ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
PARIS SCHOOL OF ECONOMICS
(Doctoral School n° 465
Économie Panthéon Sorbonne)

&

UNIVERSITÉ DE NAMUR

PhD Thesis

For the Degree of Doctor of Philosophy in Economics

Prepared and defended on September 26, 2022 by:

Nitin Kumar BHARTI

Essays on Socio-Economic Inequality and Development in India and China: A Historical & Institutional Perspective

Thesis Advisors: Thomas PIKETTY & Guilhem CASSAN

Jury

Rapporteurs

Clément IMBERT	University of Warwick
Nishith PRAKASH	University of Connecticut

Jury

Lorenzo TRIMARCHI	University of Namur
Guilhem CASSAN	University of Namur
Thomas PIKETTY	EHESS-PSE

ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
PARIS SCHOOL OF ECONOMICS
(École Doctorale n° 465
Économie Panthéon Sorbonne)

&

UNIVERSITÉ DE NAMUR

THÈSE

pour l'obtention du grade de docteur en sciences économiques

Présentée et soutenue publiquement en septembre 26 2022 par :

Nitin Kumar BHARTI

Essais sur les inégalités socio-économiques et le développement en Inde et en Chine : une perspective historique et institutionnelle

Directeurs de thèse: Guilhem CASSAN & Thomas PIKETTY

Jury

Rapporteurs

Clément IMBERT	University of Warwick
Nishith PRAKASH	University of Connecticut

Jury

Lorenzo TRIMARCHI	University of Namur
Guilhem CASSAN	University of Namur
Thomas PIKETTY	EHESS-PSE

Acknowledgement

I would like to thank Thomas Piketty and Guilhem Cassan for their guidance throughout my PhD and for making it a really wonderful and learning experience.

Thomas has been a big source of inspiration throughout the journey. I credit him for instilling the desire to pursue a PhD. During the lengthy discussions of the master's thesis, his curiosity about the context helped me structure my thoughts and ideas, which have imprints all over this thesis. Since then, to this day, his regular guidance has helped me grow as a researcher. I have learnt many things from him; my personal best is always keeping the *big-picture* glasses on. Lastly, I have been asked by several people whether he gets enough time. Absolutely yes! Therefore, I take this opportunity to thank him for always being available and supportive.

I am also very grateful to Guilhem Cassan for his support and supervision. He provided a smooth transition and welcoming environment during my move to Namur. I have greatly benefited from his detailed research approach, which has helped in plugging the loose ends in the thesis, as well his constant reminders to strategically prioritize on these projects. Our weekly meetings kept me on my toes throughout the PhD. I was truly blessed with the complementary nature of my two supervisors!

Oliver has been a great professional and personal coach to me since 2015! His formal role over the years changed from an employer to a teacher to a thesis committee member (and hopefully co-author soon!), and at every stage, I am humbled by his support. His suggestions and comments are all over the chapters, but his contributions go much beyond this thesis.

I would also like to thank all the jury members: Clément Imbert, Nishith Prakash, and Lorenzo Trimarchi. Clément Imbert and Nishith Prakash have also been very kind to me since the very beginning when I was their RA, advising on the projects, career counselling and much more. It is a great pleasure that they accepted to be part of the jury. Lorenzo has always been there to support and give advice. The chapters have improved a lot from their detailed comments in pre-defence. I would also like to thank

Lucas Chancel for his comments as part of the thesis committee. Thanks to Sutanuka Roy and Li Yang for being great co-authors. Needless to say, they have a special role in helping me finish this thesis.

I want to thank many other professors who have helped during this journey. As part of the Economic History group at PSE, I had the opportunity to interact during classes, presentations, and one-on-one meetings! Thanks to Denis Cogneau, Gilles Postel-Vinay, Jérôme Bourdieu, Eric Monnet, Facundo Alvaredo, Lionel Kesztenbaum, Pierre-Cyrille Hautcœur. I would also like to thank Karen Macours, Thomas Breda and Andrew Clark for their suggestions on the job market paper, David Margolis (and Roxana for the support during the stressful job market) and Sylvie Lambert (as director of PhD programme at PSE). I am also grateful to all the professors at UNamur for providing an excellent research atmosphere – Jean-Philippe Platteau, John-Marie Baland, Catherine Guirkinger, Christian Kiedaisch, and Marc Sagnier. I have approached all of them at different times with project-related queries, and you have been very kind to listen and give suggestions. Special thanks to Pulak Ghosh (for providing the data and hosting me at IIM-Bangalore) and Abhijit Banerjee (for his valuable inputs since early in this process).

I was lucky to be surrounded by the intelligent colleagues at PSE and UNamur. World Inequality Lab provided a big joint family at PSE. Amory, Mark, Nacho, Thomas Blanchet, Marc, Olivia, Oscar, Tom, Carmen, Rowaida, Morten, Theressa, Clara, Matthew, Lydia... The countless after-work sessions and unforgettable télétravail (in Luberon and Latour de Carol) during COVID times. The office mates (5-68) at PSE have helped through frequent discussions and chit-chats - Melanie, Victor, Julietta, Jonathan, Santiago (miss your morning mon-ami), Yajna (for being a superb job-market coach and our spiritual discussions!), Jeanne, Sebastian! Even though many of us are in different places, I am always going to remember and cherish the memories

A small "lot" from the APE Masters' gang accompanied me in the PhD programme - Alvaro, Andrea, Kentaro, and Matthias Endres, who supported me emotionally throughout the process. The journey would not have been the same without all your love and support! Thank you for taking the painful task of proofreading! I am also grateful to the study group from the Masters – Mario, Olivia, Love, Helene, Ekaterina, and Leon - and I am sure our paths will cross at some point. I was fortunate to be part of this international group! And not to forget the little Indian club in PSE- Pranav, Michelle, Aru, Juni, and Srini experiencing a similar cultural shock coming to Paris and helping each other! There were so many other PSE mates cutting across cohorts with whom I had interacted during coffee breaks, ping-pong, foosball, drinks sessions - all of which kept me in good spirit during these PhD years - Florencia, Paolo, Zhexun (spell-checked!), Enrichetta, Nachita, Rosanne, Manpreet... (it can never be complete!)

Before I close the Paris chapter, special mention to the amazing long-lasting friends from India house (where I stayed in Paris). Atulya (being a good host during all my later Paris visits), Ravi (youtube partner), Vikash (morning chai sessions), Madhukar (innumerable long discussions on politics), Krupal, Aman, Akshay, Amie, Vridhi... It was very hard migrating from this village in the middle of PhD (to Namur), but in the end, these connections kept pulling me back so often that I wonder if I even left!

In Namur, I was lucky to find another set of a very supportive group of colleagues and friends. A special mention to two ladies who became terrific friends– Clarice for including me in her extensive social network that made settling down in Namur easier. And, Andrea Renk for hours of technical and non-technical discussions on the projects, co-tutoring Applied Econometrics class, being a lovely host in her Brussels' palace (thanks to Arnaud too!) and proofreading this thesis! Others in the department made great company - Pierre (board game expert), Eric (job-market buddy), Anne Michels (superb office partner), Hugues (best BBQ host and thesis summary translator), Pablito, Ana, Julia, Georgio, Ludovic, Coline, Francois... And others who kept me in sound mind during extreme phase of lockdowns- Rohan and Shalini (for constantly feeding me good food!), Ali, Menel, Marion, Tarcius, Tai, Essam. . . .

I would also like to thank the friends who are based in India and have supported me in the process. To the "Gentlemen" club (and their other halves) for always being there and making my India trips special. You were the most challenging audience to convince with the results. Thanks to Sankar, Ankit, Paramjeet, Mayank and Rishi for listening to the research ideas! Finally, thanks to my family members for their unconditional support, without which I wouldn't have made it. To Dad and Mom, for their perpetual support and belief in all my decisions, to my sister, uncles, aunts, grandparents, awesome cousins and Adharv!

To all those I missed acknowledging...

Summary

Cette thèse de doctorat regroupe trois travaux et chapitres indépendants qui ont pour trait commun la dynamique à long-terme des institutions, des cultures, du capital humain et des inégalités économiques.

Le premier chapitre est un travail co-écrit avec Li Yang. Ce chapitre compare l'évolution de l'accumulation du capital humain et le développement d'institutions modernes d'éduc-ation en Chine et en Inde entre 1900 et 2008. Ce chapitre repose sur des nouvelles données en série temporelle portant sur des mesures d'éducation comme la participation scolaire, les diplômés, les enseignants ou les dépenses en éducation. Ces éléments sur longue période nous permettent notamment de mettre en relief des liens entre les modifications des politiques éducatives (et l'évolution d'indicateurs sous-jacents à ces politiques) et l'évolution des institutions politiques. Entre ces deux pays, les différences principales en matière d'expansion des systèmes éducatifs résident sur l'approche chinoise de type bottom-up par rapport à une approche indienne de type top-down, la diversification engagée par la Chine et la priorité accordée à la qualité de l'Inde par rapport à une approche chinoise priorisant la quantité. La première différence est que la Chine s'est concentrée autour de l'idée d'une massification de l'enseignement primaire (1900-1970), puis s'est ensuite tournée vers le développement du secondaire, et au cours des récentes décennies, l'enseignement du supérieur. De l'autre côté, l'Inde a longtemps ignoré l'éducation de masse au niveau primaire, en raison de l'influence coloniale combinée à une société plus hétérogène que la société chinoise. Ces deux trajectoires distinctes ont résulté d'un plus haut niveau d'éducation moyen en Chine à partir des cohortes nées dans les années 1960. Notre travail reporte aussi que la Chine a développé un solide système d'enseignement professionnel et a produit une plus grande part de diplômés en ingénierie issus de l'enseignement supérieur. En Chine la moitié des étudiants vont vers un système dédié à l'enseignement professionnel alors qu'ils sont à peine 10% en Inde. La part d'ingénieurs que produit chaque année le système éducatif chinois est restée proche de 30 à 45% (du total des diplômés de l'enseignement non-professionnel) sur l'ensemble du XXe siècle, contre moins de 10% en Inde. Pour autant, l'Inde a produit de nombreux diplômés en sciences sociales et humaines. Dans ce chapitre, nous conjecturons que le type de capital humain produit par la Chine était plus adapté pour développer le secteur manufacturier. Le développement du système éducatif en Inde a entrainé non seulement une plus grande inégalité d'éducation, mais également un taux plus élevé de rendement de l'éducation, avec un impact plus élevé sur l'inégalité des salaires.

Le second chapitre mesure l'inégalité des richesses en Inde de 1961 à 2018 et explore, en suivant, les déterminants de la dynamique de ces inégalités. En combinant des données d'enquêtes sur la richesse (NSS-AIDIS) et des listes de millionnaires, je produis des séries sur les inégalités de revenus. Je trouve que la période succédant la libéralisation, lorsque le taux de croissance était très élevé, est aussi celle d'une augmentation substantielle dans la concentration de richesse. D'autres études récentes utilisant le revenu retrouvent ces résultats. Dans notre estimation la plus prudente, je trouve que la part des richesses des 10% les plus riches est passée de 45% en 1981 à 61% en 2018. Dans le même temps, la part des 1% les plus riches est passée de 13% à 33%. Par ailleurs, je souligne deux faits stylisés importants : la terre est l'actif le plus précieux de la richesse des ménages (60% de la valeur totale de la richesse) dans toutes les strates, et les groupes de caste supérieure possèdent une richesse supérieure à leur part dans la population. La répartition historique des terres dans les mains de ces castes (à l'époque coloniale ou avant) semble donc toujours déterminer l'inégalité des richesses. Le chapitre montre rigoureusement cette relation en calculant l'inégalité foncière au niveau village pour 374 000 villages (pris parmi 10 grands Etats indiens). Je trouve que la part de la population de la caste la plus désavantagée (les Dalits, ou autrement appelés les Intouchables) est fortement corrélée aux inégalités au niveau village, après avoir contrôlé pour les facteurs démographiques, institutionnels et géographiques. Ce résultat illustre la forte empreinte de l'Histoire sur la répartition inégale des richesses aujourd'hui. En outre, j'explore dans ce travail un autre mécanisme potentiel derrière la croissance des inégalités : l'augmentation de l'homogamie chez les individus les plus riches. Pour se faire, j'ai élaboré un nouveau jeu de données sur les couples mariés et estimé la corrélation entre les niveaux d'éducation, le revenu et la profession des époux (même si ces données ne contiennent pas malheureusement d'éléments relatifs à la richesse, elles en sont de bons indicateurs). Mes résultats montrent une augmentation de l'appariement sur l'éducation dans la société. Bien que je ne puisse pas directement établir de relation causale, le pattern de distribution des inégalités de richesses des différentes castes suit

complètement la distribution du pattern de distribution de l'appariement sur l'éducation entre ces mêmes castes.

Le chapitre trois est un travail collaboratif avec Sutanuka Roy, basé sur un cas d'étude du plus grand Etat de l'Inde, l'Uttar Pradesh, avec une population de 200 millions d'habitants, aussi connu pour des histoires de conflits communautaires (émeutes entre hindous et musulmans). En utilisant des modèles à effets fixes, nous établissons de manière causale que l'exposition à ces évènements conflictuels pendant la petite enfance (0 à 6 ans) a un impact persistant à long terme, et ce jusqu'à l'âge adulte. Ce travail étudie notamment la prise de décision des juges dans les affaires de détention provisoire, lorsque les décisions sont le plus souvent prises sur la base d'informations limitées. Le chapitre montre que les juges sont plus susceptibles de refuser la libération sous caution (c'est-à-dire de condamner à la prison un accusé) s'ils ont été exposés à des environnements conflictuels dans leur petite enfance. Ces effets ne sont pas déterminés par une sélection particulière dans la fonction judiciaire ou des différences de capacités parmi les juges, mais plutôt par des différences dans leurs préférences pour la loi et un ordre fort et affirmé. Trois éléments de preuve nous permettent d'aboutir à cette conclusion. Le premier est l'absence d'un comportement particulièrement hostile des juges envers des accusés d'une religion différente de la leur. Aussi, nous montrons un effet plus fort si l'exposition se produit entre 3 et 6 ans (période cruciale pour le développement des préférences). Pour finir, nous montrons que nos résultats sont déterminés par des juges qui ont connu des interventions étatiques fortes et efficaces, avec un nombre élevé de confinements et un faible nombre de victimes dans le conflit, offrant à leurs yeux une solution d'un ordre fort.

This PhD dissertation is a collection of three independent chapters linked through the long-run dynamics of institutions, culture, human capital and economic inequalities.

Chapter 1 is joint work with Li Yang. This chapter compares the evolution of human capital accumulation and the development of modern education institutions in China and India between 1900 and 2018. It builds a novel education time series on various educational outcome measures like enrollment, graduates, teachers and expenditure. The long period allows us to understand the changing educational policies and outcomes with changing political institutions. The main differences between these two countries in expanding educational systems are related to - China's bottom-up vs India's top-down, China's diversifying approach, and China's prioritizing quantity vs India's prioritizing quality. The first difference is that China focused on expanding primary-level mass education (1900-70) first, later shifted to develop middle-level and, in recent decades, is expanding its tertiary-level education. Due to colonial influence combined with a more heterogeneous society, India ignored primary-level mass education for a long time. It resulted in higher average years of education in China since the 1960s-born cohort. The paper also finds China has developed a robust vocational education system and produces a higher share of engineering graduates from non-vocational tertiary education. To provide the degree of differences, at the tertiary level, half of the students go through the vocational track compared to a meagre 10% in India. The share of engineers every year the Chinese education system produces has remained close to 30-35% (of the total non-vocational graduates) in the entire 20th century compared to less than 10% in India. India produces a large percentage of humanities and social science graduates. We conjecture that China's type of human capital was more apt for developing the manufacturing sector than India lacked. The second part of the project systematically studies the education-wage inequality linkage. The development of education in India resulted in not only higher educational inequality but also a higher rate of return resulting in a more substantial impact on wage inequality.

Chapter 2 first measures the wealth inequality in India from 1961-2018 and later explores the determinants of the inequality dynamics. I combine data from wealth surveys (NSS-AIDIS) and millionaire lists to produce wealth inequality series. I find a substantial rise in wealth concentration in the post-liberalization period when the growth rate was

very high. It is in line with recent research using income. The most conservative estimate shows that the top 10% wealth share rose from 45% in 1981 to 61% in 2018, while the top 1% share rose from 13% to 33%. Two important stylized facts - land is the most valuable asset of household wealth (60% of the total wealth value) in all strata, and higher caste groups own much more wealth than their population share. The historical distribution of land in the hands of the upper caste (during colonial times or even before) still seems to determine wealth inequality. The paper shows this relationship rigorously by computing land inequality at the village level for 374k villages (universe of villages from ten large states of India). I find the share of the Scheduled Caste (or Dalits, the lowest caste group) population is strongly correlated to the village-level land inequality, controlling for institutional, geographical and demographic factors. It highlights the strong imprint of history on the current-day wealth inequality. I explore another plausible mechanism behind rising inequality - increasing homogamy between wealthy individuals. I prepare a novel dataset on married couples and estimate the correlation between husband and wife's education, income and occupation (unfortunately, this data does not contain wealth, but these are a good proxy for wealth). I find an increasing education assortativity in society. Though I do not causally establish the relationship, the patterns across different caste groups (which could be treated independently of each other given 95% of the marriages are within castes) perfectly match the level of wealth inequality within each caste group.

Chapter 3 is joint work with Sutanuka Roy. The setup is the largest state of India, called Uttar Pradesh, with a 200 million population, and the state has a history of communal conflicts (Hindu-Muslim riots). The paper uses fixed effects approach to causally establish that exposure to these conflicts environment during early childhood (age 0-6 years) has a long-term persistent impact till adulthood. It studies judges' decision-making in pre-trial detention cases where decisions are often made on limited information. The paper shows that judges are more likely to deny bail (i.e. send a defendant to jail) if they were exposed to conflict environments in their early childhood. It shows that the effects are not driven by selection into a judicial occupation or judges' ability differences but rather by changes in preferences for strong law and order. Three pieces of evidence help in arriving to this conclusion: absence of inter-group hostility behaviour among judges, stronger effect if the exposure happens during 3-6 years

of age, which is a crucial time for preference development and the result is driven by judges experiencing effective state intervention, i.e. high state-imposed lockdowns and low casualties in the conflict.

Table of Contents

List of Tables

List of Figures

General Introduction

Overview

Institutions and culture play a crucial role in economic development. The input components - capital, labour, human capital and productivity (or technological progress) - and their accumulation are necessary conditions for growth. At the same time, persistence in the cross-country differences suggests they are not sufficient. Several other factors like luck, geography, institutions (Hall and Jones, 1999; Acemoglu and Robinson, 2012) and culture (Guiso, Sapienza, and Zingales, 2006; Tabellini, 2010; Nunn, 2020) determine whether the inputs are employed or accumulated towards attaining economic development.[1]

Institution and culture are extensive terms with some overlapping features. They both are societal constructions (different from deterministic geography). They both are persistent, though culture is more persistent than institutions (Roland, 2004).[2] Culture (including values, beliefs and social norms) could be seen as a slow-moving institution. The fast-moving institutions can change quickly (but do not necessarily change often). Even though others have studied the influence of institutions and culture on economics separately, they are not independent of each other. Recent literature explicitly tries to establish that institutions can impact cultural traits (Lowes et al., 2017) or they are jointly determined (Bisin and Verdier, 2017)

One line of research contests the earlier argument that - historically, it was "institutions" that caused economic growth. Another strand of research theorizes it was first the rise in human capital that caused effective institutions to develop and eventually boosted pro-growth policies (Lipset, 1981; Glaeser et al., 2004). In other words, human capital would be a more fundamental cause of economic growth. And indeed, Korea

[1]Chapter 7 of Campante, Sturzenegger, and Velasco, 2021 provides a nice summary of the literature on the causes of modern economic growth.

[2]The persistence of cultural aspects is not necessarily economic efficiency- parents imparting cultural norms Bisin and Verdier, 2000 or vested interests of the organizations that play a role in promoting culture.

and some South East Asian countries had human capital increasing first, leading to better institutions and further development. There is consensus in that history has a pronounced impact in shaping the current day outcomes (Banerjee and Iyer, 2005; Dell, 2010; Nunn, 2008; Acemoglu, Johnson, and Robinson, 2001; Iyer, 2010). The unsettled debate is whether it occurred due to establishing good institutions first or whether it was human capital accumulation or other factors.

For a long time, societies worldwide have been unequal. A small group of elites own everything (material and power), and a large excluded section has almost nothing. The idea of equality during the French revolution transcended borders and helped achieve political equality, but economic inequality remained sidelined. The rising inequalities during the phase of economic growth were seen as a devil that would correct itself over time. However, a systematic long-run inequality analysis shows that inequalities are persistent, similar to institutions or culture. They have transformed their nature from status-based to proprietary-based (Piketty, 2014; Piketty, 2020). Today, economic inequalities are political issues in almost every country, and equitable distribution of resources is an international objective of the United Nations Sustainable Development Goals.[3]

To measure economic inequality, many variables have been used, such as consumption, wage, income, and wealth (and land). The inequality measured through consumption and wages is often lower than inequality computed using other variables, hence providing a lower bound to the existing inequality. On the other hand, consumption, wages (and even income) fluctuate more; hence the inequality measures are also volatile. Wealth distribution moves slowly. Within wealth, land distribution has been studied more extensively and shown to have adversely impacted investment in education (Galor, Moav, and Vollrath, 2009) and institutions (Sokoloff and Engerman, 2000) to eventually impacting economic growth and inequality. Deininger and Squire, 1998; Easterly, 2007

[3]Sustainable Development Goal 10 aims at reducing inequality within and among countries. This SDG calls for reducing inequalities in income as well as those based on age, sex, disability, race, ethnicity, origin, religion or economic status within a country. The goal also addresses inequalities among countries, including those related to representation, migration and development assistance.

The idea of institution and culture impacting economic development (growth and inequality) tends to be very country-specific. Colonization in the past and heightened economic globalization in the post-colonization period have reduced the differences in institutions and culture to some extent between countries, but they are still very distinct country-wise. It calls for a country-specific study to understand the similarities and divergences in their economic development. The two largest Asian countries - India and China, have always caught the world's attention due to their richness. However, centuries of colonial interventions created a hub of poverty and deprivations in these two countries in the modern historical period. In the post-colonialism period, both countries are fast improving their living standards relative to other developing countries, which has once again caught the eye of the world. The current form of institutions in these two countries has imprints of the recent (foreign) and old (traditional) past.

The first chapter of the thesis studies human capital accumulation beyond average years of education in China and India while looking deeply into the evolution of educational institutions and adopted policies in these two countries under different political regimes. The second chapter of the thesis studies wealth inequality in India and links it to the persistent caste system (culture). Finally, the third chapter of the thesis causally establishes the relationship between exposure to a particular institutional intervention (administration effectively controlling the conflict) and behavioural change in a population under study (judges becoming stricter in giving judgements).

Structure and Contributions

The thesis is composed of three independent chapters, combining several novel data sources and using different methodological tools. The data contribution is a major part of the thesis. The first chapter builds a novel dataset on the entire spectrum of education-related variables for China and India from 1900-2018. It allows for building quantitative and qualitative measures of human capital. The variables include not only the absolute number of - enrollments, graduates, teachers, and expenditure- but are also split by levels of education (primary, middle and higher), gender and type of education (general vs vocational). The expenditure pattern allows us to understand the educational investments made by these two countries over time. In the second chapter, I categorize castes into finer levels to provide basic demographic, income, consumption and wealth averages, which are completely missing in the current literature. Another general data contribution is merging two village-level census datasets for 374 thousand villages of India to explore land inequality at the lowest administrative unit possible. The third chapter adds the number of days lockdown was imposed in the communal riots from 1950-2000 in the state of Uttar Pradesh (India).

The thesis links the progress of institutions, culture and economic development. The first chapter shows how modern educational institutions developed differently in China and India. In the colonial phase (pre-1947) in India, the extractive colonial government and status-based elites (who came majorly from the upper caste or class) came together to create modern institutions in such a way as to keep the hegemony of elitist groups. The educational institutions served the colonial government with a western-educated workforce while it paid off local elites through government jobs. In China, modern educational institutions emerged due to the internal revolution, which demolished the traditional institutions responsible for creating elites (through civil services exams). Although political institutions changed several times in China and India, there was a certain consistency in educational policies (except during the decade of the Cultural Revolution in China or during 1930-50 in India when Britain was involved in wars) for a long period. This chapter also highlights that the type of human capital accumulation has the potential to explain the development of the manufacturing sector in China, service sector growth in India and higher wage inequality in India. Overall, the chapter seems to indicate that in the context of these two countries, the educational policies (and resulting human capital) played a more prominent role in economic development than the type of institutions.

The second chapter finds the link between current-day wealth inequality and the old caste system in India, highlighting the persistence of historical distribution of land ownership based on caste. The main contribution is providing evidence using village-level agricultural land inequality, which has not been done so far. The results are robust to controlling for institutional, geographical and demographic factors. It also highlights that the cultural trait of marrying within caste in conjunction with an evolving feature of marrying with similar educational levels (a proxy for economic outcome) may be a potential channel behind rising economic inequality in general and within-caste inequality in particular.

The third chapter is set up in the background of the current-day conflict erupting from the animosity between two religious groups (Hindu and Muslims) fuelled in the past through the divide and rule colonial government policies. It shows that judges exposed to effective state intervention in handling conflict during their early childhood show stringent behaviour in giving judgements.[4]

The thesis also contributes to the field of measurement. Accurate measurement often is challenging in any historical analysis due to data constraints. Comparing China and India in the first chapter added the requirement of consistency of measures between these two countries. The first chapter creates a novel measure of the Education Investment Ratio, defined as total expenditure per school-going kid population normalized by gross national income per capita). It does not require price index and exchange rates (usually difficult to get) for cross-country comparisons. It captures the government spending on its education system to develop human capital (a critical input for economic development), considering the overall economic and demographic attributes. The EIR measure also has the feature of simple multiplicative decomposition into quantity and quality components. In the second chapter, I build new measures for land ownership inequality at the village level. The first measure is the land dependency ratio, the ratio of households dependent on agriculture and households owning agricultural land. This measure is better than the usual land inequality measure (like gini) as it considers that some families are no longer dependent on agriculture. The second measure is the share of land owned by the top 1, 2 and 3 households in a village.[5]

I move to highlight the chapter specific main findings, its contributions in other areas and policy relevance.

[4]Though it links institutional action with changing preferences, it is not enough to call it changing of culture (due to the focus on a particular population group)

[5]Though I use the measures for stylized fact in this thesis, it could be used to study political economy at the village level. E.g. There is anecdotal evidence that wealthy households usually influence the decision to vote for a particular candidate in a village.

Human Capital Accumulation in China and India

The literacy rate at the world level grew from 20% to 80% in the twentieth century, highlighting a general acceptance that education is vital for the development of a nation. However, today the most pressing question for several countries is how best to develop the education system to achieve higher economic development. The resource constraint in several low and middle-income countries makes this more challenging. The essential questions, such as whether to allocate more resources to primary or higher education; vocational or general education; which disciplines in higher education; prioritizing quantitative expansion (with quality taking a back seat in the beginning) or keeping a balance etc., are difficult questions for policymakers. The answer is not straightforward for several reasons- the template of developed countries may not be incorporated now given faster technological change, country-specific constraints etc. This chapter serves as a case study of the two largest developing countries with very different histories and political setups.

One main difference worth highlighting in the general context is that China followed a bottom-up approach model, focusing on its primary-level mass education until it reached a more than 90% net enrollment rate. It is certainly not to say that middle or tertiary-level remained non-existent, but a major share of expenditure and the policy debates revolved more around primary education. On the other hand, India's education system was biased toward middle and tertiary-level education. Primary level education was ignored for a long time. This different approach led to a higher literacy rate and more average years of education in China.

The other interesting difference is related to the diversification of education. The diversification of education is measured through a share of vocational enrollments/graduates and the share of graduates from different disciplines from the tertiary-level non-vocational graduates. China produces a high share of vocational and engineering graduates and a lower share of social science graduates. On the other hand, in India, the development of vocational education was subdued and more than 90% middle-level graduates go towards standard degree courses and more than half towards social science. These results could lead one to think that, in some sense, both countries adjusted their type of human capital production to avoid direct competition at the global level. China became a goods provider, and India, a service provider at the world level.

