\partial \mathcal{L}}{\partial B_0} = (w^i h_0 + B_0 - e_0)^{-\sigma} - \beta(1+r)(\psi y_0 - (1+r)B_0)^{-\sigma} = 0$$

and rewriting these FOCs gives us two non-linear equations

$$(w^i h_0 + B_0 - e_0)^{-\sigma} = \gamma\theta A^{(1-\rho)\theta} e_0^{(1-\rho)\theta-1} h_0^{-\rho}$$
$$(wh_0 + B_0 - e_0)^{-\sigma} = \beta(1+r)(\psi y_0 - (1+r)B_0)^{-\sigma}$$

These equations have been solved on Python using `fsolve` class from the `scipy.optimize` library. Here, each grid point belongs to an individual, with different parameter values for each point. For a faster solving method, an algorithm implementing `joblibs` class from the `parallel` package on Python has been written, which can be found here. Once we solve these non-linear equations, we can then achieve the distributions for e_0 and B_0 for all individuals in the unconstrained case.

10.1.2 The Binding Case

This case is also referred to as the constrained case. Once the unconstrained case is solved for, individuals who do not satisfy the borrowing constraint are filtered out. Basically, for people whose

$$B_0 \geq (1 - D_i)w^i h_0$$

they are called binding individuals in this case. For these binding individuals, it is the case that $[(1 - D_i)w^i h_0 - B_0] = 0$, and therefore

$$B_0 = (1 - D_i)w^i h_0$$

Then, substituting this value in the Euler's Equation,

$$\begin{aligned}
u(c_0) &= \beta(1+r)u(c_1) \\
&\implies c_0^{-\sigma} = \beta(1+r)c_1^{-\sigma} \\
&\implies (wh_0 + B_0 - e_0)^{-\sigma} = \beta(1+r)(\psi y_0 - (1+r)B_0)^{-\sigma} \\
&\implies (wh_0 + (1-D_i)wh_0 - e_0)^{-\sigma} = \beta(1+r)(\psi y_0 - (1+r)((1-D_i)wh_0))^{-\sigma}
\end{aligned}$$

we get the following relationship between educational investment and cultural distance:

$$e_0 = wh_0 + (1 - D_i)wh_0 - \frac{[\psi wh_0 - (1+r)(1-D_i)wh_0]}{[\beta(1+r)]^{\frac{1}{\sigma}}}$$

And we have already seen that educational expenditure and cultural distance is negatively related, i.e., with an increase in cultural distance, the educational investment done by parents towards their children falls.