It is unlikely to be the case. First, the differences in diversification were present even before both countries opened up their economies (pre-1980). The share of engineering graduates was close to 35% in the 1960s in China, compared to less than 5% in India. The percentage of humanities and social science graduates has remained above 50% throughout the century in India. Second, reading into the educational policies provides a clear answer. Both countries focused on developing the heavy industry just after their independence/liberation. But only China simulated the Soviet-style higher education system to replace the British and American-style higher education (adopted in the Republic of China, i.e. 1911-49). It led to the abolition of all private universities, and disciplines like - engineering, teacher's training, agriculture and forestry were given more emphasis to promote industrialization, ignoring the humanities and social sciences subjects. On the other hand, India did not tinker much with its existing institutions (possibly because of the long existence of the British-established higher education system).

This chapter assumes that education remains a tool for economic enhancement. It has been true for China since the inception of modern education at the beginning of the 20th century. The objective of education in India has always remained broader than simply economy-enhancing. In the pre-independence period of India, the purposes of education enunciated by the colonial government were to impart western knowledge and culture, promote intellectual development, and raise the young generation's moral character (Wood's Despatch 1854 and Indian Education Commission Report 1882). The economy-enhancing was not even the goal. The Indian intelligentsia back then also viewed education as a tool for nation-building. Post-independence, education was seen as a factor vital to national progress and security. The first national education policy in 1968 explicitly mentions education for the economic and cultural development of the country. The Education policy of 1986 (revised in 1992) also adds the acculturating role of education as one of the roles of education other than economic development. [6] Hence, the success of educational institutions should be seen with this difference in mind.

[6]The National Education Policy of 2020: "The purpose of the education system is to develop good human beings capable of rational thought and action, possessing compassion and empathy, courage and resilience, scientific temper and creative imagination, with sound ethical moorings and values.

Lastly, the chapter shows that education explains a considerable share of existing wage inequality in India. The rate of returns to education in India started increasing post-liberalization (in the 1990s), and the highest increase occurred during 2000-2011. It is perplexing that during this period, there was an increase in the supply of educated graduates. It suggests that the education system does not meet the increasing demand for high-skilled workers. It links to the type of human capital India produces, where 50-60% tertiary-level graduates come from humanities and social sciences. In China, the rate of return to education also increased post-liberalization up to 2002. However, it decreased between 2002-2013 because of a vast increase in the supply of higher education graduates.

Wealth Inequality, Class and Caste in India

This chapter deals with inequality in detail in India. It first produces the series of wealth inequality in India from 1961-2018 and finds an increasing trend in wealth inequality (with the highest increase during 2002-12) in line with the other research on income inequality. The top 10% of Indian household wealth holders own 61% of the total household wealth in 2018, compared to 43% in 1961. The wealth concentration within the top 10% population is very high. The top 5% of Indian household wealth holders owned 50% of the total household wealth in 2018, compared to 31% in 1981. The top 1% of Indian household wealth holders hold 33% of the entire household wealth in 2018, compared to 13% in 1981. Compared to other developing countries, the wealth concentration in India is higher than in China but lower than in South Africa.

The land constitutes the most important asset in the wealth basket of Indian households, making up to 60% of the total wealth value. The paper finds an increasing land price premium associated with the land of the rich class (top 10% population) for any given type of land in rural and urban areas. It suggests an underlying bias in the spatial development process in India, where the land around the rich community is possibly developed more (E.g. construction of roads, sewage etc., in urban areas and irrigation facilities in the rural areas). Also, it finds a pattern of changing portfolios of land types, where the rich class is moving towards owning more expensive non-agricultural land.

The dynamics of wealth inequality revolve around a land distribution that is historically linked to the caste structure. The paper demonstrates, using extensive data on agricultural land area ownership at the village level, that the share of the Scheduled Caste (the lowest caste group) population is positively correlated with land inequality. The relationship holds using district (or sub-district) fixed effects capturing institutional differences and village-level geographical, climatic and demographic controls (like literacy rate, distance from town etc.). The positive association remains in each state. The paper does not claim causality, but such a strong correlation indicates a strong influence on the historical land distribution of the past based on caste.

Another interesting result from the paper is the high level of wealth inequality within each caste group. The highest within-caste inequality is present in the Forward Caste (a group of upper caste, FC) and Scheduled Tribe (ST) groups. It is interesting because the FC group, on average, is the richest (and most educated), and ST is the poorest (and least educated) of all other caste groups. The paper finds that this phenomenon could be explained by the high education and wage assortativity in these two caste groups, i.e. people marrying with similar educational (or wage) backgrounds. The high level of economic inequality within the FC group explains the demand from some upper caste groups in India to be categorized under Other Backward Caste groups to take advantage of affirmative actions.

Overall, the high level of inequality is a matter of concern in India, where the unfairness of the past against lower caste groups has not been mitigated despite government policies.

The Origins of Judicial Stringency in Bail Decisions: Evidence from Early-Childhood Exposure to Hindu-Muslim Riots in India

The third chapter is motivated by the observation that millions of people worldwide are detained without being charged guilty by the court. Such prisoners are called pre-trial detainees (or under-trial prisoners in India). Often low-income and minorities are over-represented among these pre-trial detainees. Such detentions have been shown to have negative economic and criminogenic consequences on the lives of detainees. The Indian judicial system has one of the world's highest shares of pre-trial detainees. More than 70% of total prisoners are pre-trial detainees. The decades of government efforts to reduce the share have not borne fruit. This chapter examines the origin of judicial stringency in the pre-trial decision-making process in India's largest state (Uttar Pradesh).

The third chapter is also motivated by the early-childhood literature, where there is ample evidence now that early childhood interventions have long-term impacts on cognitive skills, health outcomes, and labour market outcomes. These interventions by the government (and parents) in more formal ways endeavour towards a healthy and educated human capital. Evidence also shows that early-life exposure to a sociopolitical environment engenders the development of fundamental parameters, such as later-life social preferences, preferences for honesty, political identity, and inter-group behaviour (Cappelen et al., 2020; Abeler, Falk, and Kosse, 2021; Billings, Chyn, and Haggag, 2020; Couttenier et al., 2019; Fisman et al., 2020).

The chapter provides causal evidence that judges' exposure to the communal conflict environment (Hindu-Muslim riots) during their early childhood (0-6 years of age) makes them more stringent in their adulthood compared to other judges. It exploits the following institutional features - judges are never posted in their home districts (where they were exposed to the conflict environment), they are transferred across districts every three years, and cases are allocated based on rules. These features create quasi-random rotation of judges and quasi-random allocation of cases. The key identifying variation is driven by quasi-random assignment of judges to courts and cases to judges, combined with ex-ante heterogeneity in childhood exposure to riots among judges. The main result is that early-childhood exposure to communal conflict increases the share of pretrial detentions by six percentage points, which is 16% of the mean. It shows that the effect is not driven by differences in the ability of judges or inter-group hostility behaviour (i.e. Hindu judges' detention rate is not different with Hindu or Muslim defendants) among judges.

The heterogeneity results show no evidence of the strictness of exposed judges coming from a certain type of case. The heterogeneity by age shows that the effect is driven by the judges who are exposed between 3-6 years of age, which early-childhood literature has identified to be crucial for preference formation. The effect is also driven by the judges who were exposed to the conflict environment, where the state intervention in controlling the conflict was effective. Overall, the findings suggest that early-childhood exposure to state-imposed lockdown measures that proved effective in containing violence generated higher support for state institutions in law-and-order matters.

Bibliography

Abeler, Johannes, Armin Falk, and Fabian Kosse (2021). *Malleability of Preferences for Honesty*. Working Paper 14304. IZA Discussion Paper.

Acemoglu, Daron, Simon Johnson, and James A Robinson (2001). "The colonial origins of comparative development: An empirical investigation". In: *American economic review* 91.5, pp. 1369–1401.

Acemoglu, Daron and James A Robinson (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Currency.

Banerjee, Abhijit and Lakshmi Iyer (2005). "History, institutions, and economic performance: The legacy of colonial land tenure systems in India". In: *American economic review* 95.4, pp. 1190–1213.

Billings, Stephen B, Eric Chyn, and Kareem Haggag (2020). *The Long-Run Effects of School Racial Diversity on Political Identity*. Working Paper 27302. National Bureau of Economic Research.

Bisin, Alberto and Thierry Verdier (2000). ""Beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits". In: *The Quarterly Journal of Economics* 115.3, pp. 955–988.

— (2017). *On the joint evolution of culture and institutions*. Tech. rep. National Bureau of Economic Research.

Campante, Filipe, Federico Sturzenegger, and Andrés Velasco (Oct. 2021). *Advanced Macroeconomics. An Easy Guide*. London: LSE Press, p. 418.

Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden (2020). "The Effect of Early-Childhood Education on Social Preferences". In: *Journal of Political Economy* 128.7, pp. 2739–2758.

Couttenier, Mathieu, Veronica Petrencu, Dominic Rohner, and Mathias Thoenig (2019). "The Violent Legacy of Conflict: Evidence on Asylum Seekers, Crime, and Public Policy in Switzerland". In: *American Economic Review* 109.12, pp. 4378–4425.

Deininger, Klaus and Lyn Squire (Jan. 1, 1998). “New ways of looking at old issues: inequality and growth”. In: *Journal of Development Economics* 57.2, pp. 259–287.

Dell, Melissa (2010). “The persistent effects of Peru’s mining mita”. In: *Econometrica* 78.6, pp. 1863–1903.

Easterly, William (Nov. 1, 2007). “Inequality does cause underdevelopment: Insights from a new instrument”. In: *Journal of Development Economics* 84.2, pp. 755–776.

Fisman, Raymond, Arkodipta Sarkar, Janis Skrastins, and Vikrant Vig (2020). “Experience of Communal Conflicts and Intergroup Lending”. In: *Journal of Political Economy* 128.9, pp. 3346–3375.

Galor, Oded, Omer Moav, and Dietrich Vollrath (Jan. 1, 2009). “Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence”. In: *The Review of Economic Studies* 76.1, pp. 143–179.

Glaeser, Edward L, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004). “Do institutions cause growth?” In: *Journal of economic Growth* 9.3, pp. 271–303.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2006). “Does culture affect economic outcomes?” In: *Journal of Economic perspectives* 20.2, pp. 23–48.

Hall, Robert E and Charles I Jones (1999). “Why do some countries produce so much more output per worker than others?” In: *The quarterly journal of economics* 114.1, pp. 83–116.

Iyer, Lakshmi (2010). “Direct versus Indirect Colonial Rule in India: Long-Term Consequences”. In: *The Review of Economics and Statistics* 92.4, pp. 693–713.

Lipset, Seymour Martin (1981). *Political Man: The Social Bases of Politics*. Johns Hopkins University Press. 622 pp.

Lowes, Sara, Nathan Nunn, James A Robinson, and Jonathan L Weigel (2017). “The evolution of culture and institutions: Evidence from the Kuba Kingdom”. In: *Econometrica* 85.4, pp. 1065–1091.

Nunn, Nathan (2008). “The long-term effects of Africa’s slave trades”. In: *The Quarterly Journal of Economics* 123.1, pp. 139–176.

— (2020). “The historical roots of economic development”. In: *Science* 367.6485, eaaz9986.

Piketty, Thomas (2014). *Capital in the twenty-first century*. Cambridge Massachusetts: The Belknap Press of Harvard University Press. 685 pp.

— (2020). *Capital and ideology*. OCLC: 1119745744.

Roland, Gérard (2004). “Understanding institutional change: Fast-moving and slow-moving institutions”. In: *Studies in comparative international development* 38.4, pp. 109–131.

Sokoloff, Kenneth L and Stanley L Engerman (2000). “Institutions, factor endowments, and paths of development in the new world”. In: *Journal of Economic perspectives* 14.3, pp. 217–232.

Tabellini, Guido (2010). “Culture and institutions: economic development in the regions of Europe”. In: *Journal of the European Economic association* 8.4, pp. 677–716.

CHAPTER 1

Human Capital Accumulation in China and India in the 20th Century

With Li YANG.[1]

Abstract

The education system of a country is instrumental in its long-run development. This paper compares the historical evolution of the education systems in the two largest emerging economies- China and India, between 1900 and 2018, through a newly created time series on educational statistics at three levels of education (primary, middle and tertiary), combining reports and surveys. There are three main results. First, China has adopted a bottom-up approach in expanding its education system than India's top-down approach, resulting in a higher average years of education and lower education inequality in China since 1907. Second, China has diversified its education system more through higher vocationalization and a different mix of disciplines at the tertiary level with more engineers. We conjecture that a better mix of engineering and vocational graduates produced the human capital apt for the developing manufacturing sector. Third, the Chinese education model focuses on quantity first, whereas the Indian model focuses on quality. Finally, utilizing micro-survey data since the 1980s, we show that the education expansion strategy of India has increased inequality due to both an unequal distribution of educational attainment and higher individual returns to education.

[1]We would like to thank François Bourguignon, Guilhem Cassan, Oliver Vanden Eynde, Clément Imbert, Thomas Piketty, and Giles Postel-Vinay for their comments and support. We gratefully acknowledge the comments received during presentations held at Paris School of Economics, University of Namur, JNU Delhi and Economic History Workshop (Warwick).
Li is postdoctoral researcher at Berlin School of Economics (yangli1997@hotmail.com).

1.1. Introduction

The importance of the composition of human capital, imparted through education, for a country's long-term economic development is widely accepted by economists and policymakers alike. However, policymakers in low- and middle-income countries face a trade-off when developing educational systems. They can use their limited resources either for primary or higher education. They may focus on either vocational or general education; may produce either engineers or economists; may prioritize either quantity expansion or quality. The relevant literature guiding policymakers to these trade-offs is surprisingly very sparse. To our best knowledge, this paper is the first to study the long-term impact of education development policy. We study this long-term impact by comparing the development of the education system of the two most populous emerging nations, China and India, in the last 120 years.

China and India, with a combined 36% of the global population and 20% of the world's GDP, are the two major economies of the world today[2]. The two countries had a GDP per capita of comparable size until 1980. Then China started growing faster. Today, China's GDP per capita is double that of India's. The development of China came from the manufacturing sector. At the same time, the Indian economy benefited from service sector growth[3]. This economic divergence has attracted much more attention than the divergence in their literacy rate, which started thirty years before their GDP divergence. Both countries had about 20% literacy rate in 1950. In 1990, China had 25 percentage points higher literacy rate than India.

The first question we ask is how China and India expanded their modern education systems. We prepare a novel dataset of extensive education series utilizing multiple volumes of historical and current reports, yearbooks and censuses since the early twentieth century. We harmonize the education data series to make it comparable across time and two nations. The core variables relate to - enrollments, teachers, graduates, and expenditures. We also provide discipline-wise data in higher education. We focus on the annual

[2]Bolt and Zanden, 2020 accounting shows that these two countries were the two largest economies of the world for a large part of history. Their declining economic position in the world started with the industrialization of Europe and colonization in the 18th-19th century.

[3]According to the World Bank Data, manufacturing share difference (China-India) has remained more than ten percentage point 1950's-2010. Service share difference (China-India) was negative until 2010.

flow of the variables and link the observed patterns with the adopted educational policies prevailing at different periods.

The second question we ask is whether different paths of education expansion explain the observed differences in wage inequality in these two countries. The relevance of the question is evident with a secular expansion of education in every country, inequality becoming a non-ignorable aspect of economic development[4] and the dynamic complex relationship between them. Cross-country studies show that both the levels of education and education distribution contribute to wage inequality (Ahluwalia, 1976, Ram, 1990, Gregorio and J.-W. Lee, 2002 Castelló-Climent and Doménech, 2021). We harmonize nationally representative labour force surveys from 1980s to 2018 from these two countries to explore the relationship.

The first main finding is that, China followed a *bottom-up* mode of expansion, initially expanding its primary level mass education(starting from early 1900s pre-communism period), followed by middle level (started during communism) and finally tertiary level elite education(in post-communism, post 1980s). On the other hand, the Indian education expansion resembles a *top-down* approach. The focus was the highest at the middle level until 1950 (British Raj era), then towards tertiary level (in post-colonial socialist phase) and finally towards primary level (post-liberalization). Also, the period of pre-communism period of China was spending much more in education than the British colonial period of India.

The comparison of the evolution of enrollment, expenditure, and teachers by education level (primary, middle and tertiary) supports the above finding. E.g., China overtook India in enrollment (both absolute numbers and per kids population measured by gross or net enrollment rate) in the 1930s in primary education, 1970s in middle education and 2010s in higher education.[5] The result is robust in accounting for the differential trajectories of population and economic growth. This finding is also consistent with a careful reading through the educational policies adopted in these two countries (Section 1.2.2).

[4]UNDP 2019 report highlights the increasing income and wealth inequality in the world.
[5]It is because the modern education started in India, by Britishers, almost 50 years before China.

E.g. the debate on compulsory education and its implementation in China started during the pre-communism period, whereas a serious implementation of the compulsory education phase started post-1990 in India. Adopting different education development paths helped China impart higher human capital (average years of education) to every birth-year cohort since the beginning of the 20th century than India. To illustrate, for the cohort born in 1962, the average years of schooling in China is 8.9 years compared to only 3.4 years in India.

Second, there is more diversification of the education system in China than in India. The diversification is measured through two different statistics - share of vocational students and share of students in different disciplines at tertiary level general education. The high vocationalization has been an important feature of the education system in China, where today, almost 25% of students at middle and tertiary levels combined are enrolled for some vocational education which stands at around 2% in India. [6] This is important as recent evidence shows that vocational education is more growth-enhancing if a country is farther from the productivity frontier or if the rate of growth of frontier technology is slower.[7]

At the tertiary level, the distribution of students by disciplines in China has changed dramatically over the years. Before the 1950s, humanities and law students accounted for more than 50% of total enrollments in higher education; the second half of the 20th century saw a great expansion in the disciplines of engineering and education; since the 1980s, the shares of enrollment in law and economics (and management) is increasing. In India, the distribution of students by discipline is relatively stable. Since 1897, humanities and law students account for 60% of the enrollment, while the share of students in engineering and education is much smaller than in China. Several studies show that

[6]The suppression of vocational education during the cultural revolution has been reversed in China after opening up the economy in 1978.

[7]D. Krueger and Kumar, 2004 shows that a country's optimal education policy to provide subsidies for general versus vocational education should depend on the growth rate of the frontier technology. In particular, the European focus on specialized vocational education might have been effective during the 1960s and 1970s but resulted in a growth gap relative to the US after the 1980s when new technologies emerged more rapidly. Similarly, Aghion et al. (2005, 2009), using US states' level panel dataset, highlight that research education is more growth-enhancing in states closer to the productivity frontier. In contrast, vocational education is more growth-enhancing in the states those are farther below the productivity frontier.

engineering and science are positively associated with higher innovation and growth. [8]

Third, an analysis of the quantity-quality tradeoff suggests that China's has prioritized quantity first; in contrast, India prioritized improving quality. We create a measure education investment ratio (EIR), total expenditure per kid population normalized by gross national income per capita (GNIpc) for primary, middle and tertiary levels separately. It is decomposed into three multiplicative components: *Quant* (GER), $Qual_1$ (inverse of pupil-teacher ratio;1/PTR) and $Qual_2$ (expenditure per teacher normalized by GNIpc, a proxy for teachers' salary). In the primary (and middle) stage, till the late 1960s, China prioritized increasing GER, during which both quality measures deteriorated. From 1970, after attaining >100% GER at primary (and close to 40% GER at the middle level), it started hiring more teachers (lowering PTR and improving $Qual_1$), and post-1980, it started improving $Qual_2$ (improving teachers' relative salary to attract better talent). In India, since the beginning of the 20th century, the debate shifted towards improving the quality through increasing teachers' wages and creating a few "model" institutions (of high quality), which continued till the early 1990s. During this period, at the primary stage, there is a continuous improvement in $Qual_2$, though the lack of hiring enough teachers resulted in deteriorating $Qual_1$. Post-1990, there was a change in stance in India at the primary (and middle level) to focus on increasing quantity (GER) which came at the cost of declining quality.

Finally, we observe similarities along the dimension of the gender gap. Both countries bridged the gender gap at the primary and middle levels of education in the 20th century in enrollment. The female share in enrollment is now at par with the total population. There is an increasing trend of feminization in the teaching profession. At the primary level, the share of female teachers is now above 50% in both countries.

[8] Romer, 1990 and Mokyr, 2005 identify research engineers and engineering-minded technicians to be the key to innovation. In a recent paper, Maloney and Caicedo, 2017 argues that the density of engineers in 1880 captures historical differences in innovative capacity, which in turn explain a significant fraction of the Great Divergence in the Americas. Toivanen and Väänänen, 2016 also find the causal effect of M.Sc. engineering education on invention, using data on US patents' Finnish inventors. Their counterfactual calculation suggests that establishing three new technical universities resulted in a 20% increase in the number of USPTO patents by Finnish inventors.

We find that the education expansion contributed to the wage inequality in India more - due to higher education inequality and a higher rate of returns to education. Using harmonized household wage surveys from the 1980s to 2018 in China and India, we decompose the wage inequality by education levels. The within-group (where groups are primary, middle and higher) wage inequality of the two countries is comparable. In contrast, the between-group inequality (capturing the "education effect") in India is much higher than in China. The impact of education distribution on wage inequality (using unconditional quantile regression from Firpo, Fortin, and Lemieux, 2009) show that an increasing share of tertiary graduates is associated with increasing wage inequality in both countries. However, in China, it starts only after 2000.

The backbone of this paper is the novel dataset we build. We complement the literature on measuring development of education and human capital accumulation - Baier, Dwyer, and Tamura, 2006;Fuente and Donénech, 2000; Cohen and Soto, 2007; Morrisson and Murtin, 2009; Barro and J. W. Lee, 2013, 2015; J.-W. Lee and H. Lee, 2016[9]. Over the years, studies have expanded the coverage and improved the quality of human capital measures. However, they have remained limited to broad measures like average years of schooling and enrollment ratio. More detailed and long-run information on education development is needed to answer the questions we raise. In this paper, we bridge the gap by constructing a data set with the coverage of the whole spectrum of education variables for China and India in the last 150 years. We have unearthed multiple volumes of official education reports and education statistic yearbooks of China and India dated back to 1907, which are surprisingly under-explored in the previous literature. Available variables include not only the number of teachers, enrollments and graduates by gender, the stage of education (primary, secondary and higher education) and type of education (general education vs vocational education) but also the education expenditures by the stage of education (Refer to Appendix C.4 for details). In particular, we also provide discipline-wise data in higher education. Further, we keep human capital central in our paper rather than as a means to understand variation in national income.

[9]For pre-2000 literature Pascharopoulos and Arriagada, 1986, Lau, Jamison, and Louat, 1991, Nehru, Swanson, and Dubey, 1995

We add to the literature on the education series for China and India, which are still scattered, incomplete and often rely on second-hand sources. The papers measuring human capital related variables for India and China have relied on either of the sources: Mitchell, 1998, UNESCO[10], Gao, 2018 (for China) [11] Leeuwen, 2007 (for India); the first two for historical and the last two for contemporary time periods. We provide a detailed comparison with other studies in Appendix C.3. We improve from the existing datasets by providing a harmonized dataset for a longer time[12]. The harmonization relates to incorporating the Indian complexity of primary stage students studying in secondary schools and class XI-XII as part of college education before the 1960s and school education later.[13] The existing studies have ignored this aspect as even the published statistical reports have not harmonized the series over the years. It has led to estimating some statistics like pupil-teacher ratio and expenditure per student. To our knowledge, we are the first to provide harmonized and comparable education series for China and India for such a long period.

Our comparative study of (British) India before 1950 compared to China, which was partially colonized, supplements the comparative literature on the provision of education in British colony versus French colony (Cogneau and Moradi, 2014), British colony versus Dutch colony Indonesia (Leeuwen, 2007).[14] The decline in the public expenditure (as a share of gross national income) between 1930-45 in India created the gap between India and China. The public share in the total expenditure also declines during this period in India. We conjecture that external (European) factors like the Great Depression and World War II negatively impacted the public investment in education in India more than under direct British rule. Post-independence (i.e. after 1950), the comparative study takes the form of social democracy (India) versus communism (China) set up. The investment in education increased in India, but there was continuance in the policy of

[10] UNESCO, 1958, UNESCO, 1961b,UNESCO, 1961a

[11] Chaudhary, 2009 uses the same source as ours for India, but focus on understanding the expansion of Primary education regionally within India, similar to what Gao, 2015 does for China

[12] Mitchell, 1998 provides enrollment from the reports up to 1993 for India since 1870 and for China since 1950. The other often used historical dataset of UNESCO provides enrollment and teachers but starts from 1930.

[13] The Calcutta University (Sadler) Commission in 1922 had recommended that the dividing line between college and school should be intermediate (XII) and not matriculation (X). The National Education Policy 1966 formally gave the 12 (school) + 3(college) structure.

[14] The comparison of educational policies in British colonies versus colonizers has been studied extensively too, e.g. British India versus British (Naik, 2000, British India versus Japan (Leeuwen, 2007)

focusing more on the Middle/Higher elite education (Figure 1.4) rather than undertaking a massive mobilization towards mass Primary level education. It was partly because the enrollment at the middle level was already considerable but partly due to domestic factors. The domestic factors like higher stratification in the society as compared to China and gradual, incremental approach (compared to big-bag reforms in China) contributed towards retaining the early advancement in education in China (Chaudhary, 2009, Arnove, 1984). We highlight that both the external and domestic factors have contributed to the observed expansion mode.

We also speak to the literature on education and inequality. A vast literature highlights the growing economic inequality in China and India [15], but the identification of underlying drivers remains limited. In particular, long-term education distribution in China and India and its impact on income distribution are extremely under-studied. First, we provide long run education distribution statistics since 1907 using our annual enrollment and graduates data. We differ from the previous studies [16] by estimating the education inequality by birth cohort to better compare the education policies. Next, though we do not solve the problem of establishing a causal linkage between education and wage inequality, we systematically estimate the impact of education distribution on wage inequality.

The paper structure is as follows. In Section 2, we provide the details of the data sources and a detailed education policy review. Section 3 discusses the different strategies for education expansion. Section 4 focuses on the quantity-quality tradeoff while education expansion. Section 5 deals with the dynamics of education-wage inequality. Section 6 concludes the paper.

1.2. Context and Data

1.2.1. Timeline. The period from 1900 to 2020 has seen several changes across political, economic and social dimensions, all of which could influence the development and

[15] China: Wealth and Income Series: (Piketty, Yang, and Zucman, 2019); India: Wealth (Bharti, 2018); Income(Chancel and Piketty, 2017)

[16] Ram (1990), Thomas, Wang, and Fan (2001), Castello and Domenech (2002), and Morrisson and Murtin (2013) - estimate education inequality of the whole population (or adult population)

spread of the education system. The political situation is (possibly) the most important as the government was the most important provider of modern education. We divide the study period into three parts- 1900-1950, 1950-85 and 1985-2020 and briefly account for the existing political situation in both countries. We start from pre-1900 to better understand the starting point in both countries.

Pre-1900: Before the beginning of the 20th century, China was under the Qing dynasty since the 17th century. The traditional education system was linked to the state as it prepared students for the imperial civil service examinations to enter into bureaucracy. After the Opium War (circa 1839–1842), China was forced to open up (its port for trade) and western powers started creating their sphere of influence within China. On the other hand, India was under British colonial rule. Modern education started in India in 1813, but up to 1857, East India Company focused on expanding its territories. In 1858, the British government took direct control, and the next 50 years, which is termed a "Victorian-era", were relatively politically peaceful. The main features of this period were - financial stringency, government experimenting with different ways, and the start of publishing of statistical reports 1886-87.

1901-1950: This period was politically very turbulent at the world level (Bolshevik revolution, world wars, great depression etc.), creating financial stringency within China and India as both were still under colonialism. This period was also politically turbulent due to domestic factors in both countries. In China, the old mighty Qing empire fell in 1911, and the People's Republic of China (in short- Republic government) came to power, though the ruling remained mired with warlordism (1915-28). The end of warlordism (warlord fiefdoms and rival governments reunification) led to the creation of the Nationalist government, which experienced a Japanese invasion (1937–45) and a long-drawn Chinese Civil War (1927–49). In India, the freedom struggle was gaining ground with the involvement of the masses.[17] In response to these large-scale movements, the British government was ceding power to Indians gradually, which impacted education system.[18]. By 1935, the Education department was in the hands of Indians. In 1947 India

[17](Non-Cooperation Movement in the 1920s; Civil Disobedience Movement in the 1930s; Quit India Movement in the 1940s)

[18]The control of education departments was transferred to Indian ministers in all provinces, recruitment to the Indian Education Services was discontinued in 1924, and a new Provincial (Class 1) service was introduced.

gained independence, and in 1949, China was liberated.

1951-1985: The next 30-35 years was a period of communism in China and socialism in India. Both countries kept limited contact with outside world and followed a Soviet-style planning approach. In China, the education was nationalized and "education plans" were linked to the economic development. The major role of education was to contribute to the nation-building under the guidance of central government. The period of cultural revolution (1966-76) was a setback in the expansion of education and it is considered as a “Lost Decade of Education”.[19] In India, until 1975, education remained a state subject by Constitution, which meant state governments were responsible to manage education; this changed in 1976 and education was transferred to the Concurrent list (i.e. both Centre and State government can make laws on education). India came up with its first ever National Educational policy in 1968, proposed uniform structure of schooling for the country. In 1978, China opened its economy and in 1991 India followed. This has led to increased trade openness and globalization in both these countries.

1986-2020: By the 1980s, both countries were two well-established free countries, domestic political issues were getting sorted in a non-violent manner, the global competing forces of the cold war started dissipating, and political transitions became much more peaceful. Both countries started shifting towards a market-based economy where private sector involvement was encouraged in all sectors, including education. More and more students started going abroad for higher education, and many migrated to work after their studies.

1.2.2. Overview of Educational Policies. This section provides a detailed overview of the adopted educational policies for China and India, separately in chronological order. One can skip this long contextual subsection if the objective is to know the main results.

1.2.2.1. *China.* Pre-1900, the classic Confusion education was predominant in China, with its foremost goal to support the Imperial civil service examination.[20] The thousand years old imperial examination-based education system started to seem inadequate for

[19]Vocational and Tertiary education suffered, but Primary level education expansion wasn’t affected.

[20]The Imperial Civil exam was implemented as early as the Tang Dynasty (618-896) and had existed for more than 1000 years before its abolition in 1905.

the development of the nation, especially when China confronted western and modern forces. Two fundamental weaknesses of the traditional education system were - first, a narrow focus on Confucian study disincentivizing young talent from pursuing broad academic subjects, resulting in a retarded development of technology and a modern mentality. The second was the absence of public provision of basic schooling, thus keeping education inaccessible to many (Gao, 2015). The later part of the 19th century saw the diffusion of the European university model throughout the world under conditions of imperialism and colonialism. After the Opium War (circa 1839–1842), western education started in China through missionary activities. Supported by the Protestant and Catholic Churches, schools with Western-based curricula were established and increased gradually at all levels of education.[21] The goal of higher education was to educate talents in the areas of foreign language and military to meet the urgent need of the nation. It led to the creation of specialized colleges (more vocational) and the spread of liberal arts-based courses. The reach of missionary activities remained limited till 1900, nevertheless, they were the nucleus out of which the idea of modern education grew in China.

The following 50 years period was politically turbulent, with power changing hands several times. Despite this turbulence, there was a fair amount of continuity in the approach to the development of modern education. In 1904, Education Act laid down the general foundation of the first modern educational system in China. [22] In 1905, to incentivize modern education, the imperial civil service examination was abruptly ended, after more than 1300 years of its existence, marking the start of a transition to the modern education system officially.

The education reform took a quantum leap at the turn of the century to meet the rising challenges from the Western powers. Learning lessons from the past, the pivot of the education development plans was to increase the reach of modern education to all. The focus was more concentrated on primary education, where the promulgation and expansion of the coverage of universal compulsory education became the primary

[21]For example, St. John's University (Shanghai), one of the oldest and most prestigious universities in China, was established in 1879.

[22]The Education Act 1904 was a revised version of the Education Act 1902. The version of 1902 was never put into practice due to its unrealistic design.

tool. Initially, the compulsory education stage was conceptualized for four years, starting at age 7[23] which was to be implemented in 3 phases: providing one-year compulsory education to more than 80% school-aged children (1935-40); providing two years compulsory education to more than 80% school-aged children (1940-44); providing four years compulsory education nationwide (after 1944), (Sun, 1991, P423). [24] The expansion of compulsory education was made a core component of the local level administrative units with a clear goal of establishing one centre national primary school in each village or town and one national primary school in each Bao (Jiang, 1957, Ch 3, P115). [25] As a result, by 1945, China had 270K primary schools(32K centre national primary schools and 215K national primary schools). Furthermore, in 1947, six-year compulsory education was written into the Constitution for the first time.

During this period, the government's priority was to increase the enrollment rate, mainly at the primary level at all costs (such as establishing short term primary schools and half-day schools to name a few). The quality of compulsory education remained a second-order issue. A similar strategy continued later in Mao's period of PRC.

Another main criticism related to education during the 1920s highlighted was that liberal art, instead of science and technology, was overemphasized in higher education. For example, in 1928, more than 60% enrollment were in law and art courses. It led to course correction exercises in the 1930s, through legislation, changing the focus of the higher education from "teaching advanced academics and cultivating professional talents" to "teaching applied sciences and cultivating technical talents" .[26] In 1932, the push to develop a robust vocational education system came through passing laws to establish

[23]In 1906, the first compulsory education law was issued stipulating that "Children must go to school at the age of 7" (Li et al. 1995, Page 37). In 1912, the Ministry of Education was established, followed by the enactment of the "Primary School Law", setting the four-year elementary and primary school as the compulsory education stage (Jiang, 1957, Ch 3, P115).

[24]In 1935, the Nationalist government promulgated "Outline of Interim Measures for the Implementation of Compulsory Education", aiming to implement four years of compulsory education.

[25]The Nationalist government implemented a new local administrative system called - the "New County System" to strengthen its control over grassroots political power. Bao is an administrative unit consisting of 100 households.

[26]in 1929 the Nanjing National Government promulgated the "Republic of China's Educational Aims and Implementation Guidelines", stipulating that university and specialized higher education must focus on practical science (Wang, 1934, Ch3, P11). Following the guidelines, the focus of higher education started shifting towards science and technology.

separate and independent vocational schooling systems at the middle and higher level. [27] Further, the Ministry of Education prescribed the ratio of the number of classes in different types of secondary schools, balancing the enrollment in general and vocational education. [28] These measures led to the diversification of education during this period with exceptional attention towards the applied (professional) disciplines as well as vocational education towards a powerful industrialized nation and modern military prowess. [29]

The policies for primary and secondary education carried out during the Communism period resembled in many ways its predecessors. Popularizing compulsory education continued to be the priority of education [30]. The significant addition was that the expansion of secondary education was also to be accelerated.[31] The expansion of the primary and middle (general or ordinary education) continued even during the cultural

[27]The promulgation of "Secondary School Law", "Normal School Law", and "Vocational School Law" in 1932 led to the establishment of separate normal and vocational schools (Sun, 1991, P425). "Normal" schools are teachers' training institutes, hence a part of vocational education. "Vocational" schools are related to agriculture, industry and commerce. Before 1932, students could opt for different tracks in later years of secondary education.

[28]At junior high education level, the ratio among ordinary secondary school, normal school and vocational school is 6:3:2, and at senior high education level, the prescribed ratio was 2:1:1 (Jiang, 1957, Ch 3, P182).

[29]The narrow focus on Confucian study in the traditional education discourages the distribution and creation of knowledge or natural sciences and practical expertise, which is often believed to be the explanation of China's underdevelopment in military prowess and industrialization (Lin, 1995, Landes, 2006; Cantoni and Yuchtman, 2013).

[30]The first national primary education and demonstration education conference of the Ministry of Education (1951) proposed to enroll 80% of school-age children in primary school in 1952-1957 and provide universal basic education coverage within ten years. In 1956, the State Council passed the "1956-1967 National Agricultural Development Outline", proposing that from 1956 onwards, according to local conditions, compulsory primary education should be popularized within 7 to 12 years. In 1961, the Central Committee of the Communist Party of China approved the "Report on the Arrangement of Cultural and Educational Work in 1961 and the Future Period" by the Central Culture and Education Group, insisting that according to the different conditions of urban and rural areas, popularize primary education for school-age children. (See Zhang, 1984, P123)

[31]The first national secondary education conference (1951) concluded with urgency for medium technical talents for national defence construction, economic construction and cultural and educational construction. In 1963, the Central Committee of the Communist Party of China issued the "Regulations on Discussing the Work of Full-time Secondary Schools (Draft) and Instructions on Several Issues in the Current Education of Secondary Schools", proposing that "primary and secondary education should conscientiously implement the policy of 'walking on two legs and establish different types schools. The national organization of full-time primary and secondary schools is the major component of primary and secondary education. The government should strengthen leadership management for collective and individual schools and provide appropriate teaching materials. (See Zhang, 1984, P148-150)

revolution period - primarily through the widespread expansion of locally run or min-ban schools[32].

Accompanying the First Five-Year Plan (1953–1957), which heavily focused on developing the heavy industry, based on Russian experience and advice, the universities and colleges went through a nationwide large scale adjustment of faculties and departments after the 1950s. The Soviet-style higher education system was simulated in China to replace the British and American-style higher education systems adopted in the Republic of China. As a result, all private universities (including universities run by foreign churches) were abolished; training in engineering, teacher training, agriculture and forestry was given even more emphasis to promote industrialization; humanities and social sciences were overkilled.[33] Vocational education was also strongly emphasized through "two education systems, two labor systems". [34] The development of higher and vocational education was abruptly interrupted by the cultural revolution, during which radical affirmation action was taken to achieve equality for the poor and the uneducated. The adopted strategy was to cut off the top of the educational pyramid by lowering the quality and quantity of urban and tertiary-level education. The enrollment in university/college was stopped for the next six years; the enrolling of the graduate students was halted for 12 years. The impact was catastrophic, especially for vocational and higher education.

In 1992, the 14th National Congress of the Communist Party of China established for the first-time revitalizing China through science and education as the primary national policy. The period also saw an adoption of population control measures, which gradually impacted different levels of education by reducing the absolute number of enrollments.

[32] They were people-run schools started in the 1940s as voluntary village institutions.

[33] In May 1952, following the guideline of "focusing on cultivating industrial construction talents and teachers, develop specialized colleges, rectify and strengthen comprehensive universities", the Central Ministry of Education put forward plans for the adjustment of colleges and universities nationwide (Zhang, 1984, P251)

[34] In 1958, the State Council issued the "Instructions on Educational Work" and proposed to establish three types of schools - agricultural middle schools (in rural areas) and technical schools (in urban areas). In 1964, Liu Shaoqi proposed "two education systems, two labor systems", parttime-work-parttime-study education system. (See Zhang, 1984, P149, 180)

The focus of the development of the education policies now tilted towards popularizing higher education, though primary and middle levels were not neglected. In 1982, compulsory education was written into the Constitution of the PRC for the first time. In 1985, a series of laws regarding compulsory education was promulgated, transitioning compulsory education from 6 to 9 years.[35] The goal was to universalize nine-year compulsory education (Primary+Junior Low) nationwide by 2000. The expansion of higher education became central to setting clear targets and allocating more resources by encouraging private investment. Further, the vocational education system had to play a significant role in the expansion.

"To run education, we must walk on two legs, pay attention not only to popularization but also to quality" - Deng Xiaoping, May 1977, marks the policy shift from quantity to quality. The establishment of key educational institutions and strict examinations towards entry into these key institutions were some of the measures. [36] Meanwhile, due to the accelerating economic development since Chinese economic reform, demographic change under the birth control campaign in the 1970s, and the implementation of the one-child policy since 1980, the available resource per student improved, thereby improving the quality of education. In higher education, the main reforms provided more autonomy to universities and colleges. Also, the emphasis turned towards developing world-class universities in the 21st century through schemes like - the 211 Project in 1996, Project 985 in 1998, C9 in 2009.[37]

[35]"The Compulsory Education Law of the People's Republic of China (1986)", "Rules for the Implementation of the Compulsory Education Law of the People's Republic of China (1992)" "Education Law of the People's Republic of China (1995)"

[36]Deng Xiaoping. (1977b). Respect for knowledge, respect for talents (May 24, 1977). Central Committee of the Communist Party of China Literature Editing Committee. (1983). Selected Works of Deng Xiaoping (Volume 2). Beijing: People's Publishing House (1978) In January, the Ministry of Education issued the "Trial Plan for Running a Batch of Key Primary and Secondary Schools")

[37]By the year of 2010, there are in total 112 universities selected in Project 211 and 39 top universities selected in Project 985. In 2009, The C9 League was founded, which has been compared to other elite university groupings around the world, such as the Ivy League (US), Russell Group (UK), U15 (Canada), and Group of Eight (Australia).

The diversification of education continued with expanding higher vocational education[38] and developing disciplines of law, management and economics in degree programmes. The goal of higher vocational education was to cultivate a large number of specialized talents with both necessary theoretical knowledge and strong practical capabilities for economic development urgent needs in various fields. In 2014, the State Council proposed establishing the "Modern Vocation Education System", which features the strong interconnections between secondary and higher vocational education and between vocational and general education. In 2018, vocational education was officially endorsed to have equal importance as general education. [39] The past expansion of tertiary education had already expanded professional education (like engineering and science), hence with the increased institutional autonomy and more private participation, the emphasis is now on the under-developed disciplines like law, management and economics.

1.2.2.2. *India.* During the Victorian era in India (i.e. 1858-1900), the education system was guided by two important documents- Wood's Despatch 1854 and the Indian Education Commission (IEC) of 1882. One of the objectives of the education policies was to impart western knowledge (and culture) to the Indians, thereby creating a class of public servants. Though it was not the only aim[40], the low level of social and political awareness about formal education combined with the existing abject poverty and colonial domination made education a tool to gain economic employment in the public sector. The progress of education had to be carried out mainly through privately managed bodies, with the government playing the role of financier (through grants-in-aid), manager (through the creation of the Education department) and supervisor(through regular inspections and publishing reports). The religious neutrality and too much focus on "westernization" led to a gradual decline of indigenous forms of schooling (and Missionary-led education). There was much more attention towards the planning of secondary and higher education, and the responsibility of primary education was relegated to the local level bodies. The expansion of education was a significant feature on account

[38]Vigorously developing higher vocational education was announced for the first time by the Central Committee of the Communist Party of China and the State Council in its decision on "Deepening Educational Reform and Comprehensively Promoting of Quality Education".

[39]"The Decision on Accelerating the Development of Modern Vocational Education (2018)"; "National Vocational Education Reform Implementation Plan (2018)" by the State Council

[40]Wood's Despatch other objectives talk about promoting intellectual development, raising the moral character of the young generation, developing, spreading education among masses etc.

of *laissez faire* policy of the government. The material benefits associated with gaining degrees[41] led to a rush towards passing Matriculation examination and eventually getting Universities degrees. The growth of vocational education could not pick up, even though the policies enunciated the development of this type of education from higher classes of the secondary stage. [42]

The beginning of the century started with a big shift in education policy. The rapid pace of education expansion in the late 19th century, mainly through native Indians (evident from the Quinquinneal statistical reports), sparked the quality-quantity debate. The government attention was geared towards improving quality (through increased government control), whereas the Indian intelligentsia argued for continuing expansion.[43] First, the government policy now changed to take an active role in the provision of education. The government would now maintain "model" institutions at the primary and secondary level and begin providing grants-in-aid to collegiate education. Second, improving quality was to be implemented through several means- stricter conditions for affiliations of colleges, prescription for "recognition" of the secondary schools (by the Department of Education for receiving grants-in-aid and by a university for presenting its pupil at the Matriculation examination), prohibition of the transfer of students from unrecognized to recognized schools, increase in the inspecting staffs to enforce conditions of recognition, reducing PTR at primary level, increasing the salary of teachers, training of primary teachers, revision of curricula etc[44].

[41]The resolutions of Governor-general in Council of the 10th October 1844 gave a general preference to well educated over uneducated men in the admissions to the public service.

[42]Wood's Despatch in 1854 contemplated the provision of vocational instruction from the secondary stage, IEC 1882 recommended bifurcation of upper classes of high school, one leading to the University and the other to a more practical character, intended to fit youths for commercial and non-literary pursuits.

[43]Government Resolutions of 1904 and 1913 are the essential documents before the 1920s. Gopal Krishna Gokhle introduced a private bill on compulsory education in 1911 in the Imperial Legislative Council, which was not passed. A committee under Sir Phillip Hartog was appointed in 1929 to enquire about the organization of various aspects of education in India and suggest its overall improvement and progress. Its suggestions and recommendations influenced the government policies in later years.

[44]Government Resolution of 1913 mentions that no teachers should be called on to instruct more than 50 pupils, preferably the number should be 30 or 40, trained teachers should receive not less than Rs 12/month and they should either be eligible for a pension or admitted to a provident fund. Hartog Committee 1929 also emphasized the increasing inspecting staff, improving standard, remuneration and service conditions of teachers at both primary and secondary.

The neglect of the problems of primary education in the past was first accepted in the Hartog Committee report in 1929. Though it categorically condemned the policy of hasty expansion at the primary stage and proposed the policy of consolidation on account of enhancing quality. It highlighted the over-crowding in high school and collegiate education due to a lack of a reasonable selective system. It also pointed out excessive devolution of authorities to local government in the primary stage. It called for taking more control by the government and improving the quality (reducing wastage and stagnation).[45] The report proposed compulsory education (for four years) but without any haste, which led to the passing of compulsory education acts (for 4 or 5 years) in several provinces but covering predominantly urban areas and boys from 1921-to 37. It was the Sargent Report in 1944 that suggested increasing the compulsory education period to 8 years (from age 6-14) which is valid till today in India.

The spread of vocational education continued to suffer despite the government reiteration several times. Hartog Committee in 1929, too, recommended diverting pupils towards industrial and commercial careers through a more diversified curriculum in the middle-level vernacular schools and technical education in universities. The fast spread of secondary and higher education based on literary education (which meant employment in government) led to an issue of educated unemployment in the 1930s! The lack of professional education got much more highlighted due to Wars when more technical persons were required. Later in 1936-37, the British government had to call two experts from Britain to study and formulate the expansion of vocational education. [46]

After Independence, National Policy on Education (in 1968, 1986 and 2020; NPE in short) and Five Year Plans (from 1st to 12th) are the crucial documents providing insights on the adopted educational policies.

[45]Wastage implied students were not finishing primary stage and dropping out. Stagnation meant a repeat of the classes for more than one year. The reasons behind wastage and stagnation, according to the report, were illiterate parents, single teacher schools, lack of trained teachers, and poor methods of teaching.

[46]A. Abott and S.H. Wood's report came out after four months of a tour in Delhi, Punjab and the United Provinces on vocational education in India and made several suggestions. The most important was drawing parallels between general and vocational education and treating vocational at par with general education.

This period tilted the shift towards quantity, with massive expansion in enrollment numbers at all levels, allocating a larger share of resources in establishing the institutions. Under the quality reforms, teachers' quality was considered the most critical factor (NPE 1968), which led to increased emphasis on the teachers' training and their emoluments at all levels of education. [47] The second component of quality improvement is through establishing "model" schools and autonomous colleges as pace-setting institutions. [48] The third focus was on science education and research as it was considered an essential factor for the growth of the national economy, which led to the inclusion of science and mathematics as an integral part of general education till the end of the school stage.

The lack of seriousness toward primary level education continued for the next 40 years after independence. Unfortunately, no comprehensive study was undertaken on primary education as it was done for secondary and higher education immediately after independence. [49] The provision for the primary/elementary education in the first National Policy on Education in 1968 was also a simple reiteration of the existing Constitutional provision of free and compulsory education up to the age of 14 years,[50] and reduction of prevailing wastage and stagnation. The first 7 Five Year Plans (FYP) from (1951-to 90) kept re-iterating the goal but shied away from allocating enough resources or outlining a concrete plan to achieve them. [51] The rapidly growing population combined with relatively slow economic growth did not help either. Later, the goal was split into 3

[47] Mudaliar Commission 1952 also suggested improving the quality of teachers and recommended increasing the share of post-graduates for teaching at higher secondary schools.

[48] V Five Year Plans (1974-79) recommended establishing one model comprehensive secondary school in each district and one model primary school in each community development bloc. In addition, 10% of the institutions were selected at all levels for intensive development. It was quite similar to the policy in 1900 but argued using an analogy of the "seed-farm" technology with three steps- the first step is to establish a quality number of institutions, in the next step excellence percolating to a larger group of second-level and finally excellence generated in these two groups to spread in every educational institution.

[49] All-India Commission on Secondary Education under Dr A Lakshmanswami Mudaliar's chairmanship was set up in 1952-53 to examine the prevailing system of secondary education and suggest measures for its re-organization and improvement. Indian University Education Commission under Dr S. Radhakrishnan was established in 1948-49 for a similar purpose for higher education.

[50] It is essential to highlight here that the placement of free and compulsory education was made under Directive Principles of State Policy which does not make it a justiciable right. In simple words, the government can not be held accountable in the courts for not being able to implement the provision.

[51] The share of elementary education was 56% in the first FYP, which decreased to 35% in the II FYP and remained like that up to VII FYP (1985-90). III FYP states that "The progress in establishing new schools during the first two Plans was relatively greater in respect of middle and high schools than in the case of primary schools"

phases- universal provision of schools, universal enrollment and finally, universal retention - always focusing on educationally backward regions and classes, keeping in check the disparity. During this period, secondary education also had unplanned growth and suffered from lower resource allocation. The thrust was more on the expansion of higher education and research capabilities. It was done through an increased share of plan expenditure, strengthening science and technology, and setting up research centres.[52] The rapid development of higher educational institutions created a situation in that VI FYP changed its stance towards increasing coordination and maximizing their utilization.

The academic nature of the secondary schools (from class IX onwards) and the lopsided development of liberal education in higher education were well-known issues by now. Hence, the government announced several measures to diversify education. NPE 1968 emphasized education development for teachers training, agriculture, industry (technical education) and other workers through traditional, part-time and correspondence courses. Multipurpose schools were established on the recommendation of the Mudaliar commission in 1952.[53] All the FYP documents emphasized on the development of basic vocational courses starting from the secondary stage (class IX onwards), to increase vocational courses enrollment (after class X), but the enrollment share in vocational remained abysmally low.[54] 6th FYP (1980-85) reiterated that - "There has been an undesirable growth of facilities for general higher education, especially at the the under-graduate stage in arts, commerce and humanities, and in the consequent increase in the incidence of unemployment among the educated"

The period (1950-90) saw a massive expansion of educational institutions and enrollments at all levels of education, partly due to the increasing social demand for education and partly due to the adopted policies. The education remained very academic in nature, and the government became the foremost education provider. The diversification happened at the very top of the education ladder. The progress of vocational education

[52]Scientific Policy Resolution in 1958 was adopted, which established National laboratories, Indian Council of Agricultural Research, Indian Council of Medical Research, Indian Council of Social Science Research and Department of Atomic Energy, to name a few.

[53]The purpose of these schools was to provide terminal courses in technology, commerce, agriculture, fine arts and home science, intending to divert students into different walks of life and reduce the pressure on university entrance (Pg 443 of Naik, 2000)

[54]The share of enrollment at higher secondary remained less than 10% during the 1970s.

never took off, and the rush toward degree programmes continued.

The following NPE of 1986 emphasized the universal enrollment and universal retention of children up to 14 years of age, like all the previous government documents. However, this time government was ready to walk the talk. The planned allocation of resources started rising, and the government started several schemes. The major among them was Sarva Shiksha Abhiyan (SSA) and Mid-Day Meal (MDM) - the expansion of elementary education was now in mission mode. The opening of the economy started many foreign-funded projects in the 1990's - District Primary Education Programme (DPEP) in 1994, Mahila Samakhya Programme in 1998, and Janshala in 1998 - were the major ones. [55] Gradually, the Government of India took over these programmes. The Constitution of India was amended in 2002 to make elementary education a justiciable Fundamental Right, and the Right to Education Act 2008 was passed. The provision of an informal type of education (started gradually in the 1970s in V FYP) also helped expand at all levels, including open universities, distance learning and correspondence courses.

During this period, till the end of XI FYP, expansion continued in the education system, outshining any measures undertaken for quality. The GER at the primary level (up to class V) crossed 100% in the early 2000s, and the higher education system entered into the "mass" phase by 2011 (i.e. crossing the threshold of 15% GER). Several academic research started highlighting the poor learning outcomes of Indian students, which shifted the debate towards quality from XII FYP (2012-17) onwards. One significant departure from before is that the quality started to be seen from the learning-outcome based approach compared to the input-centric and credential-based approach before. The past strategies relied on increasing teachers' salaries (to attract better human capital), establishing training institutions for ample production of teachers, and training of teachers since independence. Unfortunately, these did not translate into the learning outcomes of

[55] DPEP was launched in 1994, assisted by WB, European Commission, DFID, the Netherlands and UNICEF. It was the main vehicle for the spread of primary education and was rapidly spread so that by 2000, it covered 50% of children in the primary stage in over 271 districts in 18 states (10th FYP, pg 5). Janshhala

children.[56] The introduction of ICT from the middle level onwards was another important feature during this period. Further, after the 2010s, there was an increased focus on the consolidation of institutions (as rapid expansion resulted in the opening of institutions working at low capacity, creating thinning of resources).

The vocationalization of education received a great impetus in policymaking during this period. However, the problem of never achieving the set target remained throughout.[57] This impetus came in the background of the opening up of the economy in 1990, changing nature of jobs, increasing prominence of the service sector, thus making the pre-existing system obsolete, and poor skill development in the country[58], a golden opportunity of the "demographic dividend", and increasing mismatch between supply and demand leading to higher educated unemployment issue. The first national policy - National Skill Development Policy, came in 2009 to guide the skill development strategies covering institution-based skill development.[59] Some other major reforms during this period were the early introduction of vocational courses[60], bringing the service sector into the domain of vocational education[61], standardization of skill qualifications to facilitate mobility from vocational to general education, and vice-versa. [62] There has been improvement in diversification in higher education concerning the expansion of professional disciplines. The government policy of opening the door for private players at a higher level increased access in specific disciplines that were lagging before, like

[56]The annual ASER reports and other studies highlighted the issue of learning outcomes at the primary stage of education.

[57]e.g. the NPE 1986 set a target of 10% of the higher secondary enrollment towards vocational streams by 1990 and 25% by 1995. XI FYP revised the target to 25% by 2011, but at the beginning of 2012, XII FYP highlighted that only about 4.8% of students are enrolled in vocational streams.

[58]XI FYP: "According to NSSO data, only 5% of the population of 19-24 age group in India have acquired some sort of skills through vocational education, compared to 96% in Korea.

[59]The formal skill development through vocational educational institutions was one of the aspects of the policy. It also covered non-formal, self-employment, and entrepreneurial development. Later, National Policy for Skill Development and Entrepreneurship 2015 was launched, superseding the 2009 policy.

[60]XII FYP proposed to begin vocational courses after eight years of education, instead of 10 years before.

[61]XI FYP: "Greater emphasis will be placed on the services sector and, therefore, on soft skills, computer literacy and flexi-time." Prior to that, the vocational education through polytechnics ′ year diplomas was related to conventional subjects such as civil, electrical and mechanical engineering.

[62]The expansion of vocational education and training had taken place in a very decentralized fashion. It has led institutions to have their own standards in terms of duration, curriculum, entry requirements, title, certifications etc., which created the problem of establishing equivalence of certificates/diplomas/degrees in different parts of India. The National Skills Qualification Framework comprises ten levels, each representing a different level of complexity, knowledge and autonomy.

engineering, management, medicine, and IT, where students are willing to pay substantial fees. It increased the diversification of disciplines in higher education. There is an impeccable growth in technical education after 2000.

1.2.3. Current Education Level Structure. The education system has gone through transformations, but the current education structure is as follows. We divide education life into three broad stages: Primary, Middle and Higher/Tertiary. In both countries, primary and middle are school level education for the first 12 years. It is equally split in China with six years of primary education and six years of middle education. Whereas in India, primary education is for five years, and middle education is seven years [63]. The middle level education in China is split into Junior Low and Junior High for three years each. Whereas in India, it is split into 3+2+2. The first three years after primary is called upper primary, the next two years are secondary (finishing with Matriculation examination - in the past, it was an exam to enter into University/College.), following two years of education are called Senior Secondary (which ends with Intermediate exam/12th exam which is now the entrance exam for college.) In both countries, there is an option to go through vocational education after 8/9 years of school education for 2-3 years of course. At the tertiary level, China demarcates vocational colleges providing diplomas and universities providing standard degrees (Bachelors, Masters etc.) Vocational courses are for three years, whereas the minimum years for the first standard degree (bachelor's) is four years. Master's is for three years, and doctoral studies for three years. In India also, there is this demarcation between vocational and standard degrees, but institutionally it is not so strong (partly because of lesser development of vocational studies). Bachelor's degrees can range from 3-6 years depending on stream, master's is of 2 years, and PhD takes a minimum of five years.

1.2.4. Data. We use statistical/administrative datasets to create the long-run series (1897-2020) of educational outcome measures.[64] Both countries have a rich tradition of producing statistical reports. In the last 150 years, both countries have gone through several politico-economic transitions. The challenge was to make a coherent time series. We

[63]NPE 1968 tried to create a 10+2+3 years common structure for all India, with the first five years primary, following three years upper primary and two years of secondary education. In the past, there were several differences; e.g. some provinces/states had four years of primary and some five years. The uniformity of education structure came gradually.

[64]For some later years, education surveys are used in India due to the unavailability of some statistics in the reports

also exploit expenditure reports, budget documents, and educational surveys to compute stage-wise expenses and get a public-private split. Finally, we exploit employment surveys from both countries to perform an education-inequality(wage) analysis.

1.2.4.1. *Educational Statistical Reports.* : Both countries produce a rich set of regular reports providing information on enrollment, graduates, teachers and expenditure. For China, we use "Compilation of Materials on Modern Chinese Education History"- general and higher education, Statistical Digest of the Republic of China 1935-1947 and Education Yearbooks. For India, the pre-independence period is covered by the "Progress of Education in India"- quinquennial reports and post-independence from "Education in India: Annual Reports", UGC reports, AISHE (2010 onwards). We extract information such as enrollment, graduates, teachers, and expenditure (if provided), from these reports by different stages (Primary, Middle and Tertiary). In China - primary, secondary and higher-level schools are properly differentiated, i.e. the schools are synonymous with the stage of education. However, in India, schools can have mixed stages. E.g. primary stage students can be studying in secondary or senior secondary school type [65]. Further Intermediate stage (which are class XI and XII) were part of college studies for a long period [66] and gradually integrated into school education. Hence important care has to be taken in comparison at stage-wise. Total enrollment and graduates are readily available in the reports. However, teachers, expenditure, and public-private distribution are usually present at the school-type level instead of the level of education. We impute the total teachers and expenditure at the primary stage by adding the numbers present in primary schools and an estimated (number) of the teachers teaching primary class students in non-primary schools. The imputation is based on the assumption that statistics like teachers per student and expenditure per student in primary school are the same for primary stage kids in non-primary school types. Please look at appendix C.2 for details about data and appendix C.4 for variable creation.

1.2.4.2. *Surveys.* : We use standard nationally representative surveys for our education-inequality analysis. They provide information on completed education level (degree),

[65]The categorization of a school into primary, secondary and tertiary depends upon the highest standard in the school. So a school up to class 10^{th} is called Secondary school, up to class 12^{th} is called Senior Secondary school

[66]National Education Policy of 1968 recommended Class XI-XII to be part of school education

wage earnings and other demographic characteristics. For China, we use CHIP (Chinese Household Income Project) datasets, and for India, NSS's (National Sample Survey) Employment and Unemployment thick round of surveys and Periodic Labour Force Survey (PLFS). We restrict the sample to a population between 20-60 years (working age) old, having regular salaried jobs and positive income for the inequality analysis. For China, we are left with sample size of 18,411, 12,091, 15,784, 21,055 and 20,200 for the years 1988, 1995, 2002, 2013 and 2018 respectively. For Indian surveys, we have 34,952, 30,332, 35,495, 37,660, 39,792, 34,992, 37,645 and 40,665 sample size for the years 1983, 1987-88, 1993-94, 1999-00, 2004-05 and 2009-10, 2011-12 and 2018-19 respectively. We create three categories of education, namely- primary, middle and tertiary and compute the log of annual and daily wages. [67]

1.3. Progress of Modern Education System : A Tale of Two Countries

The phenomenal expansion of the education system during the 20th century in both countries depicts an exemplary case of the "Human-Capital century" (Goldin, 2001). Table 1.1 presents the average values of enrollment, teachers, expenditure and tertiary-level graduates by period. E.g. the total enrollment in China has gone up from an average of 5 M during 1900-25 to 227M during 2000-18. Similarly, it went up from 6 M to 253 M during the same period in India. Both countries' current gigantic education system absorbs billions of dollars, employs millions of teachers, and generates millions of high-skilled workforce every year. There are significant underlying differences in the trajectories behind the veil of similarity of the overall expansion. We shed light on the differences in education development strategies through analysing the long time series of educational statistics and adopted educational policies.

1.3.1. Bottom-Up vs Top-Down Expansion. We make a case for bottom-up versus top-down strategies analysing long-run evolution of enrollment, expenditure allocation and teachers' recruitment at different education levels.

1.3.1.1. *Total, Gross and Net Enrollments by education level.* Primary education is defined as the first 5/6 years of education. [68] Top part of Figure 1.1 presents the evolution of total enrollment in the primary stage. At the start of century (in 1907), benefiting from

[67]Since Indian surveys capture information on weekly wages, we compute annual wage as simply as 52*Weekly wages

[68]The definition slightly varies across the years and regions. Since we take our data from government reports, we are limited by the inconsistencies present in those reports.

the head-start of modern education, India had 4.3 M [69] enrollment compared to 0.87 M in China. In the 1930s, India lost this lead due to the rapid expansion of primary level enrollment in China. The continuation of the faster expansion resulted in China having 10M more enrollment at the primary level by the time both countries gained freedom from colonial subjugation. [70] [71]

China maintained its lead in the second half of the 20th century despite internal politico-socio-cultural hiccups. China had 40M more enrollment than India by 1985. The highest primary stage enrollment in China was in 1976, with 147 M enrollment compared to 66M in India (less than half of China!). It is important to emphasize that the cultural revolution period of 1966-76 in China did not deter primary stage mass expansion. The Net Enrollment Rate (NER) was more than 90% (Bottom part of Figure 1.1) by 1985, i.e. before the economic liberalization of China. India continued with its slow but steady growth. The peak of enrollment in India came much later in 2011, with 140 M enrollment, but NER was lower than 90%. In both countries, population control measures have led to the arrest in the total enrollment figures (2016- India 124M; China 99 M) at the primary level. [72].

The top part of Figure 1.2 shows the evolution of total enrollment at the middle stage. The pattern is similar to in primary level - India starting with a higher level of enrollment, China catching up and finally, population control measures arresting enrollment growth. The main difference is that the lead of India is maintained till the 1970s. The catching-up by China came 40 years later (1930's for the primary stage). China had 17 M more enrollment in the middle stage by 1985 (once again cultural revolution had no significant impact on the middle level enrollment). Looking at the gross enrollment rate (GER) we see the same trend- India having higher GER at the middle level till the 1970s. In recent year data, the GER is higher for China at the middle level (CH: 90% and IN:70%) (Refer Appendix Figure 1.12) The impact of population control is seen in the

[69] For India pre 1900; 0.6M in 1871, 2.1M in 1881, 2.8 M in 1887, 3.1M in 1892, 3.4M in 1897 and 3.6M in 1902

[70] One extra year of schooling in China at the primary stage is not enough to make up this huge gap.

[71] The net effect of the partition of India in 1947 was a reduction of 3 M in enrollment in 1947. The independence of India came with the split of India into Pakistan and Bangladesh. Also, several princely states that were not part of British India became part of the new India

[72] China adopted One-Child Policy in the 1980s, and India achieved replacement level fertility in Census 2011

decline in the numbers after 2000 in China. India is still having an increasing curve and will remain so for the next decade due to late population control and higher primary enrollment (feeding into the middle level).

The bottom part of Figure 1.2 shows the evolution of total enrollment at the higher stage. The catching-up comes 30 years after the catching-up in the middle stage or 70 years after the catching-up in the primary stage. China overtook India in the 2000s with a massive expansion of tertiary/higher level enrollment. In China, the number increased from 7.5 M in 1998 to 26M in 2003. For the last year of data, China has 18M more students enrolled in higher education (HE) than India. The enrollment is increasing at the tertiary stage in both countries and will continue to do so in the coming years. The gross enrollment rate (GER) at tells the same story. India had higher GER at the tertiary level till the early 2000s. In recent year data, the GER is higher for China at the tertiary level (CH: 30% and IN:23%) level.

The trend of the expansion of total teachers is similar to the trend of enrollment. The upper part of Appendix Figures- 1.13; 1.14 and 1.15 present the total number of teachers in the primary, middle and higher stages of education from 1912-2018.[73] China started having much more teachers at the primary level in the 1930s and middle level in the 1970s.[74]

The pattern of the catching-up of China in 1930 (for primary), 1970(for middle) and 2000(for higher) indicates a **bottom-up** mode of expansion of the education system in China. It is important to note that reduction of gender gap (at primary level) in China, as a potential reason behind the observed pattern of bottom-up, is valid only between 1965-75. By 1976, the last year of cultural revolution in China, the proportion of girls was close to the population share (Upper part of the Appendix Figure 1.20). Pre-1950, if anything, India had a slightly more female share. If we ignore some aberrations, both countries have nearly similar shares until 1967. The divergence started only after 1967, when China already had 80% NER.

[73]Because of the presence of mixed institutions in India, we estimate the teachers at Primary and Middle stage. The details are present in Appendix C.4

[74]At the higher level, China has more teachers than India since 1950 (i.e. prior to 60 years, China surpassed the enrollment in higher education).

1.3.1.2. *Expenditure allocation by education level.* Figure 1.3 plots the total (and public) expenditure as a share of Gross National Income (GNI). Both countries were spending less than 1% of GNI in 1930. In the next 20 years, it increased in China but decreased in India, creating a gap of 1 percentage point, which remained till mid-1960. The period of cultural revolution saw a reduction of the share in China, thus overturning the gap in favour of India. In recent years, the total expenditure stands at 6% in India and 5% in China. The pattern remains similar looking at the public component only - the slowdown in India during 1930-50, the slowdown in China during 1966-76, and higher spending in India after the 1960s. The similar trend of the total and public expenditure is not surprising because of the predominance of the public share in total expenditure (See Appendix Figure 1.16 for public-private share). In recent decades, the public spending in both countries has been $\sim$4.5% of total GNI.

The share of total expenditure going to different stages of education (Figure 1.4) supports the claim of bottom-up and top-down strategies. At the beginning of the 20th century, the expenditure share was highest for the primary level mass education and gradually shifted towards the middle level from the late 1970s in China. The difference between the expenditure share in the primary and middle stage remained positive till 1976, with a narrowing gap - 44 percentage points (pp) in the 1910s; 14pp in the 1950s; and almost 0 pp in 1976. Post-1980, the spending share is higher in the middle stage. In contrast, the expenditure share in India's primary and middle stage is very similar at 40-45% till the 1950s and 35-40% till the 1980s. The share has always remained slightly higher in the middle stage.

In the pre-1950 period, the allocation to primary level was 60% (of total expenditure) in China compared to 40% in India; middle-level allocation was higher in India at 51% compared to 25% in China, and tertiary level allocation was higher in China (14%) than in India (8%). It highlights the greater emphasis on middle-stage education in India. During 1950-85, the resource allocation to primary in China kept declining, with a corresponding rise at the middle and higher stages. The average allocation during 1965-85 in China declines to 39% at primary, increases to 39% and 21% at middle and tertiary levels, respectively. In India, between 1950-85, there was a huge increase in allocation towards the tertiary level. It increased dramatically from 8% before 1950 to 28% during 1965-85.

The primary and middle share both declined during this period relative to pre-1950. It highlights the development of the higher education sector in post-independence India, neglecting the primary level. Finally, during 1986-2018, the allocation share was stable. The allocation towards primary, middle and higher levels remains at around 30%, 39% and 30% in China and 33%, 45% and 23% in India. The expansion of higher education in China after the 2000s does not change the relative share at primary and middle levels, which shows that the expansion of higher education did not come at the cost of primary or middle levels. (Refer Table 1.2).

To summarize briefly, in India, there was more focus on the middle stage in the first half of the 20th century. Later, post-independence, there was an increasing share of expenditure going towards higher education, when the primary level was expanding, and NER was close to 60%. This skewed expenditure pattern relates to the **top-down** model of expansion of India.

1.3.1.3. *Educatin and Nation Building.* The circumstances under which the modern education system originated had a significant role in the fast/slow expansion of the primary level mass education. In colonial India, the objective was to produce a small western-educated workforce to help run the country's administration. It led to several policies unfavourable to the primary level mass education. First, it meant more years of education. Four-five years of education was not enough to equip someone for handling public administration. There was also a general disinterest of the colonial government towards a major expansion of primary level education (due to financial stringency), which is evident through lesser debates on primary education, lesser reforms, and transferring of responsibility to local level bodies (without resources). Second, the educated workforce was required to be know English (official language), a language foreign to Indians. This was implemented by adopting native languages as a medium of instruction at the primary level and English at later stages of education, thereby creating a structural break. Appendix Table 1.8 shows that almost 40-50% of the schools in the lower-middle stage (Grade VI-VIII) were of English medium. This structural break must have acted as a deterrent to enrollment at the primary level. Third, the traditional education system was wholly neglected and liquidated by 1900 instead of their transformation (Naik, 2000). Finally, the government efforts were also mirrored by the native Indians who came from the higher caste, and education became a tool to acquire government jobs.

Despite the introduction of modern education in China almost half a century later than in India, the development and expansion of modern education reflected the will of a nation to employ education as an important vehicle for achieving national development and military advance from the very beginning. The new system emphasized getting rid of the inaccessibility of education by establishing schools in all villages and implementing compulsory education laws. Just like in many other post-colonial nations, in India, not until its independence (1947) was education granted the nation-building mission. The compulsory education law was first introduced in China in 1906. In contrast, in India, it started in the 1920s. However, the coverage and enforcement remained limited till the 1950s [75] in both countries. The decentralization of primary education provided an opportunity to the local elites to take up the role of providing primary schools as a way to continue their social status [76]. The number of primary schools in China by 1945 (270K) became double that in India (140K).

Other than this, there were influential domestic factors in India which inhibited the growth of primary level mass education. Empirical analyses [77] have highlighted high (caste and religious) diversity in India combined with the decentralization of primary school management as one of the causes behind the poor provisioning of primary schools in the early 20th century. The socio-demographic factors like one language more homogeneous culture also benefited the spread of primary level education faster in China.[78]

[75]Interestingly, compulsory education was introduced in England in 1870, and by 1902 it was effectively enforced in all parts of the country.

[76]Gao, 2015 shows that counties which had a higher proportion of gentry, i.e. traditional scholars who had passed Civil Services Exam with a degree, increased the provisioning of primary schools. Also, see Chaudhary et al., 2012 which shows that in India, the provinces where elites were non-landed, the provisioning of primary schools was higher.

[77]E.g. Chaudhary, 2009 show that districts with high religious and caste diversity had fewer total (mainly primary) schools. The effect is due to lower provision of private primary schools in culturally diverse districts.

[78]Alesina et al., 2003 and Fearon, 2003 both rank India higher than China for ethnic, cultural or linguistic diversity. Fearon, 2003 ethnic and cultural fractionalization score for India is 0.811 and 0.667 compared to 0.154 and 0.154 for China respectively. The score for ethnic, linguistic and religious fractionalization is 0.42, 0.81 and 0.33 for India compared to 0.15, 0.13, and 0.66 for China Alesina et al., 2003.

Post-1950, the main distinction between these two countries of foreign ruling versus domestic ruling became absent. After the end of colonial domination, both countries included compulsory education in their constitution.[79] The main difference was that in India, it was a non-justiciable right, i.e. state cannot be brought to the court for its non-implementation. It was only in 2002 that it was upgraded to Fundamental right[80], i.e made justiciable in nature. Lastly, after 1982, China introduced nine years of compulsory education, one year more than in India. The independent Indian government's educational policies could have re-directed more resources towards mass-level primary education. However, one only observes a modest and gradual improvement for the next 40 years. Indeed, the imbalances of the last 100 years (i.e. 1857-1947) created a top-heavy education system (Arnove, 1984), but the independent government shied away from allocating a higher share towards primary education. One observes some improvement in the first 15 years (increasing relative primary share). However, the focus changed towards higher education in the early 1960s with an increasing allocation towards higher education, thus diverting the resources away from the weakly developed primary education system. Liberated China successfully universalised primary education by implementing compulsory education laws (increasing it to 6 years and later to 9 years).

[79]China inserted 6-years of compulsory education in its constitution. India: Article 45, as part of Directive Principles of State Policy, provides free and compulsory education for children until the age of 14 years.

[80]86th Amendment to the Constitution of India inserted Article 21A, and later Right to Education 2009 law was made for its implementation.

1.3.2. Diversification of Education. This section deals with broadly two aspects of diversification - development of vocational education and expansion of different disciplines at higher level of education.

1.3.2.1. *Vocational and Non-Vocational.* Vocational education and training are considered an integral component of UNESCO's global Education for All initiative. [81] It defines vocational education as the education/training which aims to equip people with knowledge, know-how, skills and/or competencies required in particular occupations or, more broadly, in the labour market. In both countries, broadly there are middle and higher level vocational education (parallel to non-vocational or standard education). The start of vocational education has gone through several changes in both countries.[82] Due to changing of start point of vocational education and data availability constraints[83] we combine the middle and higher stage vocational students for analysis.

China sends a much larger proportion of students towards vocational education than India. The vocational share was close to 80% at the start of modern education in China, highlighting the importance of the vocational track in the origin of education (See Figure 1.5). The share declined gradually to 20% by 1950 as non-vocational education expanded. In India, the share was close to 5% in the pre-1950 period. Post-1950, there was a surge in the vocational share in both countries in the initial years but it did not sustain. In the first fifteen year, i.e. 1950-65, the vocational share increased and remained close to 30% in China, much higher than in India (7%). In the following decade, Cultural Revolution (between 1966-1976) brought a almost complete stop of the vocational education in China, while vocational education remained unpopular and the share declined in India. Post-1980's, with opening of the economy, there was a strong resurgence in the vocational education track in China. It came from the middle-level vocational upto 2000 and from the tertiary-level vocational after 2000. Today almost 25% of the students in China are enrolled in vocational track out of total middle and higher stage combined, which is 2% in India. The higher level of vocationalization of Chinese educational system could

[81] Education for All is a global movement by UNESCO aiming to meet the learning needs of all children, youth and adults.

[82] In China: Post-1980, vocational education starts after junior low, i.e. after nine years of compulsory education. Before 1945, vocational education could have started after primary education. In India, currently, vocational education could start from class IX (i.e. after eight years of compulsory education). NPE 1986 had suggested that vocational education start after class X.

[83] E.g. the Quinquennial reports in pre-independent India provide combined figure (middle and higher) for vocational schools.

also be seen from the absolute enrollments. In 2017, China had 30.4 M students enrolled in vocational education compared to 3.3 M in India. If one looks within tertiary-level, close to 50% of the students go towards the vocational track, which is 40 percentage points more than India.[84] The expansion of higher education in China after the 1980s is driven by the expansion of vocational education.

In pre-1950 China, the development of vocational education is closely related to the objectives of the modern education to develop industrialized and modern military prowess. At middle-level, the educational policies provided clear guidelines, like separate establishment of vocational schools in specific ratio to non-vocational education to name a few, to develop vocational education. At tertiary-level, Qing's dynasty in the late 19th century established specialized colleges, to develop military needs of the country, which formed the backbone of tertiary level vocational education system in the 20th century modern education. In 1932, passing of Vocational Education Law formed an independent vocational education system at tertiary-level in China.

In pre-1950 British India, the main drive for creating vocational education was to reduce the flow of students towards university education and not industrialization. In the 1920s, unemployed graduates started surfacing in the reports as one of the problems of the education system (Hartog Committee 1929). Under the colonial policy, India was seen as a raw material provider for the manufacturing industry in Britain and a consumer of the finished products (Wood's Despatch 1854).[85] Further, the education remained limited to well-off population (mostly upper caste and class which aimed to get service under Government) where demand for vocational training was low. [86] Several

[84]The difference is much starker, as the data for India also include some secondary level vocational enrollments.

[85]"...secure to us a larger and more certain supply of many articles necessary for our manufactures and extensively consumed by all classes of our population, as well as an almost inexhaustible demand for the produce of British labour."

[86]Other paramount factors impeding the growth of the vocational education in the colonial period are-expensive nature (possibly one of the reasons behind government "model" schools not including vocational courses at secondary stage, thereby no example to replicate for privately managed institutions to follow); and knowledge of English acting as a vocational course in getting employment; industrial education catering primarily to Europeans in India and to the Anglo-Indian community; a bureaucratic orientation which was staffed with men from liberal arts and technicians were looked down upon, lack of attention from the universities in India, no curriculum reforms, lack of a concrete colonial policy and neglect of indigenous systems of vocational learning (Naik, 2000, Singh, 2001).

of these factors continued to hamper vocationalization post-independence in India.

Comparing the vocational education and training system in China and India post-1990, Mehrotra, Gandhi, and Kamladevi, 2015 highlights that the success of China is - the decentralization of vocational education management,[87] presence of state-owned enterprises ensuring industry participation in the VET system, mandatory participation of industries through Vocational Education Law in 1996, better teachers training and recruitment system[88], and financial assistance to students and making tuition-free after 2009 at middle-level vocational education. Many of these features have also been adopted in the new policy measures undertaken in India.

1.3.2.2. *Tertiary Level Disciplines: Engineering vs Humanities.* In this sub-section, we analyse the diversification within non-vocational or standard degree programmes (leading to Bachelors, Masters and PhD) in different disciplines. Bachelors remain the predominantly offered degree in both countries, though the share is declining as more and more students are continuing to Masters/PhD. [89]

We create eight comparable broad categories of different disciplines pursued in higher education- Humanities, Law, Education, Science, Engineering, Medical, Agriculture and Others. The humanities category is probably the most heterogeneous category with sub-disciplines like history, philosophy, economics, geography, and MBA/BBA, to name

[87]The course contents of most of the vocational subjects are decided by a mix of national, local level government and industry participation whereas in India it is decided centrally. The other one-third of course content is general education (nationally decided; ensures mobility from vocational to general), one-third nationally decided on certain trade-related content and one-third trade-related content but locally decided. (Mehrotra, 2014)

[88]There are strict guidelines in China which require teachers at vocational secondary schools to be at least vocational graduates, and those only with post-graduate vocational degrees and the respective occupational certificate can teach at vocational, undergraduate classes. In India, in the ITI system, most of the trainers were merely ITI graduates.

[89]Figure 1.17 provides the evolution of shares of Bachelors, Masters and PhDs by enrollment and graduates. After 2000, close to 80% Bachelors, 19% Masters and less than 1% PhD degrees are offered in India, compared to 89%, 10% and 1.5% in China. In China, before 1980, almost all the graduates were at Bachelors level with Masters remaining below 1%. Due to the top-heavy structure of the education system of India, even in pre-1950, close to 8-10% Masters degrees were awarded. To keep in mind, absolute numbers at all the levels in India have always remained higher because of - the expansion of higher education earlier and a larger share of students pursuing degree (non-vocational) courses.

a few[90]. The rest of the categories are self-explanatory; "Others" include all the sub-disciplines which cannot be clubbed in the existing categories. Figure 1.6 provides the share of graduates from these disciplines in both countries. There is a stark difference in the type of graduates both countries produce.

The brain-drain of the top-notch engineers from the 1980s and the impressive growth of engineering discipline over the last 10-15 years in India has created the perception of India being the *land of engineers*. Comparing the share of engineering graduates in the last 120 years shows that China has consistently produced a much higher share of engineering graduates than India. China produces $\sim$35% Engineering graduates every year compared to 15% in India today. The share of Engineering graduates was less than 5% before 2000 in India.

The share of Science graduates is higher in India than in China throughout the period. The shares have fluctuated around 10% in China and 20% in India. While comparing the share of Education graduates, the situation reverses. It is higher in China, hovering around $\sim$ 15-20% compared to 8-10% in India. Finally, the share of Medical graduates is much higher in China at 10-12% compared to just 2-3% in India.

The largest share of graduates in India comes from Humanities. 60% of total graduates belong to Humanities, compared to only 20% in China today. Further, the share of Humanities graduates has remained relatively high for the entire duration. Splitting the Humanities category into Arts (leading to Bachelor/Masters in Arts) and Commerce (leading to Bachelors/Masters in Commerce), the two significant streams which are combined for comparability with China, shows that Arts graduates have declined from 65% in 1897 to 34% in 2018 and Commerce has increased from 0% to 21% in 2018.

The share of Law graduates has seen a considerable decline in both countries. In India, the share of Law graduates used to be around 20% at the beginning of the 20th century, which has dropped to 1-2% today. In China, the share of Law graduates used to be 35% in the 1930s and has dropped to 5% today. Another stream which has seen a

[90] In the Indian context, we club Arts and Commerce and in China, we club Humanities and Language for comparability

consistent decline in Agriculture. The drop is starker in China, from 15% in 1912 to 2% today. In India, the share of Agriculture graduates has remained 1-3%.

In summary, lack of vocationalization and a lopsided development of humanities in the non-vocational category have remained the two prominent features of the Indian education system. The over-reliance on the humanities courses is attributed to - the continuation of the colonial legacy and accommodation of the surge in higher education in the 1960s in India through cheaper modes of education. The expansion of commerce courses is also partly due to its less expensive nature. The expansion of engineering and other professional subjects started only post 2000, primarily through private sector involvement. On the other hand, China has diversified more into vocational education and more professional course disciplines in higher education.

1.3.2.3. *Education-Growth.* The combination of more engineering and vocational students (combined with more educated mass of population) possibly helped China to generate the human capital that was more apt for building a manufacturing base (apart from the trade openness and other policy measures). Whereas India wanted to increase its manufacturing sector, it was and even today is restricted by the type of human capital it generates. In this paper, though we do not perform any causal analysis, we support the argument through an analysis of recent literature on the importance of the composition of education crucial for the economic development.

The divergence in economic development between the two nations started in 1990. Based on the World Bank data, China's GDP per capita (in PPP terms) was at par with India until 1990, while by 2020, China's GDP per capita became more than 2.5 times that of India (17211:6504).

The composition of human capital and its impact on growth emerged very recently, after more than two decades of debates on the importance of education on growth. Recent studies suggest that the composition of human capital plays a critical role in explaining economic growth. Motivated by the idea of division of labour, Joshua, 2015 shows that after considering the composition of human capital and imperfect human capital substitution, human capital variation can account for the large income differences between rich and developing countries. (For detailed discussion see Jones, 2014;

Caselli and Ciccone, 2019).

Regarding the debate on whether mass education or elite education is more growth-enhancing, several empirical studies suggest the impacts of an increase in different stages of education (primary, secondary and tertiary education) vary according to the level of a country's development. In particular, while primary and secondary education appear to be related to growth in the poorest and intermediate developing countries respectively, it is tertiary education that is important for growth in developed countries.[91] Vandenbussche, P. Aghion, and Meghir, 2006 proposed an endogenous growth model which separates the contribution of human capital to productivity growth into a level effect and a composition effect. They show that holding the composition of human capital constant, an increase in aggregate level is always growth-enhancing. However, holding the level constant, growth-enhancing properties of human capital depend both on the composition and distance from the technological frontier. In particular, higher education investment should have a bigger effect on a country's ability to make leading-edge innovations. In contrast, the focus on primary and secondary education seems warranted for developing countries.

Similarly, existing literature shows that a country's optimal education policy in the form of subsidies for general education vs vocational education should depend on its distance to the technological frontier. General education is always more growth-enhancing when the country is closer to the frontier, whereas a larger emphasis on vocational education is more growth-enhancing for the countries that are farther from the productivity frontier (see D. Krueger and Kumar, 2004; Vandenbussche, P. Aghion, and Meghir, 2006 Aghion et al., 2009). This thread of literature implies that in the early phase of the development of a country, a bottom-up model of expansion combined with a strong vocational education system could be more growth-enhancing than the top-down model with limited vocational education development. However, such findings are inconclusive. Another strand of study provide evidence of increasing higher education having stronger effect on growth compared to primary and secondary education (see Gyimah-Brempong, Paddison, and Mitiku, 2006; Castello-Climent and Mukhopadhyay,

[91](See, Wolff and Gittleman, 1993; Gemmell, 1996; McMahon, 1998; Petrakis and Stamatakis, 2002; Sianesi and Reenen, 2003; Papageorgiou, 2003; Self and Grabowski, 2004; Pereira and St. Aubyn, 2009)

2013; Castelló-Climent, Chaudhary, and Mukhopadhyay, 2018). Thus, further research is still needed before we can make any conclusion.

Another debate on education composition and economic growth is about talent allocation among different disciplines in higher education. The common belief that teaching and research on science and engineering in higher education will drive economic growth has been widely accepted (Woodhall, 1992). Murphy, Shleifer, and Vishny, 1991 shows that countries with a higher proportion of engineering college majors grow faster, whereas countries with a higher proportion of law concentrators grow slowly. Meanwhile other studies identified that engineer and engineering-minded technicians are the key to invention and transfer of technology (see Romer, 1990; Mokyr, 2005; Hanson, 2008; Toivanen and Väänänen, 2016; Maloney and Caicedo, 2017). Our results show that in the second half of the 20th Century, there was a great expansion of engineering and education disciplines in higher education in China. In a very drastic comparison, in India, humanities and law students account for more than 60% of the enrollment since 1897. The literature suggests that the difference in allocating the talents between China and India could contribute to different economic development paths of the two nations.

1.4. Choice between quality and quantity

In the planning of the education system, both quantitative and qualitative aspects are important. Under given resource constraints often one is chosen over the other. This section tries to understand the choice adopted by China and India at different stages over time. The quantitative aspects are relatively easier to measure- like gross enrollment ratio. On the other hand, the educational quality measures are more difficult to measure. There are broadly two ways to measure quality- input-based standards (such as pupil-teacher ratio, quality of teachers and classroom infrastructure) and outcome-based measures (such as cognitive skills and test scores) (Azam and Kingdon, 2015). The limitation of data availability for a comparable outcome-based measures for the long time frame under analysis, restricts us to create input-driven measures.[92]

1.4.1. EIR measure and its Decomposition. We create a measure - Education Investment Ratio (EIR) which takes into account *economic evolution* and *demographic transition*.[93] It is comparable across years and countries (standardizing with economic and demographic factors). The EIR is defined as the ratio between total education investment per child population (age between 6 to 22 years old) and per capita national income. [94]

$$EIR = \frac{\text{Total Expenditure/Population}_{6-22}}{\text{GNI/Total Population}}$$

where GNI is gross national income. The EIR has increased in both the countries and the value of EIR has remained higher in China than in India till the 1960s (See Appendix Figure 1.18. Between 1930-50, the EIR in China increased from 3% to 8%, whereas in India, it stagnated at around 2.5% during this period.[95] The gradual increase in India started from post-independence but remained below 8% up to 1985. The slowdown

[92]China and India do not participate in the Trends in International Mathematics and Science (TIMSS). The participation of China and India in OECD's Programme for International Student Assessment(PISA) is not representative of big countries like China and India. India participated only in 2009, where only two Indian states participated. Similarly, China participated in 2015 and 2018 with regions from Beijing, Shanghai, Jiangsu and Zhejiang.

[93]Post 1970, a fast pace reduction in the fertility rate in China, created the divergence in the demographic structure of these two countries. Hence not taking this feature into account, could be misleading.

[94]For instance, if the total education expenditure is equal to 4% of GNI and the children population is equal to 20% of the total population, then EIR will be equal to 4%/20% = 20%. Intuitively, this means that each child receives an equivalent of 20% of per capita national income in the education investment, i.e. the equivalent of a 20% part-time teacher paid at per capita national income.

[95]This stagnation is crucial because this period saw an increase in the expansion of enrollment and teachers, which implies a deterioration in the quality component of education.

during 1966-76 reduced EIR in China, and both countries had a similar level in 1985. The next difference in these two countries appears after 2000 when it increases in China and drops in India. The drop in India is driven by relatively lesser allocation towards education relative to the growth in GNI per capita. In China, within a decade between 2000-15, EIR doubled from 12% to 25%, whereas in India, it increased from 15% to 18%.

It is a very simple measure to understand education investment (relative to the economic and demographic size) in a country. A higher value means more investment. Though if countries are at different levels of education development then this measure in itself will be less informative. E.g. two countries where in one all kids (in 6-22 years) go to school (say at x cost per student) versus in other only 50% go to school (but with 2x cost per student) could have the same value, everything else identical. Since China and India had not only different starting years for education development but also they adopted different strategies (bottom-up versus top-down), we compute EIR at primary, middle and higher level separately. Further, in order to understand choice between quantity-quality, we decompose EIR at each level as follows:

$$\begin{aligned}
EIR_P &= \frac{\text{Expenditure}_P/\text{Population}_{6-11/12}}{\text{GNI}/\text{Total Population}} = \underbrace{\frac{\text{Enrollment}_P}{\text{Population}_{6-11/12}}}_{GER_P} * \underbrace{\frac{\text{Expenditure}_P/\text{Enrollment}_P}{\text{GNI}/\text{Total Population}}}_{Quality_P} \\
EIR_M &= \frac{\text{Expenditure}_M/\text{Population}_{11/12-18}}{\text{GNI}/\text{Total Population}} = \underbrace{\frac{\text{Enrollment}_M}{\text{Population}_{11/12-18}}}_{GER_M} * \underbrace{\frac{\text{Expenditure}_M/\text{Enrollment}_M}{\text{GNI}/\text{Total Population}}}_{Quality_M} \\
EIR_H &= \frac{\text{Expenditure}_H/\text{Population}_{18-22}}{\text{GNI}/\text{Total Population}} = \underbrace{\frac{\text{Enrollment}_H}{\text{Population}_{18-22}}}_{GER_H} * \underbrace{\frac{\text{Expenditure}_H/\text{Enrollment}_H}{\text{GNI}/\text{Total Population}}}_{Quality_H}
\end{aligned} \tag{1}$$

The first component, as it turns out is nothing but GER, capturing the *quantitative* part of education expansion. It is simply total enrollment over the population size in a given cohort. The second term- *Quality*, captures how much a country spends per enrolled student relative to its per capita economic development. It as an input-based *quality* measure. The benefit with *Quality* is that it provides a statistic that is comparable across time and space without the need of exchange rate and price index (often difficult to found in long-run).

Quality could be further decomposed into *Quality*1 (Teacher per student) and *Quality*2 (expenditure per teacher as share of GNI per capita, proxy for teachers' relative salary). The intuition behind this is that for a given level of expenditure per student, there are two ways a country could strategize to spend. It can either hire more teachers (at lower cost, maintaining better pupil-teacher ratio) or hire less teachers (at higher cost in pursuit of attracting better talent towards the education sector). These two are also well-known input-based quality measures, often targeted by policymakers and is reverberated in the policy documents of China and India.[96]

$$\begin{aligned} Quality_j &= Quality1_j * Quality2_j \\ &= \frac{\text{Teachers}_j}{\text{Enrollment}_j} * \frac{\text{Expenditure}_j/\text{Teachers}_j}{\text{GNI/Total Population}} \\ &= \underbrace{\frac{1}{PTR_j}}_{Quality1_j} * \underbrace{\frac{\text{Expenditure}_j/\text{Teacher}_j}{\text{GNI/Total Population}}}_{Quality2_j} \end{aligned} \tag{2}$$

where $j \in P, M, H$ for Primary, Middle and Higher education respectively.

*Quality*1 is inverse of PTR. A lower class-size has been shown to have positive impact on learning outcomes.[97] Additionally, the positive impact of small class-size tends to be higher among minority and lower socio-economic backgrounds students. [98]. *Quality*2 is a proxy for teachers' relative salary. It is a proxy as part of the total expenditure goes into developing and maintaining infrastructure, creating better working conditions for teachers. In some sense, it is a broader measure than teachers' salaries. It also signals the attractiveness of the education sector relative to the overall economy. A higher value

[96] A meta-analysis by Glass and Smith, 1979 used 77 studies dating back to as old as 1900's in support of a lower pupil-teacher ratio.

[97] After the Project STAR of the 1980s in the USA, causal evidence started pouring in regarding the impact of class size on students' achievement. In post-1990, several papers have found a positive causal impact of reducing class size on achievement scores. A. B. Krueger, 1999 using STAR data found the effect to be .20 s.d. for kindergarten, .28 sd in class I, .22 sd in class 2 and .19 sd in class 3. Case and Deaton, 1999 finds strong and significant effects of pupil-teacher ratios on enrollment, on educational achievement and test scores for numeracy in South Africa.

[98] A. B. Krueger, 1999 finds larger impact for black students; Angrist and Lavy, 1999 finds that reducing class size induces a statistically significant and substantial increase in test scores for 4th and 5th graders

implies a better qualitative measure for both components. To simplify, essentially, we decompose the EIR computed at primary, middle and higher levels separately into three multiplicative parts as below.

$$\begin{aligned} EIR_j =& Quantity_j * Quality1_j * Quality2_j \\ =& \underbrace{GER_j}_{Quantity_j} * \underbrace{\underbrace{(1/PTR_j)}_{Quality1_j} * \underbrace{\frac{\text{Expenditure}_j/\text{Teacher}_j}{\text{GNI/Total Population}}}_{Quality2_j}}_{Quality_j} \end{aligned} \tag{3}$$

where $j \in P, M, H$ for Primary, Middle and Higher education respectively.

1.4.2. "Prioritizing Quantity" vs "Prioritizing Quality". We make a case that China's strategy has been to prioritize quantity (even at the cost of quality) and after achieving a certain expansion level, it starts improving quality. The development strategy in India, has been to maintain balance between quantity and quality, and to some extent even prioritizing quality. Second, China produces more teachers at low cost (thereby keeping PTR low or better $Quality1$) whereas India produces less teachers but at a higher cost per teacher (better $Quality2$). Table 1.3 shows the average values of decomposition components at different stages by periods. The lower part of the Appendix Figures- 1.13; 1.14 and 1.15 plots the pupil-teacher ratio from 1912-2018.

Primary Stage: In recent years, both countries have had similar EIR_P, and even the quantity (GER_P) and quality ($Quality_P$) components are very similar. The journey to reach this similarity has been very different. Following its "quantity first quality later" approach, China first achieved more than 100% GER. In contrast, with its balancing "quantity with quality" approach, India crossed 100% GER for the first time almost 40 years later than China.

During pre-1950, there was a rapid increase of GER in China (closing the gap with India) with a declining quality level. However, it continued to have better quality ($Quality_P$) level with a narrowing gap, which occurred due to quality decline in China and quality increase in India. The $Quality_P$ drops from 17.5% in 1932 to 10.5% in 1936 in China, whereas it increases from 6% to 8.5% in India between 1902-1937.[99] The drop in $Quality_P$ occurred in two ways- first in the 1920s by deteriorating $Quality1_P$ (PTR doubled in the 1920s (from 13 to 26) and remained close to 30 until 1950) and then in the 1930s from the declining $Quality2_P$ (400% in 1932 to 286% 1936). In India, the $Quality_P$ increase happened entirely from the $Quality2_P$ increase (134% during 1900-30 to 229% in the 1930s) as $Quality1_P$ declined (PTR increased slightly from 27 to 30 between 1912-50).

The trend of increasing quantity (with decreasing quality) remained the dominant feature in China till it reached 100% in the 1970s. On the other hand, India consistently improved in quality and quantity from 1950-85, resulting in higher EIR_P than China in the 1970s. [100] The declining quality in China occurred- first due to $Quality1_P$ (PTR was 33) up to 1960 (after which it stabilizes or improves slightly) and later from the declining $Quality2_P$ (reduces to 145% compared to 344% in the 1950s). The decreasing PTR in China up to the 1980s came from the massive expansion of teachers (to meet the enrollment demand). In India, the $Quality_P$ increase came entirely from the $Quality2_P$ increase, as there was a consistent deterioration in the $Quality1_P$ (due to increasing PTR).[101] The improvement in $Quality2_P$ shows increasing salaries of teachers and their working conditions. This highlights the stark difference in the adopted strategies by China and India. China recruits teachers at a low salary, whereas India hires teachers at a high pay scale (in expectation of attracting better-quality personnel).

Post 1985, EIR_P is at similar level in both countries, reaching 15% in 2006 and close to 20% in the 2010s. But this similarity masks the difference. The increasing EIR_P after 1985 in China comes entirely from increasing both the components of $Quality_P$ as GER stabilized at 110% after the 1980s. PTR improved due to decreasing enrollment (an artefact of demographic factor - reduction of fertility level) and $Quality2_P$ improved reaching 300% compared to 145% during the 1960s. In India, during 1985 -2000, the increase in EIR_P was due to $Quality2_P$ increase (with an almost stagnant GER at 80%) and after 2000 majorly from quantity increase (GER crossing 100% mark). $Quality_P$ is close to 20%

[99] The expenditure details for China during 1939-49 has 4 data points 1931, 1932, 1935 and 1936.

[100] In China, EIR_P went below 6% in the 1970s, whereas, in India, EIR_P continued increasing, overtaking China in the 1970s and remained higher till 2006.

[101] In India, PTR was 35 in the 1950s, 38 in the 1960s and 70s, 41 in the 1980s, and 45 in the 1990s. In China, PTR was close to 28 in the 1970s, 23 in the 1980s and 90s, and 19 after 2000.

in both countries in recent years though India spends more than double exp/teachers (relative GNIpc), whereas it has double PTR (38 in India compared to 19 in China: 2015).

Middle Stage: Similar to the primary stage, the trend of increasing quantity (with decreasing quality) remained till the 1970s in China (GER touched 50% in 1980). China started with a very high $Quality_M$, so even with the quality decline, it remained higher than India till the 1960s. In pre-1950, the expenditure per student was almost twice the size of GNIpc in China, whereas it was on average 0.7-0.8 times (of GNIpc) in India. [102] The $Quality_M$ declined to 25% in 1965 and 12% in 1976, which occurred mainly due to a reduction in the $Quality2_M$ (one-tenth during 1965-85 compared to the 1930s level) and not so much from the $Quality1_M$ (though PTR increased from 14 during 1900-30 to 20 during 1965-85). It was in the 1970s that China surpassed India in GER, but $Quality_M$ fell below India. It is to be highlighted that till 1985, there was also a quality decline in India (but slower than China) due to a reduction in both $Quality1_M$ (PTR 17 during 1900-30 to 26 during 1965-85) and $Quality2_M$ (488% in 1965-85 from 1498% during 1930-49).

Post-liberalization, there is a reversal in the declining $Quality_M$ in both countries, and it has stabilized since 2010. There is improvement in both $Quality1_M$ (PTR reducing to 16) and $Quality2_M$(increasing to 400% from 289% before) in China. In India, it is only the improvement in $Quality2_M$ which led to raising $Quality_M$. $Quality_M$ is close to 25-30% in both countries in recent years though India spends more than double on exp/teachers relative to GNIpc (700% in India and 400% in China), whereas it has double PTR (25 in India compared to 14 in China: 2015). The quantitative advantage in China is visible with close to 90% GER (70% in India).

Higher Stage: The higher stage expansion is still ongoing in both countries with a declining quality. In pre-1950, since the higher education was very limited (GER<0.1%) the cost was very high in both countries. In 1930s exp/student was 7 times GNIpc in India and 16 times GNIpc in China. The decomposition shows that both quality components were better in China- a better PTR and a better (exp/teacher w.r.t GNIpc) during this period. A part of it is due to higher diversification in China than India, with China

[102] Another way to look at it was when GER was 2% in both countries (1900-30 in IN and 1930-50 in CH), then the quality measure was three times in China than in India.

developing a more expensive form of streams. $Quality_H$ was on declining trend in both countries.

Post-1950, in the higher stage, the dominating factor has been the quantitative expansion (GER) along with declining $Quality_H$ in both countries. First, during 1950-65, there was a rapid quantitative expansion in India and later post-1980s dramatic expansion started in China. Throughout the period $Quality_H$ is on decline in both countries. In 1985, India had double GER than China, though the $Quality_H$ was double in China.

Interestingly, PTR was just ten up to the 1990s in China, which allowed the possibility of a rapid expansion of enrollment in higher education in the 2000s. It decreased $Quality1_H$ in China (PTR doubled from 15 during 1986-00 to 30 during 2001-15). Post-2000, $Quality2_H$ became higher in India. The quantity rapidly increased in China and surpassed India in the early 2000s. The GER was 15% in 2000 (11.3% in India) reaching 60% by 2015 (34% in India) in China. As the higher education system is expanding in both countries, there is a declining trend in quality, though still, it is higher in India than in China.

In summary, China's strategy of prioritizing quantity (GER) was very strong during communism period and reached its peak during the cultural revolution. It helped China bringing more school-going age kids to schools much earlier than India, though the quality was possibly not great. On the other hand, India's policy of expanding education while prioritizing quality which started during the colonial period (from early 1900s) continued as late as the 2000s. It is to be noted that during pre-1950, even though the debate centred around maintaining quality, the quality was almost half of China (at all stages) - due to scarce resource allocation towards education, keeping the teachers' salary (especially at the primary level) very low. Today, India has higher quality measure at all stages.

1.4.3. Teachers' Wages Rank Percentile. : This section highlighted China's approach of of quantity first and quality later. A prudent opinion would be that during the quantity expansion phase (increasing number of teachers and students) in China, the tool for maintaining quality was the pupil-teacher ratio. On the other hand, the Indian approach has been quantity expansion while maintaining quality, where tools for maintaining quality were more expensive - employing good quality teachers (emphasis on minimum qualifications of teachers), their training and higher remuneration. [103]

The surveys after the 1980s allow to compute the percentile of teachers' wages among all the professions in both countries (Refer Figure 1.8). In the 1990s, the teachers' (for all three levels) wages were at a higher rank than in China. The teachers' salaries in 1995 were at 65, 48 and 33 percentile in China for higher, middle and primary levels. The corresponding percentile rank of Indian teachers' wages were 91, 77 and 60. In the The gap between China and India was possibly similar in the 1980s (China data is not available to make comparison, but the rank percentile in India was similar). The policy change to focus on quality post-1990s in China led to a vast increase in the wage rank of teachers at all levels, and it became similar to India at the middle and higher stages in the early 2000s. The wage rank of teachers became higher than in India at the primary level (CH:69 and IN:60). By 2010, there was a complete reversal. The wage rank of teachers at all levels became better in China than in India. The gap is significant at the primary level (CH: 86 and IN:62) in 2011 and 2018.

In recent years, with causal evidence highlighting that the traditional input-based measures are not being reflected in the learning outcomes (especially in the developing countries) has increased the skepticism towards input-based measures.[104]. Though it does not mean that spending and resources never matter (Hanushek and Woessmann, 2012) and in developing countries these are the first step towards the quality improvement. It is still an open question how much of better PTR in China or better-wages of teachers in India actually reflected in the learning outcomes?[105] Two useful remarks. Several studies in India have highlighted the problem of low attendance rate of students and teachers in school. Since the enrollment expansion happened earlier in China than

---

[103]The other tool has been setting up "model" institutions since the 1900s to serve as examples for private bodies. Similarly, China also started establishing "key" institutions, though only after 1980. These institutions were more expensive to set up, eating up the scarce resources.

[104]The World Conference on Education for All-1990, stressed that the quality of education should be learning outcomes and input-driven measures are simply the *means* The adoption and implementation of MDG post-1990s by several developing countries resulted into accelerated enrollments, which further pushed towards outcome-based quality measures.Since then there has been more and more focus towards outcomes-based (learning) measures (Dundar et al., 2014)

[105]We neither argue that input-based quality measures are better nor that input-driven measures are going to be reflected in the learning outcomes of the students.

India at all stages, the input-based measures are going to be the upper bound on the outcome-based quality measures.

Finally, the quantity-quality discussion is also related to the strong emphasis on developing a research-oriented higher education in India just after independence, leaving fewer resources for primary and middle education (slowing down the expansion), which had to be devoted to maintaining the quality of the school stage as they feed into the higher education.

1.5. Education-Wage Inequality

This section shows how education development has impacted the distribution of wages in these two countries. Since the education level and education distribution both impact wage distribution, first we present the evolution of education level and education inequality. Next, we study the dynamics of education and wage inequality.

1.5.1. Education Inequality. The distribution of education is essential both for welfare and production consideration. Various empirical papers have studied the relationship between increased education and education dispersion. Ram, 1990 categorically points out that it is an empirical question- increasing education (average years of schooling) may not always decrease education inequality if the increase is concentrated in the middle/tertiary level of education.

In pre-1950, the growth of literacy rate was very slow in India. From 1901-1951, the literacy rate grew by just 13pp (5% in 1901 to 18% in 1951). The first data point of literacy rate in China in 1950 shows that both countries had very similar literacy rates (China: 20%; IN:18%). The difference in literacy rate (China-India) increased from 2pp in 1950 to its peak at 27pp in 2000. The literacy rate was 91% in China and 64% in India in 2000. The gap started narrowing, and in 2015, the literacy gap remained at 20pp. China is close to 96% and India at 76%. (See bottom part of Figure 1.7)

Next, we compute cohort-wise average years of education(AYS), absolute education inequality (Standard Deviation in Schooling; SDS) and relative education inequality (Gini). [106] The methodological details of the computation are provided in the Appendix C.5.1. Figure 1.9 presents the evolution of AYS, Gini and SDS.

[106]The interpretation is that the cohort born in, say, the year 1950 had expected average years of schooling of x years.

AYS has consistently increased with a simultaneous decline in the education Gini in the past. Both countries have approximately similar AYS for the 1950-born cohort. The bottom-up expansion of education in China results in a faster gain in the AYS and a faster decline in education inequality for the later cohort. The 1950-cohort had an expected AYS of 2.1 years in both the countries, but for the 1966-born cohort, the expected AYS is 8.6 years in China compared to only 3.6 years for India! The decline in the absolute measure of dispersion, i.e. SDS, starts in the 1960 born cohort in China and remains lower than in India for subsequent years. The education expansion in India for a long time continues with increasing SDS.[107] The relative measure of education dispersion, i.e. Gini index, has remained higher in India with a clear diverging point from 1950. The cohorts born in China and India after 1960 have faced very different educational opportunities and education dispersion. The decline in post-1990 in India reflects the effort after 2000 to bring all young kids to primary schools.

1.5.1.1. *Gender Education Inequality.* The census in both countries provide basic literacy rate by gender. The female literacy rate gap (w.r.t total literacy rate) was close to -14pp in 1980s in both countries. (China 1982: Tot-66% and Female-51% (India 1981: Tot-44% and Female-30%). The gap reduced to 2pp in China (2015) and 8pp in India (2011).

We compute female share in enrollment (=female enrollment /total enrollment) at all three stages of education from our dataset. The gender gap is the difference between the female population share and female enrollment share, capturing over/under-representation of females in education relative to the female population. Both countries started with a very high gender gap and have made a consistent effort to narrow the gap over the last 100-120 years. In the primary stage, the divergence between both countries started after the 1950s due to China's rapid expansion of primary education. The divergence at the middle and higher stages between China and India began earlier in the 1930s due to higher female drop-outs in India at the primary level.

[107] Appendix Figure 1.19 presents the existence of the Educational Kuznets curve in both countries. The peak of the highest SDS comes at AYS of 7.85 in China and 7 in India. However, this peak represents a very different cohort in the two countries- 1959 for China and 1987 for India.

Similarly, we compute the female share of teachers and the gender gap among teachers. The female share of teachers also started with a very low base in both countries. There are some evidence that female teachers helps in increasing female enrollment. [108] In recent years *feminization* of the teaching profession has resulted in more female teachers, especially in the primary and middle stages. At the primary stage, in 1952, female teachers were less than 20% in both the countries, implying a gender gap of 30pp. Over the years, a steady increase of female teachers has pushed the share to 50% in India and 62% in China today. As we saw in the enrollment, the divergence starts from 1950 at the primary level. In the middle stage, the share went up from almost 0% to 60% female teachers in China during the 20th century. In India, the share of female teachers in the middle stage has reached 41% (the latest data is from 2011). At a tertiary stage, the share of female teachers has also increased, from 21% (14%) in China (India) in 1965 to 49% (42%) in 2016. The share of female teachers remained relatively similar in both countries at the middle and tertiary stages during the 20th century. The difference comes after the 2000s, with China having a higher share of female teachers.

A detailed description is provided in the Appendix C.1.1.

1.5.1.2. *Caste Education Inequality.* Caste-based society of India can be traced to be one of the oldest types of ternary society[109] which during colonial policies got rigidified/codified through censuses (Piketty, Yang, and Zucman, 2019). To overcome the historic injustice, independent government of India rallied behind the policies of positive affirmation targeting lower/backward castes - by reserving seats in public schools/colleges (or in govt. jobs) according to their population proportion. Hence, caste is one of the important stratifying agent in Indian society even today. [110]

Figure 1.23 presents the evolution of the enrollment share of Scheduled Castes (SC) and Scheduled Tribes(ST) at Primary, Upper Primary (Class VI-VIII) and Secondary (Class IX-XII) level. Several important points worth emphasizing. First, the level of

[108] Andrabi, Das, and Khwaja, 2013 finds that construction of public girl's secondary schools resulted into more private primary schools in later years, by augmenting the local female teacher supply.

[109] A ternary society as defined in Piketty, Yang, and Zucman, 2019 divides society into 3 major social groups with different status, functions and rights. India has a variant of this form with- priests(Brahmins), warrior(Rajputs), Merchants(Vaishyas), Labourers(Shudras) and Outcastes (Scheduled Caste).

[110] Bharti, 2018 has shown the concentration of income or wealth into higher/forward caste section of the society.

under-representation in both the groups (i.e difference from their population share) increases with the increase in stage of education. We see, the gap closing first in primary stage, then in upper primary and then in secondary stage of education. Second, comparing both the groups we see Scheduled caste (Dalits) has slightly better performance in closing the gap. Third, even after 15 years of independence in 1965, both the groups were heavily under-represented. It is only in 2000's we see that the percentage of enrollment in elementary schooling (upto class VIII) is at par with their population share, after 60-70 years of independence!

1.5.2. Earnings / Wage Inequality. The income inequality in both countries is increasing, and India has a higher level of income inequality. The share of income accruing to the Top 10% population in China increased from 30% to 41% and India from 35% to 56% from 1985 to 2015 (Source: World Inequality Lab). Both the components of income, i.e. *wage income* and *capital income* can be impacted by the level of education. The availability of individual-level data only for the wage income since the 1980s allows us to study the changing education-wage inequality relationship [111]. We narrow the sample to the working-age population (20-60 years) employed in long-term salaried jobs and positive income.

The average real daily wage ($ 2018 level) is increasing in both the countries (See Table 1.4). It is almost three times in China (32.2$/day) than in India (11.3$/day) in 2018. Female share in India is very low. It is only 13-22% compared to $>$ 40% in China. The education composition in both countries has changed over time with rising share of higher education graduates in the salaried class. In both countries, the percentage has increased from 13-14% to 36-37% from the 1980s to 2018. It is possibly due to the increasing education level (among population) and the rising demand for skilled labour (Acemoglu, 1998). There are a corresponding decline in China for primary (14% to 8%) and Middle (74% to 56%) level graduates. Interestingly in India, the decline is seen only in the primary-level graduates (44% to 19%) whereas the share of middle-level graduates is stable at $\sim$ 45%, which suggests job-polarization. China has transitioned towards the service sector from the manufacturing sector, and now both countries have more than 70% employment in the service sector. The manufacturing sector share is stagnant at around 25% in India. The daily wage ratios (Tertiary/Middle) and (Middle/Primary)

[111] The only income survey available for India is the IHDS panel survey of 2005 and 2011.

have increased in both countries, though at a faster rate and a higher level in India (See Appendix Table 1.9).

The wage inequality measures computed from the wage surveys (Refer Table 1.5 for Theil's index and Appendix Table 1.10 for other measures) are in line with the recent evidence on evolution of inequality in both the countries[112]. We decompose the Theil's inequality index into between and within components by education groups, where groups are formed as primary, middle and tertiary. The between-component which captures the "education effect" is higher for India. The percentage share of between component has remained 25-30% (i.e. education groups explain almost one-third of the wage inequality) in India for all the survey years. In China, in the 1980s and 90s the between component was just 1%. After 2000 the between component increased and has remained around 15-20% in China. It suggests that the link between education and inequality is higher in India, though the difference is narrowing down over the years due to the increasing education effect in China. The large between-group difference between the two countries overshadows the difference in within-group[113]. The results are similar using Mean Log Deviation.

Education and earnings inequality are interconnected in a very complex dynamic way. We focus here on the central elements. The first element is that both - level of education and education dispersion affect earnings inequality (Gregorio and J.-W. Lee, 2002). The theoretical model[114] predicts an unambiguous positive association between education inequality (as measured by SDS) and earnings inequality and an ambiguous effect of increase in average schooling on earnings inequality (due to covariance with the rate of return to education). Since SDS is higher for India after the 1950 cohort, the impact on earnings inequality will also be higher if everything else remains the same. The effect of the increasing education level is ambiguous because it also depends on its

[112] China: Wealth and Income Series: CH(Piketty, Yang, and Zucman, 2019); India: Wealth (Bharti, 2018); Income(Chancel and Piketty, 2017)

[113] The between-group inequality is two-times in India than China in 2018 whereas the within-group inequality is almost same

[114] According to the Human capital theory model

$$Var(lnwage_s) = \bar{r^2} Var(S) + \bar{S^2} Var(r) + 2\bar{r}\bar{S} Cov(r,S) + Var(u)$$

where S is the years of schooling, r rate of return to education and the bar represents average. u is the random component covariance with the rate of return (RoR) to education.

We estimate RoR[115] using extended Mincer's equation (with tertiary and primary graduates dummy and base as middle graduates). We run the standard regression:

$$ln(dailywage)_i = \beta_0 + \beta_1 College_i + \beta_2 Primary_i + \beta_3 age_i + \beta_4 age_i^2 + \mu X_i + Prov_i + \epsilon_i \tag{4}$$

,

where $College_i$ and $Primary_i$ are dummies for college and primary level graduates, respectively. Other controls (X_i include- gender and region (urban/rural) and provinces/state Fixed effects. The dependent variable is the log of daily wage (in real 2018 $), capturing the productivity.[116] The main coefficient (100*β) is plotted in the Figure 1.10 and the full result is presented in the Table 1.6. The upper part of the graph plots the coefficient for College (i.e. tertiary graduates). The lower part plots the coefficient for primary graduates compared to middle-level graduates. Several interesting observations emerge. The wage effect for a college education is always higher in India than in China. Within India, the wage effect remained almost constant in the 1980s and started increasing post-liberalization, with the highest increase observed between 2000-2011. In 2011, the average impact of education on HE graduates' wages reached 76% compared to 44% in 1983 and 51% in 1999). In 2018, it declined to 65%. Within China, the wage effect also started increasing post-liberalization. The average wage effect was 9% in 1988, which jumped to 27% in 1995, 46% in 2002, 36% in 2013 and 50% in 2018. As expected, the coefficient on primary education w.r.t middle level is negative in both countries but more negative for India than China.

[115]We are using the term Rate of return and Wage effect interchangeably, though by definition they are slightly different. Strictly speaking, the raw coefficient of the Mincer's equation is wage effect. Whereas return to education takes into account the years of education. Since the education structure remained same in both the countries in the analysis period the evolution of return to education will remain qualitatively similar.

[116]The daily wage is computed for China using the information on the total wages earned in a year divided by the total working days. Indian labour force surveys (except 2018) have collected information on working days (full-day, half-day or no work) and wages earned with the last seven days reference. The daily wage is simply weekly wages divided by total working days. In 2018, it captured monthly wages and working hours for the past seven days. If the number of hours was less than 4 hours, we assume it to be half-day work. We compute weekly working days and multiply by 4 to get monthly working days. The daily wage is the monthly wage divided by estimated monthly working days.

The increasing RoR combined with the rising average education implies positive covariance of the term $2\bar{r}\bar{S}Cov(r,S)$ for India. It then suggests that expansion of education had an overall positive impact on inequality according to the human capital theory model (all the terms are positive). The RoR for the decade (2002-2013) in China is negative, whereas average education increased, making the covariance term negative. The higher wage inequality in India(than in China) is due to higher education inequality and the positive relationship between the expansion of education and RoR.

1.5.3. Impact of Educational Expansion on Inequality. Next, to pin down the impact of education on wage inequality, we estimate the unconditional partial effect (UPE) on different distributional statistics following Firpo, Fortin, and Lemieux, 2009. [117] The estimation process has two steps. In the first step Recentred Influence Function(RIF)[118] is estimated depending on the distribution function under consideration. The estimated RIF is used as the dependent variable in OLS regression in the next step.

$$\text{RIF\_i} = \beta_0 + \beta_1 College_i + \beta_2 Primary_i + \beta_3 age_i + \beta_4 age_i^2 + \mu X_i + \rho Prov_i + Indus_i + Occup_i + \epsilon_i \tag{5}$$

For the first step, here we use annual real wages (in $ 2018) instead of daily wage, as it is more suitable for inequality analysis. Our main interest is coefficients β_1 (and β_2), which capture the effect of increasing the proportion of tertiary (primary) graduates on the expected change in the unconditional distributional statistic. We add industry and occupation fixed effects, too, following Firpo, Fortin, and Lemieux, 2009. The coefficients with the variance of log wage statistic are provided in Table 1.7 (See full tables in the Appendix Tables 1.11 - 1.16).

The coefficients for India are positive, significant and have remained stable, depicting a consistent positive impact of education on wage inequality. The positive sign for both β_1 and β_2 shows that increasing the population share of tertiary and primary graduates is positively associated with rising inequality. Over the years, the effect of β_1 (relative to

[117] Firpo, Fortin, and Lemieux, 2009 argues UQR to be more relevant for policy perspective compared to the conventional Conditional quantile regression (CQR). CQR computes RoR at different quantiles of wages where quantiles are conditional on the covariates. However, it does not capture the impact on unconditional quantile.

[118] Influence functions are statistical tools to compute the influence of an individual observation on the distributional statistic. RIF(y,v)=IF(y,v)+v, where v is the distributional statistic of y

the mean of the RIF) has declined from 52% in 1983 to 25% in 2018. The coefficients are also similar for China, suggesting increasing share of tertiary (and primary) graduates is increasing inequality. The main difference with India is that the impact has grown over the years. In the 1980s and 90s, the effect was much smaller. In 2018, one unit increase in the share of primary graduates was associated with 50% increase in inequality statistic (which was 14% in 1988). Similarly, one unit increase in the share of tertiary graduates was associated with 25% increase in inequality statistic (which was 14% in 1988) in 2018.

Using quantiles as distributional statistics provides further insights into which portion of the wage distribution drives wage inequality. Intuitively, it means estimating the wage effect at different quantiles of wages. Figure 1.11 presents the coefficients on tertiary graduates (β_1) from UQR for China and India at ten data points (for better clarity). The positive coefficients at all deciles imply a positive effect of higher education at all earnings levels. In the 1980s and 90s, the curve was almost flat for China (and close to the average wage effect computed from Mincer's OLS coefficient), resembling a "controlled" wage structure of the communism period. It suggests there is no discernible differential impact of tertiary education along the wage distribution. In India during this period, the curve is monotonically increasing with quintiles. [119]

Post-2000, the UQR curve changes drastically in both countries. In China, the curve became similar to the UQR curve of India's 1980s/90s (monotonically increasing with deciles). In India, the curve becomes inverted U shaped for India. The coefficient rises to a peak at the 60th-70th percentile and then declines for higher quantiles (but remains higher than the lower quantiles). It suggests that tertiary education decreases the wage dispersion between the top and the middle of the wage distribution but increases between the middle and bottom wage distribution. Another interesting observation is that the coefficients at lower wage quantiles in China and India are similar; the difference is at the higher wage quantiles.

1.5.4. Demand-Supply Mismatch? The increasing wage effect of HE is perplexing for India due to increasing graduates/ enrollment in HE. The growing supply of HE graduates should lead to a decline in the wage effect. It is somewhat evident in China

[119] The coefficient is below the average effect, computed through Mincer's OLS, up to - 60 percentile in 1983; 50 percentile in 1987 and 1993; 40 percentile in 1999, 2004, 2009 and 2011; and 50 percentile in 2018.)

(the dip in RoR from 2002 to 2013) due to the large supply of HE graduates (trumping the demand-side factors). The increasing wage effect implies that the demand for high skilled (educated) workers is not met in the growing supply of HE graduates. A peek into the wage ratio (tertiary/middle) by cohort strengthens the case. The wage effect for the younger cohort (Age 26-30) increases faster than the older cohort (Age 45-60) in India. In China, the high supply of tertiary graduates is resulting in a decline in wage effect for the younger cohort compared to the older cohort (Refer Appendix Figure 1.27).

It could be due to the lack of synchronization between the market demand and the college graduates' supply. In the section on discipline-wise graduates, we notice that the share of graduates from different disciplines is more dynamic in China than in India. Further, a very high percentage of graduates come from Humanities (Arts and Commerce). These issues may be the reason behind the issue of unemployability of the graduates in India if one believes in the skill-enhancing effect of colleges. If educational degrees merely serve the purpose of signalling, then it would mean the hierarchies of colleges play a more critical role. It could be the effect of both, and a more in-depth analysis is required to pin down the reasons.

1.6. Conclusion

The progress of the modern education system in China and India has followed different paths. The challenges and opportunities created by the different politico-socio-economic environments in these two countries have led to adoption of various education policies. It, in turn, has shaped the evolution of the education system. The path of education development in China aligned better with the economic development, possibly leading to a higher growth rate after the 1980s with a lower level of inequality. The path of educational development in India was a riskier choice in the beginning as the country was far from the technological frontier. The case study of China and India provides insights for other developing countries in building their education system with a rider that the 21st century will not be the same as the 20th century.

Bibliography

Acemoglu, Daron (1998). “Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality”. In: *The Quarterly Journal of Economics* 113.4, pp. 1055–1089.

Aghion, L. Boustan, C. Hoxby, and J. Vandenbussche (2009). “The Causal Impact of Education on Economic Growth: Evidence from U.S.” In: *U.S. Harvard University, Mimeo*.

Ahluwalia, Montek S. (1976). “Inequality, poverty and development”. In: *Journal of Development Economics* 3.4, pp. 307–342.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (June 1, 2003). “Fractionalization”. In: *Journal of Economic Growth* 8.2, pp. 155–194.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja (Apr. 1, 2013). “Students today, teachers tomorrow: Identifying constraints on the provision of education”. In: *Journal of Public Economics* 100, pp. 1–14.

Angrist, Joshua D. and Victor Lavy (1999). “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement”. In: *The Quarterly Journal of Economics* 114.2. Publisher: Oxford University Press, pp. 533–575.

Arnove, Robert F. (1984). “A Comparison of the Chinese and Indian Education Systems”. In: *Comparative Education Review* 28.3. Publisher: [University of Chicago Press, Comparative and International Education Society], pp. 378–401.

Azam, Mehtabul and Geeta Gandhi Kingdon (2015). “Assessing teacher quality in India”. In: *Journal of Development Economics* 117, pp. 74–83.

Baier, Scott L., Gerald P. Dwyer, and Robert Tamura (2006). “How Important Are Capital and Total Factor Productivity for Economic Growth?” In: *Economic Inquiry* 44.1, pp. 23–49.

Barro, Robert J. and Jong Wha Lee (Sept. 1, 2013). "A new data set of educational attainment in the world, 1950–2010". In: *Journal of Development Economics* 104, pp. 184–198.

Bharti, Nitin (2018). *Wealth Inequality, Class and Caste in India, 1961-2012*. hal-02878149. Publication Title: World Inequality Lab Working Papers. HAL.

Bolt, Jutta and Jan Luiten van Zanden (2020). "Maddison style estimates of the evolution of the world economy. A new 2020 update". In: p. 44.

Cantoni, Davide and Noam Yuchtman (Sept. 1, 2013). "The political economy of educational content and development: Lessons from history". In: *Journal of Development Economics* 104, pp. 233–244.

Case, Anne and Angus Deaton (1999). "School Inputs and Educational Outcomes in South Africa". In: *The Quarterly Journal of Economics* 114.3. Publisher: Oxford University Press, pp. 1047–1084.

Caselli, Francesco and Antonio Ciccone (Mar. 2019). "The Human Capital Stock: A Generalized Approach: Comment". In: *American Economic Review* 109.3, pp. 1155–1174.

Castello-Climent, Amparo and Abhiroop Mukhopadhyay (2013). "Mass education or a minority well educated elite in the process of growth: The case of India". In: *Journal of Development Economics* 105 (C). Publisher: Elsevier, pp. 303–320.

Castelló-Climent, Amparo, Latika Chaudhary, and Abhiroop Mukhopadhyay (2018). "Higher Education and Prosperity: From Catholic Missionaries to Luminosity in India". In: *The Economic Journal* 128.616, pp. 3039–3075.

Castelló-Climent, Amparo and Rafael Doménech (Jan. 14, 2021). "Human capital and income inequality revisited". In: *Education Economics* 0.0, pp. 1–19.

Chancel, Lucas and Thomas Piketty (July 1, 2017). "Indian income inequality, 1922-2015: From British Raj to Billionaire Raj?" In: *wid.world* 2017/11, p. 71.

Chaudhary, Latika (Mar. 2009). "Determinants of Primary Schooling in British India". In: *The Journal of Economic History* 69.1. Publisher: Cambridge University Press, pp. 269–302.

Chaudhary, Latika, Aldo Musacchio, Steven Nafziger, and Se Yan (2012). "Big BRICs, Weak Foundations: The Beginning of Public Elementary Education in Brazil, Russia, India, and China". In: p. 61.

Cogneau, Denis and Alexander Moradi (Sept. 2014). "Borders That Divide: Education and Religion in Ghana and Togo Since Colonial Times". In: *The Journal of Economic History* 74.3. Publisher: Cambridge University Press, pp. 694–729.

Cohen, Daniel and Marcelo Soto (Mar. 1, 2007). "Growth and human capital: good data, good results". In: *Journal of Economic Growth* 12.1, pp. 51–76.

Dundar, Halil, Tara Béteille, Michelle Riboud, and Anil Deolalikar (May 1, 2014). *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities*.

Fearon, James D (2003). "Ethnic and Cultural Diversity by Country*". In: p. 28.

Firpo, Sergio, Nicole M. Fortin, and Lemieux (2009). "Unconditional Quantile Regressions". In: *Econometrica* 77.3. Publisher: [Wiley, The Econometric Society], pp. 953–973.

Fuente, Angel de la and Rafael Donénech (Oct. 12, 2000). "Human Capital in Growth Regressions: How much Difference Does Data Quality Make?" In: Publisher: OECD.

Gao, Pei (Nov. 2015). *Risen from Chaos: What drove the spread of Mass Education in the early 20th century China*. Working Paper 89. EHES.

— (2018). "Risen from Chaos: The Development of Modern Education in China, 1905-1948". In: *Australian Economic History Review* 58.2. Publisher: Economic History Society of Australia and New Zealand, pp. 187–192.

Gemmell, Norman (1996). "Evaluating the Impacts of Human Capital Stocks and Accumulation on Economic Growth: Some New Evidence†". In: *Oxford Bulletin of Economics and Statistics* 58.1, pp. 9–28.

Glass, Gene V. and Mary Lee Smith (1979). "Meta-Analysis of Research on Class Size and Achievement". In: *Educational Evaluation and Policy Analysis* 1.1. Publisher: [American Educational Research Association, Sage Publications, Inc.], pp. 2–16.

Goldin, Claudia (2001). "The Human-Capital Century and American Leadership: Virtues of the Past". In: *The Journal of Economic History* 61.2. Publisher: Cambridge University Press, pp. 263–292.

Gregorio, José De and Jong-Wha Lee (2002). "Education and Income Inequality: New Evidence From Cross-Country Data". In: *Review of Income and Wealth* 48.3, pp. 395–416.

Gyimah-Brempong, Kwabena, Oliver Paddison, and Workie Mitiku (Feb. 1, 2006). "Higher education and economic growth in Africa". In: *The Journal of Development Studies* 42, pp. 509–529.

Hanson, Mark (2008). *Economic Development, Education and Transnational Corporations*. Routledge & CRC Press. (Visited on 05/22/2022).

Hanushek, Eric and Ludger Woessmann (2012). "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation". In: *Journal of Economic Growth* 17.4, pp. 267–321.

Jones, Benjamin F. (Nov. 2014). "The Human Capital Stock: A Generalized Approach". In: *American Economic Review* 104.11, pp. 3752–3777.

Joshua, John (2015). "The Accumulation of Human Capital as a Factor of Production". In: *The Contribution of Human Capital towards Economic Growth in China*. Ed. by John Joshua. London: Palgrave Macmillan UK, pp. 26–50.

Krueger, Alan B. (1999). "Experimental Estimates of Education Production Functions". In: *The Quarterly Journal of Economics* 114.2. Publisher: Oxford University Press, pp. 497–532.

Krueger, Dirk and Krishna B. Kumar (June 1, 2004). "Skill-Specific rather than General Education: A Reason for US–Europe Growth Differences?" In: *Journal of Economic Growth* 9.2, pp. 167–207.

Landes, David S. (June 2006). "Why Europe and the West? Why Not China?" In: *Journal of Economic Perspectives* 20.2, pp. 3–22.

Lau, Lawrence J., Dean Jamison, and Frederic F. Louat (Mar. 31, 1991). *Education and productivity in developing countries: an aggregate production function approach*. Policy Research Working Paper Series 612. The World Bank.

Lee, Jong-Wha and Hanol Lee (Sept. 1, 2016). "Human capital in the long run". In: *Journal of Development Economics* 122, pp. 147–169.

Leeuwen, Bas (Jan. 1, 2007). "Human Capital and Economic Growth in India, Indonesia, and Japan: A Quantitative Analysis, 1890-2000". In.

Lin, Justin Yifu (1995). "The Needham Puzzle: Why the Industrial Revolution Did Not Originate in China". In: *Economic Development and Cultural Change* 43.2, pp. 269–292.

Maloney, William F. and Felipe Valencia Caicedo (2017). *Engineering Growth: Innovative Capacity and Development in the Americas*. 6339. Publication Title: CESifo Working Paper Series. CESifo.

McMahon, Walter W. (1998). "Education and Growth in East Asia". In: *Economics of Education Review* 17.2. Publisher: Elsevier, pp. 159–172.

Mehrotra, Santosh, ed. (2014). *India's Skills Challenge: Reforming Vocational Education and Training to Harness the Demographic Dividend*. Delhi: Oxford University Press. 340 pp.

Mehrotra, Santosh, Ankita Gandhi, and A Kamladevi (2015). "China's Skill Development System: Lessons for India". In: *Economic and Political Weekly* 50.28. Publisher: Economic and Political Weekly, pp. 57–65.

Mitchell, B. R. (Jan. 1, 1998). *International Historical Statistics: 1750-1993*. Macmillan Reference Ltd.

Mokyr, Joel (2005). "The Intellectual Origins of Modern Economic Growth". In: *The Journal of Economic History* 65.2. Publisher: Cambridge University Press, pp. 285–351.

Morrisson, Christian and Fabrice Murtin (Sept. 2009). *The Century of Education*. 0109. Publication Title: CEE Discussion Papers. Centre for the Economics of Education, LSE.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1991). "The Allocation of Talent: Implications for Growth". In: *The Quarterly Journal of Economics* 106.2. Publisher: Oxford University Press, pp. 503–530.

Naik, J. P. (Jan. 1, 2000). *A Students' History Of Education In India: 1800-1973, 6/E*. Delhi: Macmillan Publishers India.

Nehru, Vikram, Eric Swanson, and Ashutosh Dubey (Apr. 1, 1995). "A new database on human capital stock in developing and industrial countries: Sources, methodology, and results". In: *Journal of Development Economics* 46.2, pp. 379–401.

Papageorgiou, Chris (2003). "Distinguishing Between the Effects of Primary and Post-primary Education on Economic Growth". In: *Review of Development Economics* 7.4, pp. 622–635.

Pereira, João and Miguel St. Aubyn (Feb. 1, 2009). "What level of education matters most for growth?: Evidence from Portugal". In: *Economics of Education Review* 28.1, pp. 67–73.

Petrakis, Panagiotis and D. Stamatakis (2002). "Growth and educational levels: a comparative analysis". In: *Economics of Education Review* 21.5. Publisher: Elsevier, pp. 513–521.

Piketty, Thomas, Li Yang, and Gabriel Zucman (July 2019). "Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015". In: *American Economic Review* 109.7, pp. 2469–2496.

Ram, Rati (1990). "Educational Expansion and Schooling Inequality: International Evidence and Some Implications". In: *The Review of Economics and Statistics* 72.2. Publisher: The MIT Press, pp. 266–274.

Romer, Paul M. (1990). "Endogenous Technological Change". In: *Journal of Political Economy* 98.5. Publisher: University of Chicago Press, S71–S102.

Self, Sharmistha and Richard Grabowski (2004). "Does education at all levels cause growth? India, a case study". In: *Economics of Education Review* 23.1. Publisher: Elsevier, pp. 47–55.

Sianesi, Barbara and John Van Reenen (2003). "The Returns to Education: Macroeconomics". In: *Journal of Economic Surveys* 17.2, pp. 157–200.

Singh, Madhu (2001). "Reflections on Colonial Legacy and Dependency in Indian Vocational Education and Training (VET): A societal and cultural perspective". In: *Journal of Education and Work* 14.2, pp. 209–225.

Toivanen, Otto and Lotta Väänänen (2016). "Education and Invention". In: *The Review of Economics and Statistics* 98.2. Publisher: MIT Press, pp. 382–396.

UNESCO (Jan. 1, 1958). *World Survey of Education Primary Education: II*. United Nations Educational, Scientific and Cultural Organization.

— (Jan. 1, 1961a). *World Survey of Education Higher Education: IV*. United Nations Educational, Scientific and Cultural Organization.

— (Jan. 1, 1961b). *World Survey of Education Secondary Education: III*. United Nations Educational, Scientific and Cultural Organization.

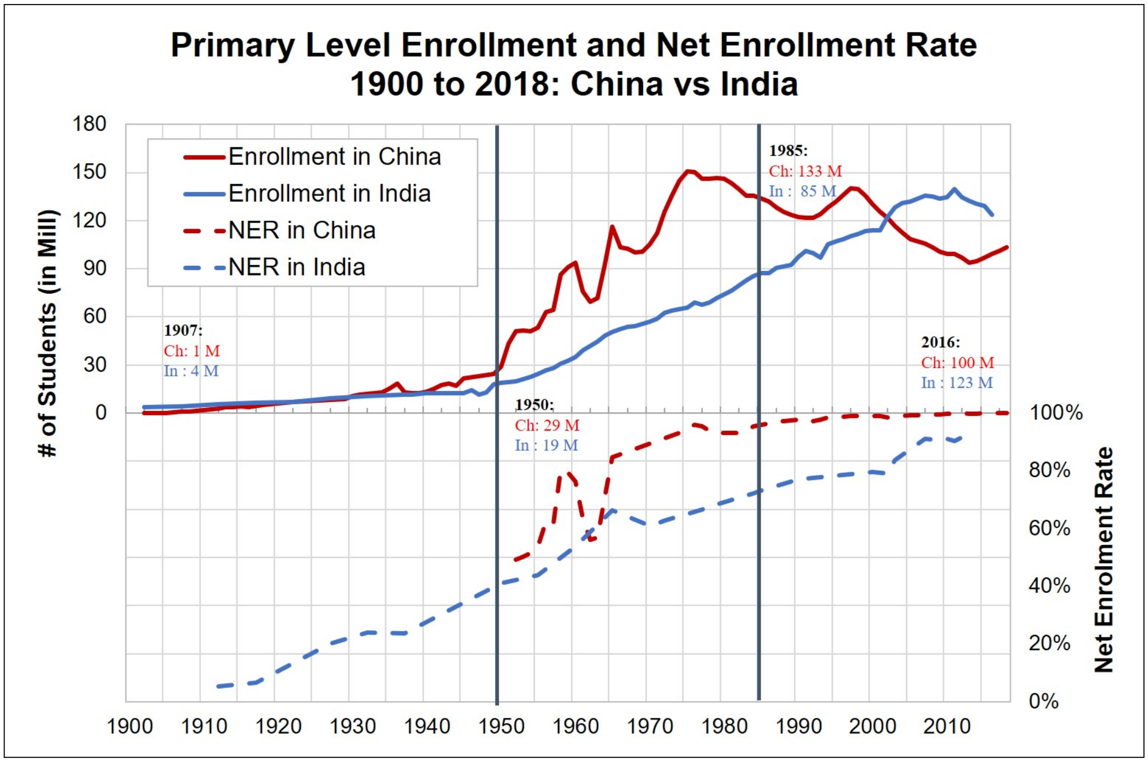
Vandenbussche, Jérôme, Philippe Aghion, and Costas Meghir (June 1, 2006). "Growth, distance to frontier and composition of human capital". In: *Journal of Economic Growth* 11.2, pp. 97–127.

Wolff, Edward and Maury Gittleman (1993). "The role of education in productivity convergence: Does higher education matter?" In: *Explaining economic growth*. Ed. by Bart van Ark and Dirk Pilat. Elsevier Science Publishers, pp. 147–167.

Woodhall, M. (1992). "Economic development and higher education". In: *The Encyclopedia of Higher Education* 2, pp. 889–895.

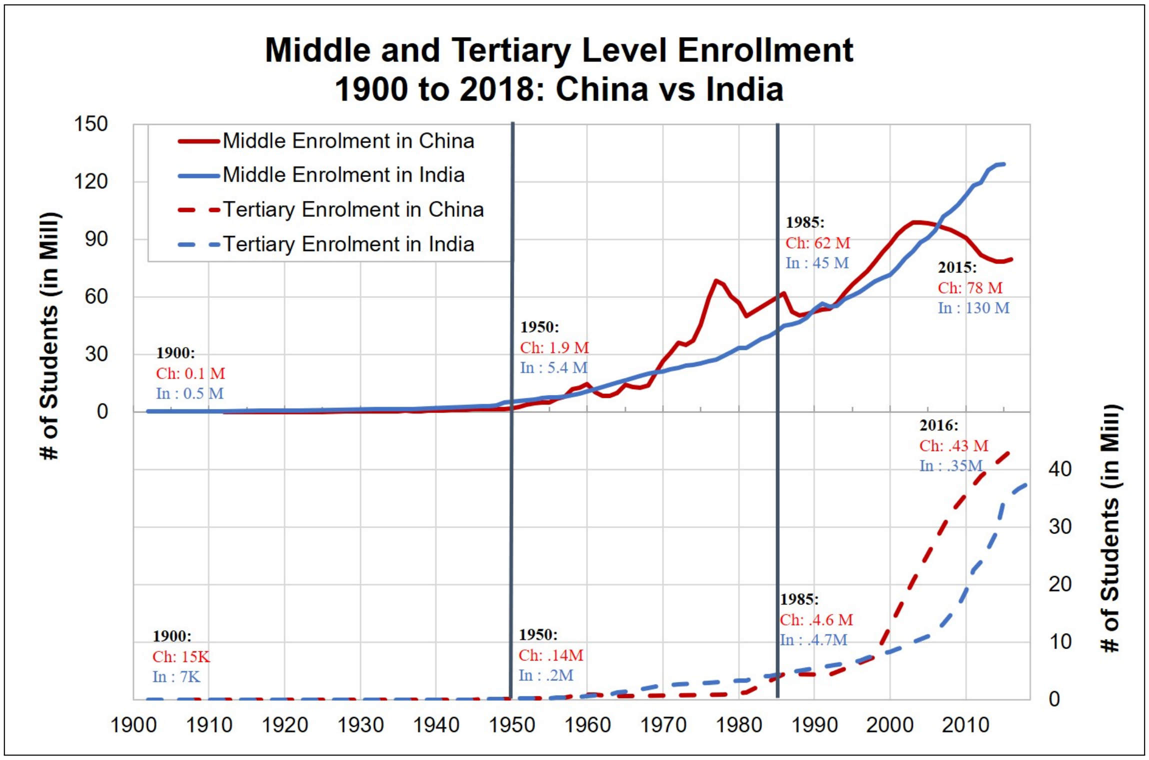
1.7. Figures

Figure 1.1. Evolution of Total enrollment and Net enrollment Rate at Primary Stage



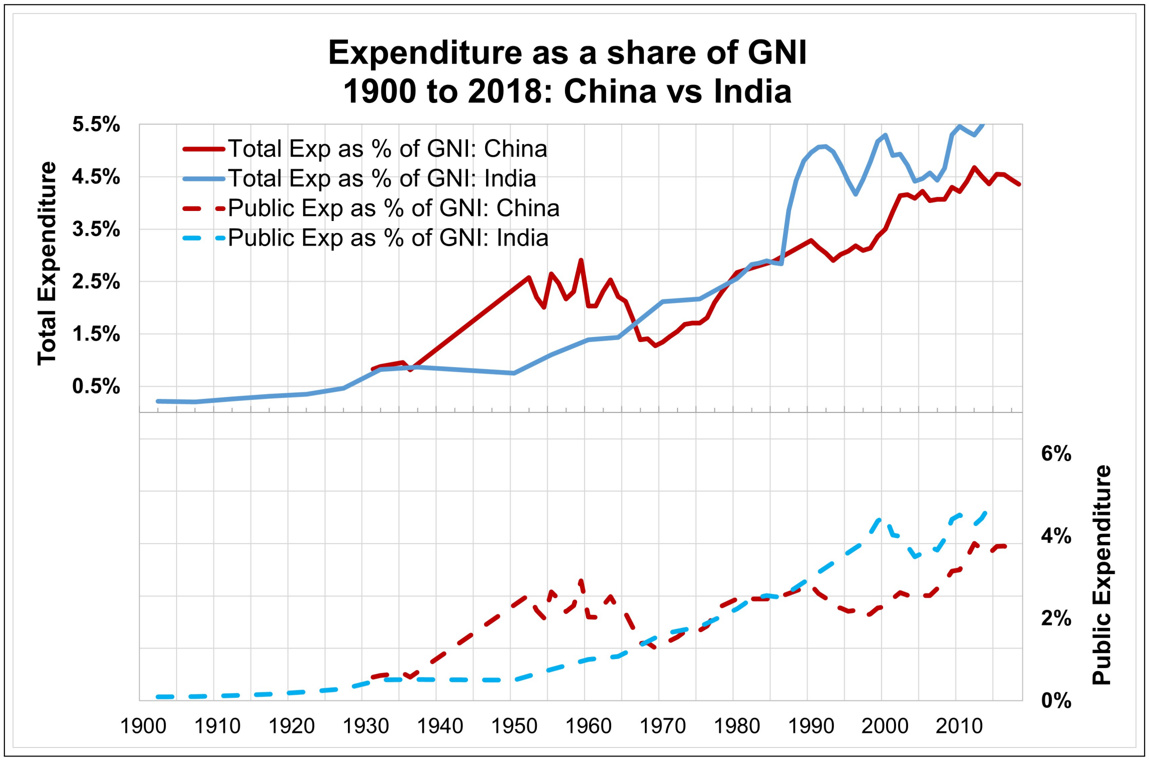
Notes: The figure plots the evolution of total enrollment (top part) and Net Enrollment Rate (bottom part) at the primary level of education in China and India from 1900-2018. The enrollment at the primary level in China surpassed India in the 1930s. The NER in China was $>$ 90% compared to $<$ 80% in India at the time of economic liberalization (China in 1978 and India in 1990).

FIGURE 1.2. Evolution of Total enrollment at Middle and Higher Level enrollment



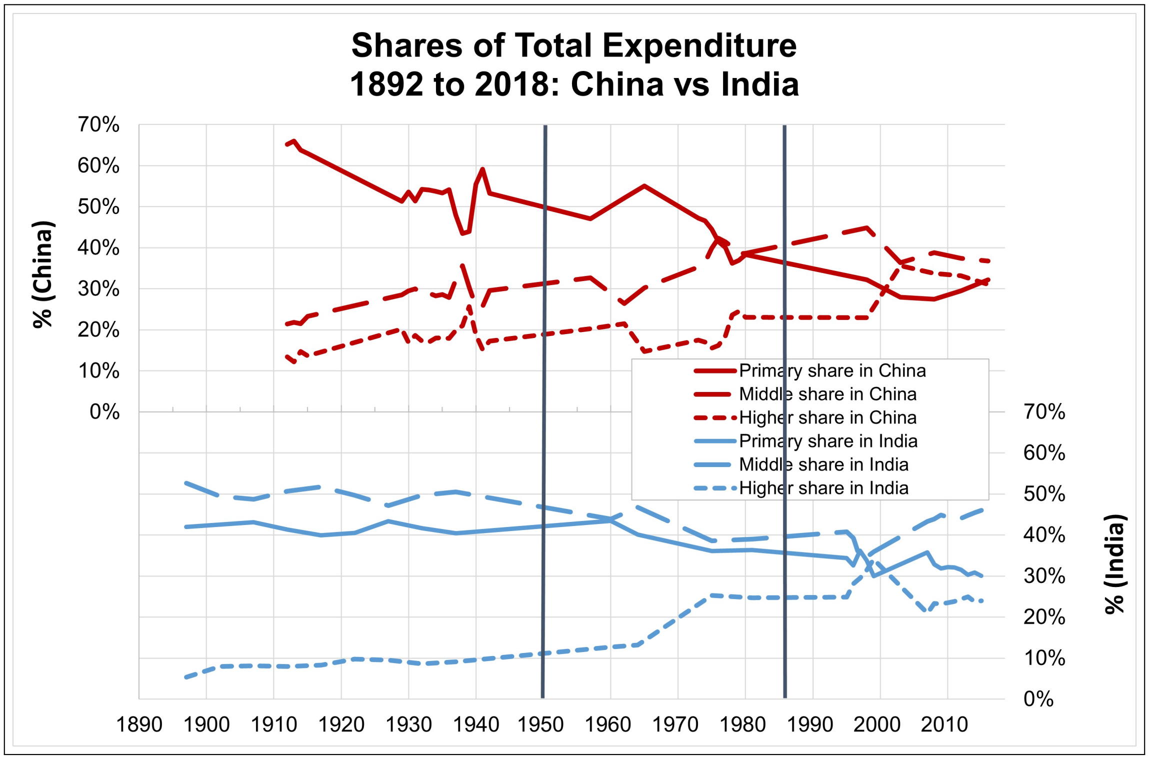
Notes: The figure plots the evolution of total enrollment at the middle (top part) and higher (bottom part) level of education in China and India from 1900-2018. The middle and higher-level enrollment in China surpassed India in the 1970s and 2000s.

FIGURE 1.3. Total Expenditure and Public Expenditure as % of GNI



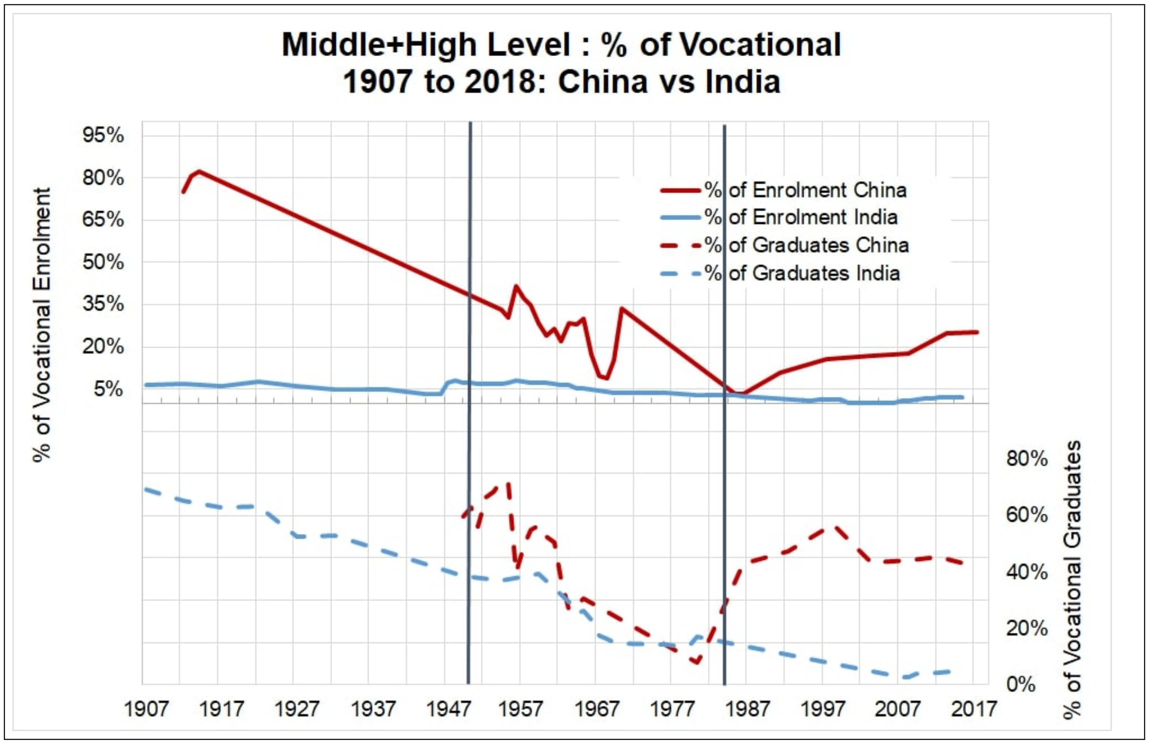
Notes: The figure plots the total education expenditure (top part) and public education expenditure (bottom part) as a share of Gross National Income. The slump in the 1930s and '40s is evident in India. In both countries, in recent years, public spending has been around 4.5% of GNI on education.

FIGURE 1.4. Share of Expenditure in Primary and Middle Stage



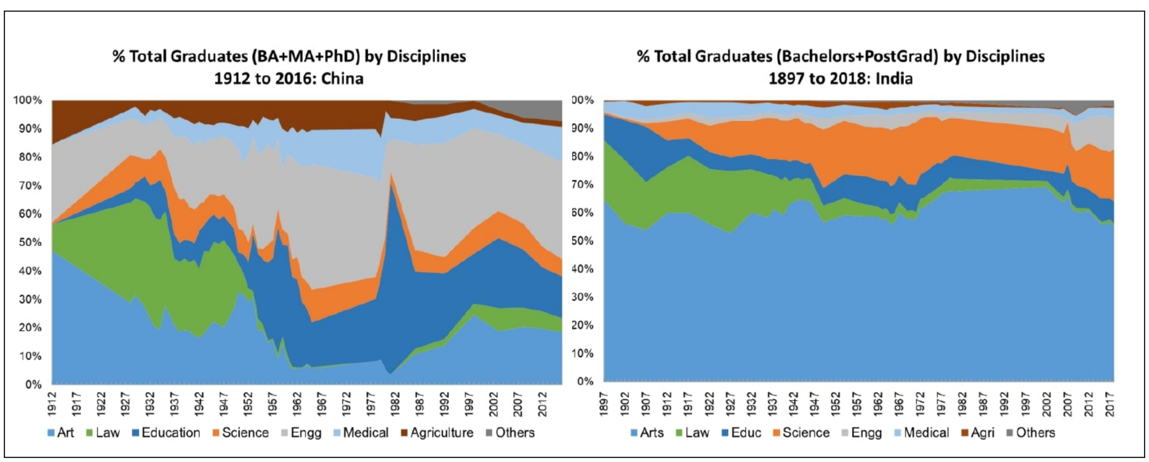
Notes: The figure plots the evolution of the expenditure share at primary, middle and higher stages of education in China (top part) and India (bottom part) from 1897-2018. The expenditure share was the highest in China's primary stage before the 1970s middle stage after the 1980s. The expenditure share is always highest in the middle stage in India, and during the 1960s, the share in higher education increased dramatically.

FIGURE 1.5. Vocational Education Share in Enrollment and Graduates



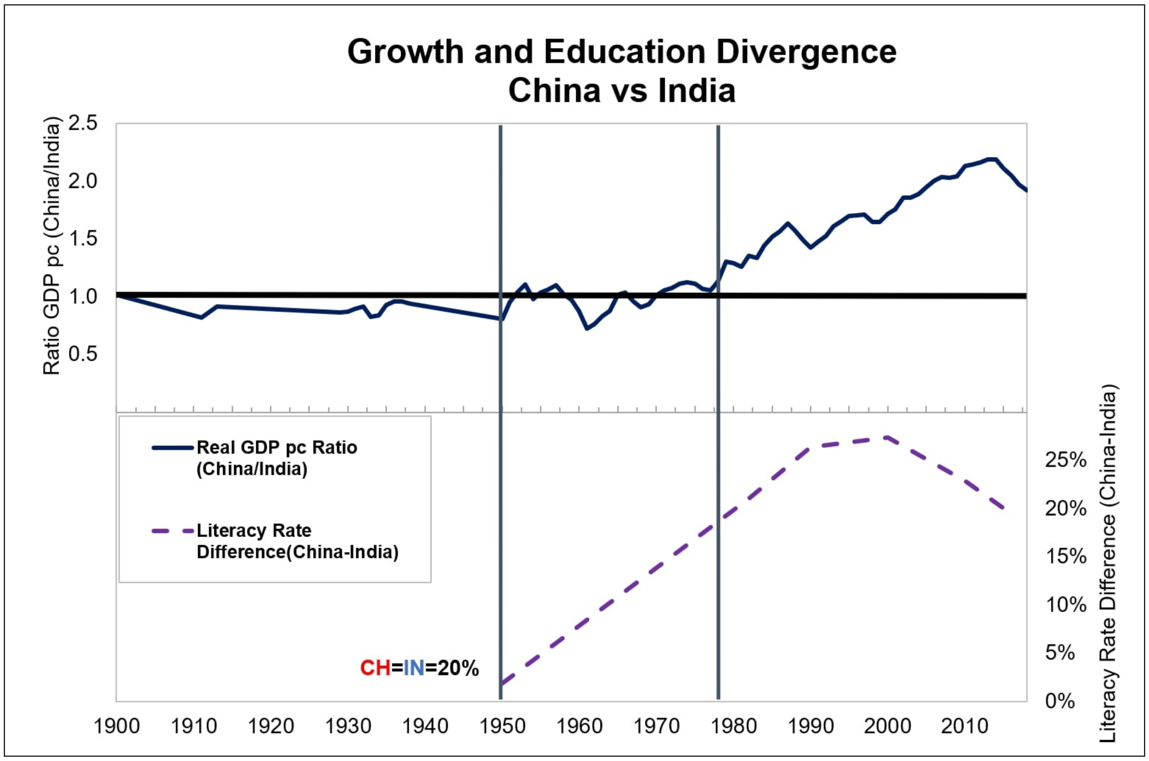
Notes: The figure plots the share of vocational enrollment (top part) and graduates (bottom part) combined for Middle and Tertiary levels of education. China has a higher share of vocational enrollment. It was negatively affected during 1966-76, but the vocational share has consistently increased after opening up its economy. India has a meagre share of vocational education.

FIGURE 1.6. % of Graduates by Discipline at Higher Education



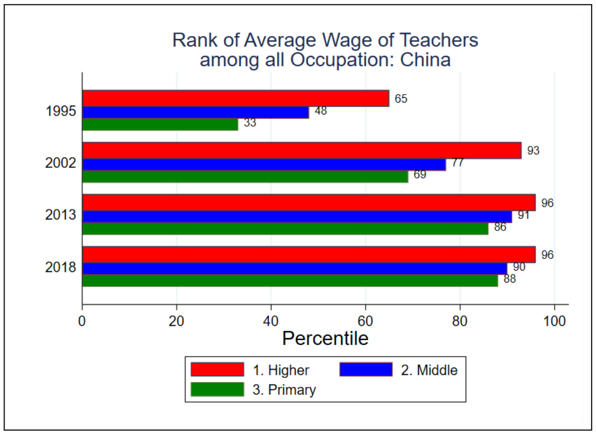
Notes: The figure depicts the share of graduates in different disciplines in China and India between 1897-2017. Both countries produce a very diverse mix of graduates. China has a higher percentage of engineers, and India produces a higher share of Humanities graduates.

FIGURE 1.7. Economic and Educational Divergence between China and India

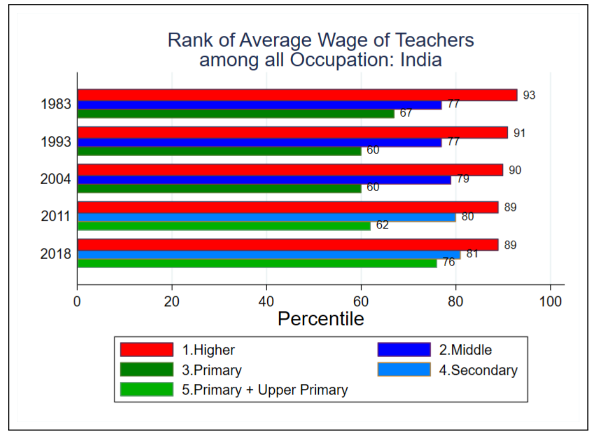


Notes: The top part presents the evolution of the ratio of GDP per capita (China/India) from 1900-2018. The bottom part shows the difference in the literacy rate between China and India from 1950-2018. The literacy rate divergence started 30 years before the economic divergence between these two countries.

FIGURE 1.8. Rank Percentile of Average Wage of Teachers



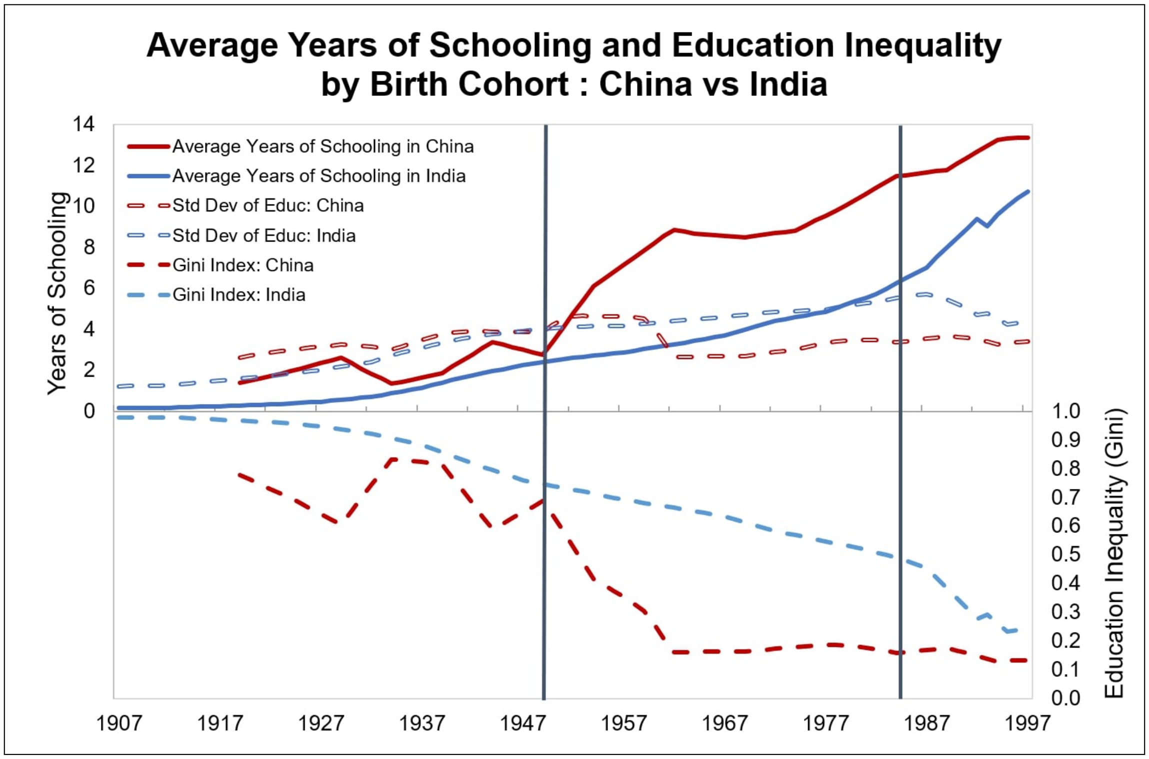
China



India

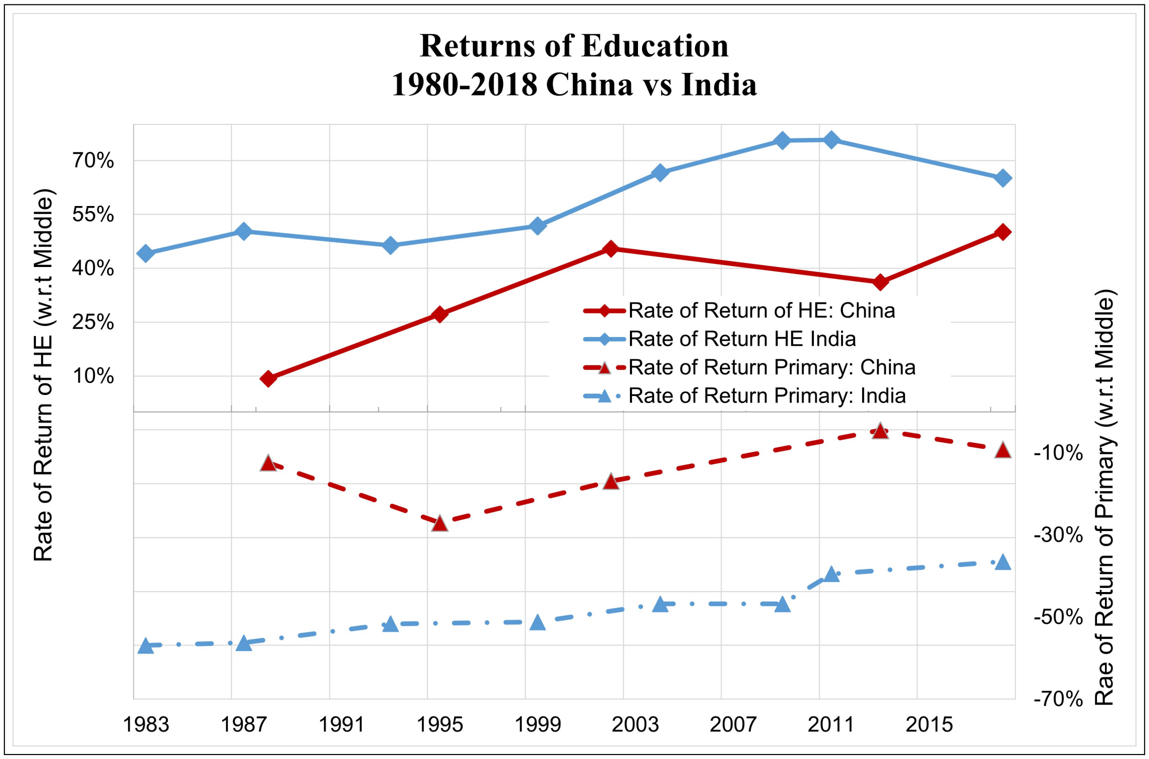
Notes: The figure plots the percentile of primary, middle and tertiary level teachers' salaries in the wage distribution of the salaried population (in 20-60 years) and positive income. In the 1990s, the percentile of teachers was lower in China at all levels than in India. Post-2000, after China started focusing on quality, the rank percentile of teachers' salaries increased (and became similar to India). The surveys of 2011 and 2018 do not allow to split Primary (Class I-V) and Upper Primary (Class VI-VIII) teachers' occupations. In 2018, there was a big jump in the primary+middle level teachers' salary rank in India.

FIGURE 1.9. Average Years of Education and Education Inequality



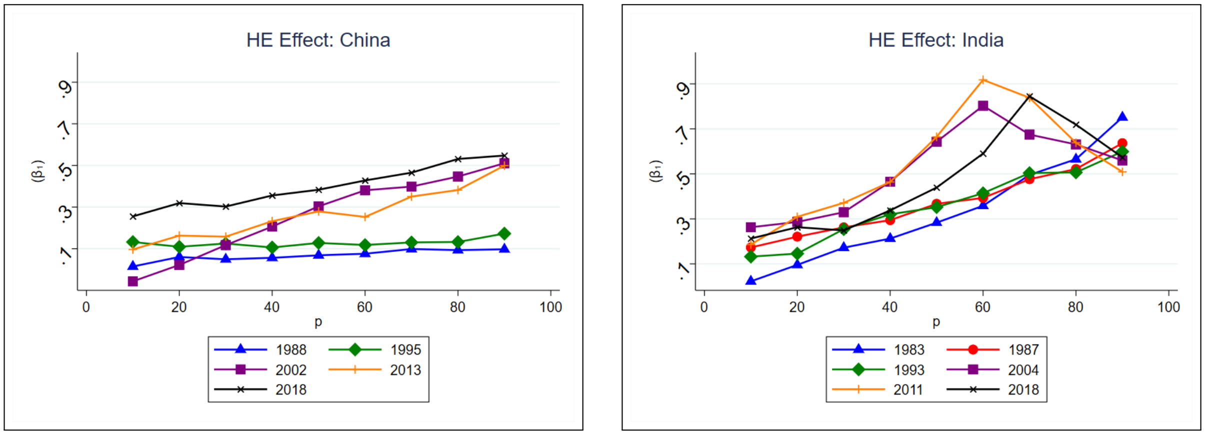
Notes: The figure plots the evolution of average years of schooling and education inequality measures (standard deviation of education and gini) by birth cohort. Due to the expansion of mass-level primary education first in China has better average years of education and lower education inequality.

FIGURE 1.10. Wage effect from Mincer's equation



Notes: The figure plots the wage effect for college graduates (top part) and primary graduates (bottom part) compared to the middle-level graduates. It is simply the coefficients on the college and primary dummies in Equation 4 (standard Mincer's equation). The tertiary wage effect is always higher in India (than in China) and has increased over the years. In China, there was a drop between 2002-2013 because of a rapid increase in the supply of tertiary graduates.

FIGURE 1.11. Wage Effect of Higher Education at different deciles



China India

Notes: The figure plots the coefficients on the tertiary graduates from the Unconditional Quantile Regression (Equation 5) with quantile as a distribution statistic. It captures the tertiary wage effect (compared to the middle level) at different wage quantiles. In China, there was no differential wage effect at different wage quintiles in the 1980s and 90s, though it was monotonically increasing in India. Post-2000, there is a shift in the curve in both countries. The UQR curve of China becomes similar to India's in the 1980s and 90s. The UQR curve of India becomes inverted U shape, with a peak near 60-70 percentile.

1.8. Tables

Table 1.1. Expansion in Education: Average Flow of Variables

	Enrollment (in Mill)		**Graduates (in Mill)**		**Teachers (in Mill)**		**Expenditure (Nom USD in Mill)**	
	CH	IN	CH	IN	CH	IN	CH	IN
1900-1925	5	6		0.0	0.3	0.3	37	27
1926-1950	17	14	0.0	0.0	0.6	0.5	89	86
1951-1985	134	75	0.2	0.6	5.3	2.3	2,848	1,760
1986-2000	199	166	1.4	2.4	10.0	4.5	24,131	8,933
2001-2018	227	253	7.7	6.0	12.3	7.4	242,645	60,875

Notes: This table provides impeccable evidence for the term "Human-capital century" for the 20^{th} century. The education system has become gigantic - absorbing billions of dollars, providing direct employment to millions of teachers and staff and generating millions of high-skilled labour force every year.

Table 1.2. Revenue share by Stages : Average over time periods

% Relative Share	Primary		Middle		Higher	
	CH	IN	CH	IN	CH	IN
1900-49	61%	40%	25%	51%	14%	8%
1950-65	51%	44%	29%	38%	19%	18%
1966-85	39%	37%	39%	34%	21%	28%
1986-2006	30%	35%	41%	44%	29%	21%
2007-2018	30%	33%	39%	45%	32%	23%

Notes: The table reports the average value of revenue allocation share at the primary, middle and higher stages of education. In India, pre-independence, the middle share was the highest and post-independence, there was a big jump in higher education. There is a gradual transition from primary to the middle to the higher stage of education in China.

Table 1.3. EIR Decomposition : Average Values

Decomposition of EIR (= GER*Quality)

	Primary				Middle				Higher			
	Quantity (GER)		Quality (Exp/Stud /GNIpc)		Quantity (GER)		Quality (Exp/Stud /GNIpc)		Quantity (GER)		Quality (Exp/Stud /GNIpc)	
	CH	IN	CH	IN	CH	IN	CH	IN	CH	IN	CH	IN
1900-30	8%	18%		5%	0.4%	2%		67%	0.10%	0.08%		492%
1930-49	28%	32%	15%	8%	2%	6%	184%	79%	0.23%	0.27%	1565%	634%
1950-65	70%	55%	11%	7%	8%	16%	67%	20%	0.94%	1.96%	703%	178%
1965-85	110%	80%	5%	9%	39%	29%	15%	18%	1.32%	6.68%	572%	140%
1986-00	103%	90%	10%	14%	58%	42%	23%	32%	10.8%	10.1%	132%	118%
2001-15	107%	102%	15%	15%	80%	60%	25%	25%	38.9%	19.2%	64%	84%

Decomposition of Quality (= Quality1*Quality2)

	Primary				Middle				Higher			
	1/Quality1 (PTR)		Quality (Exp/Teachers /GNIpc)		1/Quality1 (PTR)		Quality (Exp/Teachers /GNIpc)		1/Quality1 (PTR)		Quality (Exp/Teachers /GNIpc)	
	CH	IN	CH	IN	CH	IN	CH	IN	CH	IN	CH	IN
1900-30	16	27		134%	14	17		1032%	9	15		7087%
1930-49	27	31	360%	229%	16	19	2467%	1498%	8	13	11224%	6965%
1950-65	33	35	344%	244%	22	23	1550%	460%	15	19	10127%	3556%
1965-85	29	39	145%	320%	20	26	289%	488%	7	20	4875%	2746%
1986-00	23	44	228%	659%	16	30	397%	1017%	15	22	2160%	2818%
2001-15	19	42	292%	629%	16	31	405%	796%	29	22	1590%	1915%

Notes: The table presents the average values of quantitative and qualitative components of EIR by period for all three levels separately. The quantitative component (GER) in China becomes more than in India at primary (from 1950), middle (from the mid-1960s) and higher (from 1990s) levels at different periods. The qualitative component (expenditure per student/GNIpc) in India becomes more than in China at primary (from the 1960s), middle (from 1960s) and higher (from 2000s) levels at different periods. The second part of the table presents the average values of PTR and exp/teachers/GNIpc by period for all three levels separately.

Table 1.4. Income, Education and Demographic Characteristics

PANEL A: CHINA	(1)	(2)	(3)	(4)	(5)
Years	1988	1995	2002	2013	2018
Avg Annual Wage ($ 2018)	825	1,578	2,081	4,961	7,807
Avg Wage per Day ($ 2018)	2.85	5.62	8.38	21.41	32.20
Age	37.5	38.3	39.5	37.9	38.3
Female	0.46	0.45	0.37	0.38	0.42
Primary	.14	0.08	0.11	0.13	0.08
Middle	0.74	0.72	0.70	0.67	0.56
Tertiary	0.13	0.20	0.20	0.20	0.36
Primary Indus(%)	2.4	2.4	2.0	3.2	2.5
Secondary Indus(%)	50.1	48.3	44.8	48.1	34.0
Service Indus(%)	47.1	49.3	53.2	48.7	63.5
Urban Share(%)	93.4	85.9	57.7	36.7	70.7
Observations	18,411	12,091	15,784	21,055	20,200

PANEL B: INDIA	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years	1983	1987	1993	1999	2004	2009	2011	2018
Avg Annual Wage ($ 2018)	1,426	1,849	1,982	2,579	2,545	3,296	3,355	2,921
Avg Wage per Day ($ 2018)	3.96	6.10	5.60	7.24	7.00	9.10	9.22	11.33
Age	35.5	36.5	36.8	37.2	36.6	36.5	36.3	36.6
Female	0.13	0.17	0.14	0.15	0.19	0.18	.19	.22
Primary	0.44	0.37	0.33	0.28	0.30	0.22	.22	.19
Middle	0.42	0.43	0.46	0.49	0.45	0.45	.42	.44
Tertiary	0.14	0.20	0.21	0.23	0.25	0.33	.36	.37
Primary Indus(%)	11.0	3.97	5.63	6.6	4.2	2.8	2.3	2.0
Secondary Indus(%)	24.3	27.2	25.8	22.8	25.0	23.9	24.6	22.6
Service Indus(%)	64.8	68.8	68.6	70.1	70.9	73.3	73.1	75.4
Urban Share(%)	60.7	81.4	62.5	53.0	62.1	65.4	64.7	60.2
Observations	34,952	30,332	35,495	37,660	39,792	34,992	37,645	40,665

Notes: The table presents average values computed from the labour force surveys from both countries for the working age population (20-60 years old) employed in regular salaried jobs (and having positive income). Survey weights are used for all calculations. The labour composition has changed drastically in both countries, with increasing tertiary level graduates and declining Primary graduates. The share of tertiary graduates has increased from 13-14% in the 1980s to 36-37% in 2018 in both countries. Middle-level graduates have remained constant in India, and only the share of primary-level graduates has declined. In contrast, the share of primary and middle-level graduates in China has declined.

Table 1.5. Theil's Index and Decomposition

	(1)	(2)	(3)	(4)	(5)	(6)
	Theil's		**Between**		**Within**	
Years	China	India	China	India	China	India
1980's	0.123	0.25	0.001	0.067	0.121	0.18
			(1%)	*(28%)*	*(99%)*	*(72%)*
1990's	0.225	0.26	0.003	0.070	0.222	0.19
			(1%)	*(26%)*	*(99%)*	*(74%)*
2000's	0.342	0.40	0.067	0.12	0.276	0.28
			(20%)	*(30%)*	*(80%)*	*(70%)*
2010's	0.222	0.41	0.033	0.119	0.189	0.295
			(15%)	*(29%)*	*(85%)*	*(71%)*
2018	0.300	0.34	.049	.085	0.251	.252
			(16%)	*(25%)*	*(84%)*	*(75%)*

Notes: The table presents Theil's Wage inequality measure and its decomposition from the labour force surveys from both countries for the working age population (20-60 years old) employed in regular salaried jobs (and having positive income). Survey weights are used. Col (1) and Col (2) present Thiel's index. Col (3) and Col (4) present the between-components of decomposition (and in bracket the percentage share) by education group, where the groups are primary, middle and tertiary. Col (5) and Col (6) present the within-components of decomposition (and in bracket the percentage share). The between-component of wage inequality that captures the education effect is much higher in India compared to China.

TABLE 1.6. Evolution of Wage Effect in China and India

PANEL A:China	(1)	(2)	(3)	(4)	(5)
Years	1988	1995	2002	2013	2018
Primary	-0.124***	-0.270***	-0.169***	-0.0450***	-0.0919***
	(0.00927)	(0.0270)	(0.0202)	(0.0158)	(0.0179)
Tertiary	0.0926***	0.272***	0.455***	0.361***	0.501***
	(0.00721)	(0.0105)	(0.0144)	(0.0139)	(0.0119)
Observations	18,337	12,084	15,464	21,049	20,189
R-squared	0.320	0.254	0.431	0.132	0.188
Mean Dep Var	0.94	1.52	1.79	2.84	3.14
Province FE	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes

PANEL B: INDIA	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years	1983	1987	1993	1999	2004	2009	2011	2018
Primary	-0.569***	-0.562***	-0.516***	-0.512***	-0.468***	-0.468***	-0.395***	-0.365***
	(0.0103)	(0.0122)	(0.0129)	(0.0152)	(0.0142)	(0.0179)	(0.0166)	(0.0149)
Tertiary	0.441***	0.503***	0.464***	0.518***	0.666***	0.755***	0.757***	0.651***
	(0.0132)	(0.0132)	(0.0146)	(0.0168)	(0.0151)	(0.0197)	(0.0149)	(0.0118)
Observations	34,915	30,323	35,478	37,628	39,774	34,988	37,641	40,657
R-squared	0.427	0.412	0.294	0.404	0.462	0.425	0.433	0.378
Mean Dep Var	1.16	1.58	1.52	1.76	1.65	1.90	1.95	2.13
Province FE	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes

Notes: The table presents results from Mincer's Equation 4 for China (Panel A) and India (Panel B) from the labour force surveys from both countries for the working age population (20-60 years old) employed in regular salaried jobs (and having positive income). Survey weights are used. The outcome variable is log daily wage in real dollars ($) and the main explanatory variable is a dummy for tertiary education. The controls include age, age square, gender dummy and urban dummy (1/0). All equations include state/Province fixed effects.

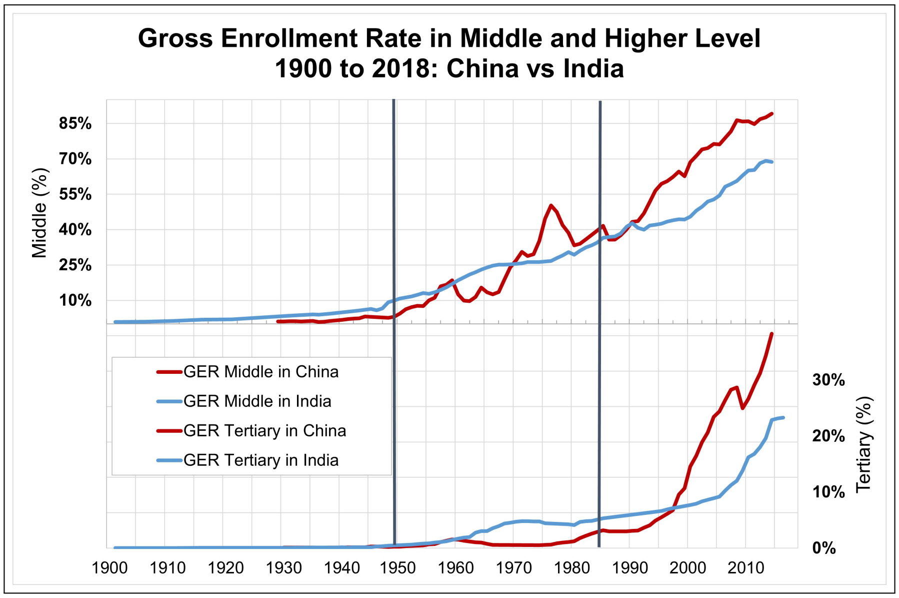
TABLE 1.7. Unconditional Partial Impact of Education on Inequality (Annual Wage)

	Primary		**Tertiary**	
Years	China	India	China	India
1983		.183***		.343***
1987	.027***	.222***	.027***	.221***
1993/95	.147***	.296***	-.015	.280***
1999		.229***		.183***
2002/04	.443***	.249***	.414***	.208***
2009		.262***		.225***
2011/13	.253***	.211***	.215***	.203***
2018	.375***	.180***	.186***	.159***

Notes: The table presents the coefficients on the dummy for primary and tertiary graduates from running RIF linear regressions (Equation 5) with distribution statistic as Variance of log of Wage. These coefficients are Col(2) or Col (7) of Appendix Tables (1.11-1.16). The regressions include Industry and Occupation FE. The standard errors are robust standard errors.

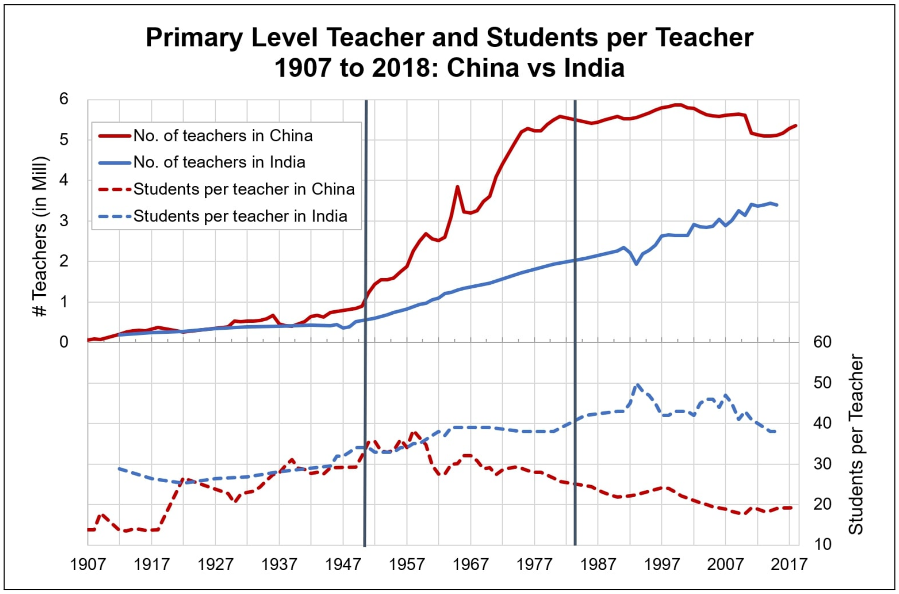
A. Appendix Figures

FIGURE 1.12. Gross Enrollment Rate at Middle and Higher Level



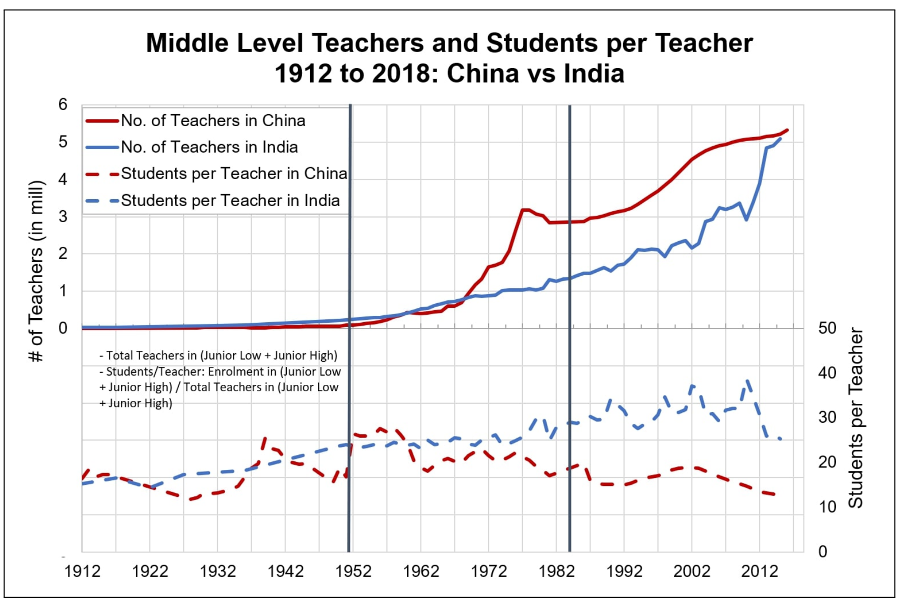
Notes: This figure plots the evolution of gross enrollment rate at middle and higher level in China and India from 1900-2015. The base population for middle level is the population of kids between 12-18 years for China, 11-18 years for India. The base population for higher level 18-24 years for both countries.

FIGURE 1.13. Primary Level Teachers



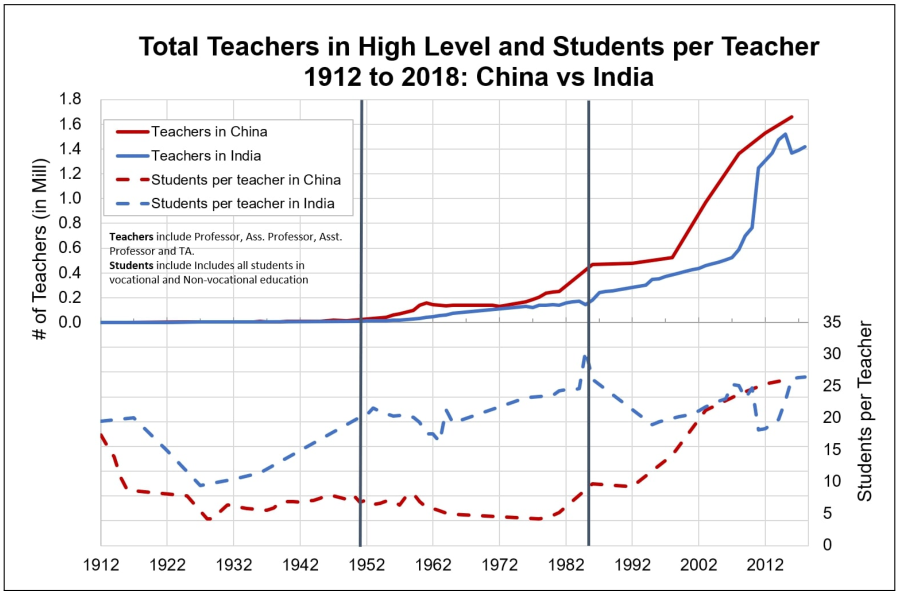
Notes: This figure plots the evolution of total teachers (top) and pupil-teacher ratio (lower) at the primary stage from 1907-2018 in China and India. The number of primary teachers has remained higher in China since the 1930s. The PTR at the primary level has been better in China since the 1960s.

FIGURE 1.14. Middle Level Teachers



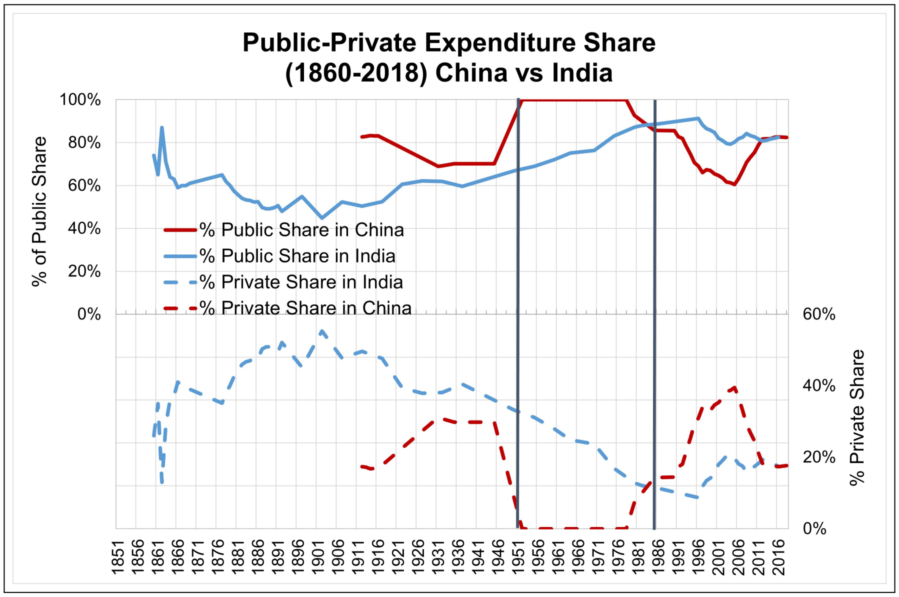
Notes: This figure plots the evolution of total teachers (top) and pupil-teacher ratio (lower) in the middle stage from 1912-2018 in China and India. The number of middle teachers has remained higher in China since the 1960s. The PTR at the middle level is also better in China since the 1960s.

FIGURE 1.15. Tertiary Level Teachers



Notes: This figure plots the evolution of total teachers (top) and pupil-teacher ratio (lower) in the tertiary stage from 1912-2018 in China and India. The tertiary-level teachers has remained higher in China since the 1950s, and the PTR at the tertiary level has been better in China throughout the 20th century. The PTR deteriorated in the late 2000s due to the rapid increase in tertiary level enrollment.

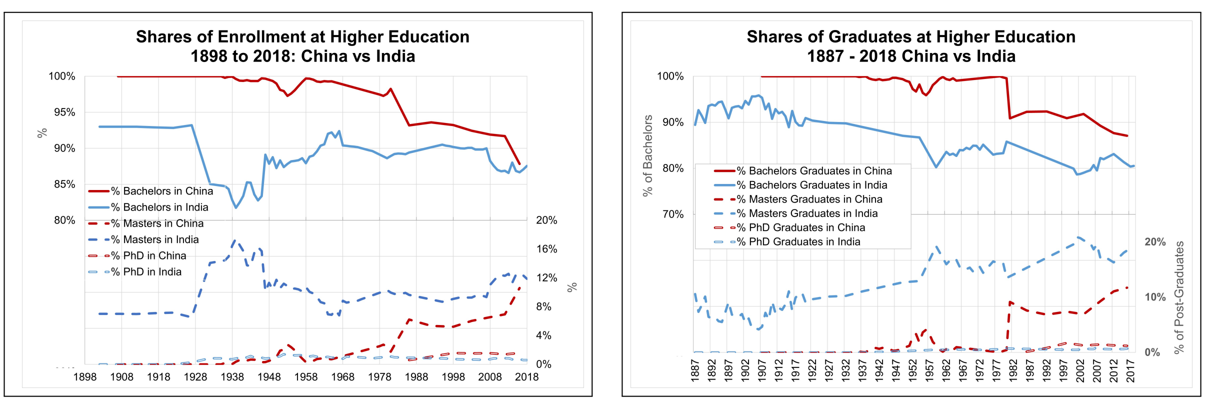
FIGURE 1.16. Public-Private Share of Total Expenditure



Notes: Total expenditure on education is split into the public and private components. At the beginning of the century, the private share in education was highest at 60% in India. It kept decreasing for the next 80-90 years, i.e. before liberalization, when it hit the lowest 10%. Post-1990s, there is a reversal, and the private share stands at around 20%. It is due to an increase in the private institutions (i.e. increasing fees) and more focus on government towards elementary education.

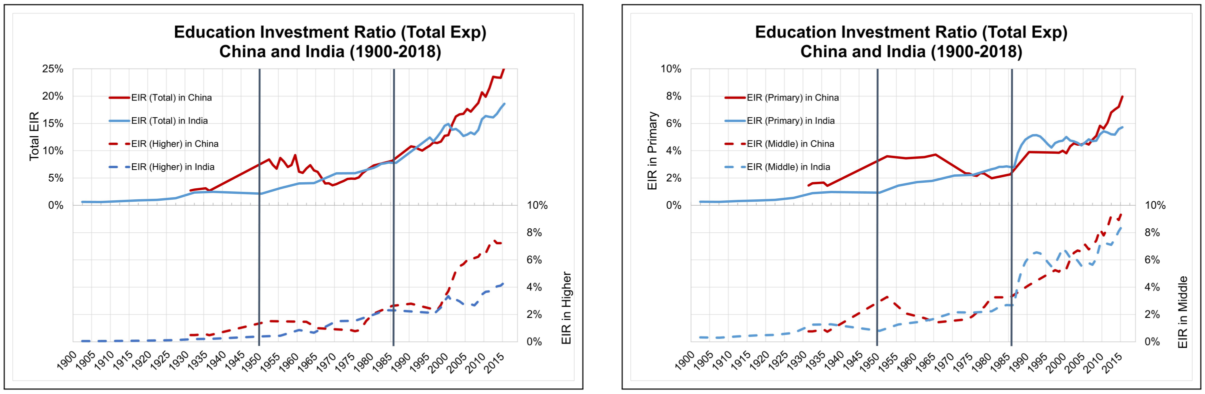
FIGURE 1.17. Share of Enrollment and Graduates at Higher Education

Enrollment Graduates



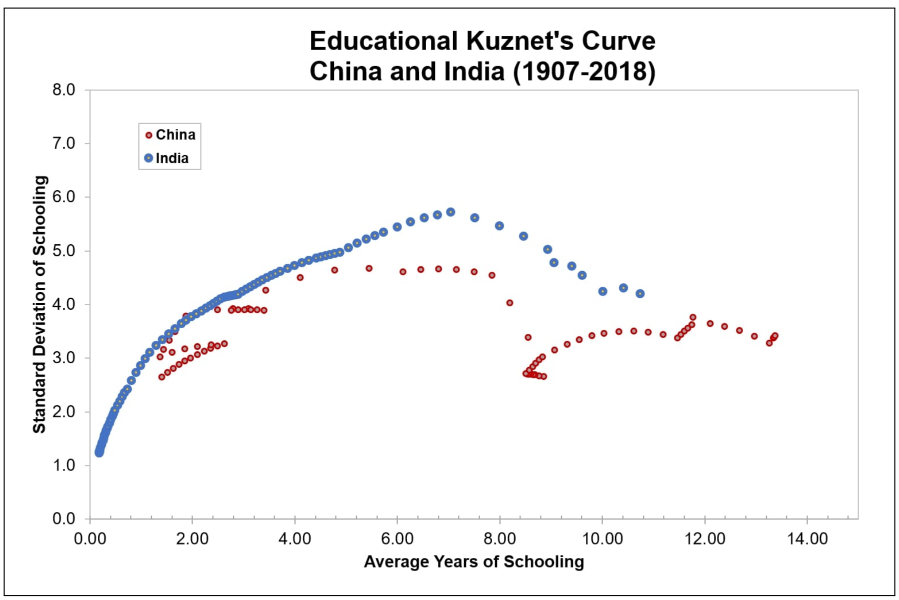
Notes: This figure plots the evolution of share of Bachelors, Masters and PhD's at higher level in China and India from 1900-2018. After 2000, close to 80% Bachelors, 19% Masters and less than 1% PhD degrees are offered in India, compared to 89%, 10% and 1.5% in China.

FIGURE 1.18. Evolution of EIR with Total Expenditure



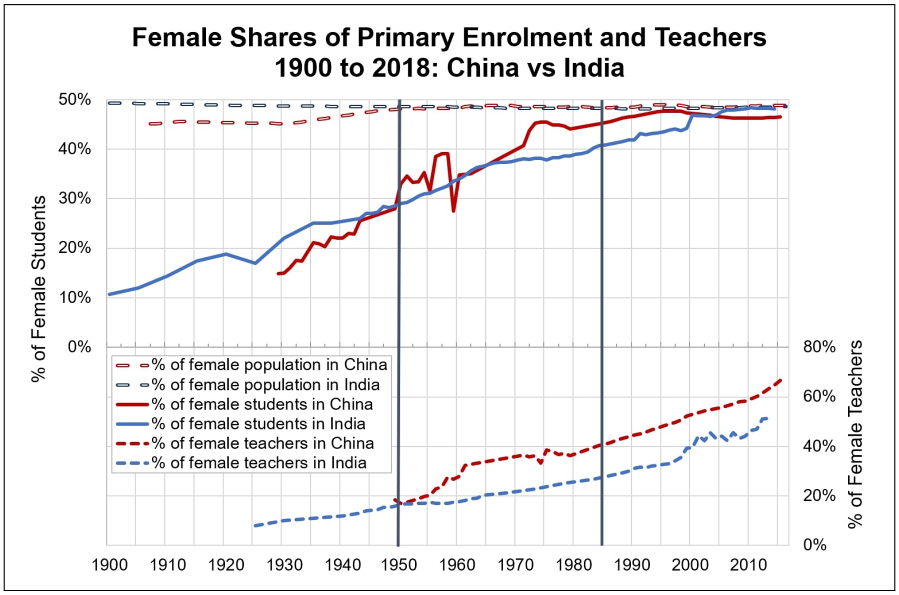
Notes: The figures plots the evolution of EIR using total expenditure. The left figure plots for total EIR and EIR at higher level whereas the right figure plots for EIR at primary and middle level of education.

FIGURE 1.19. Education Kuznet's Curve



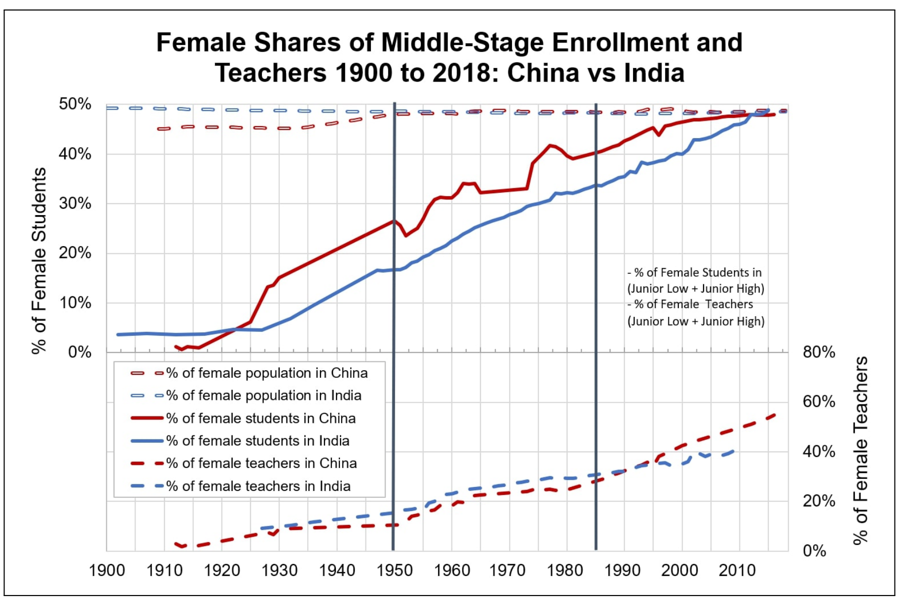
Notes: This figure plots the evolution of Standard Deviation in Schooling and Average years of schooling by birth cohort. It shows the existence of inverse U curve, also called Education Kuznet's curve in both the countries. The drop starts at 7 years in India compared to 7.85 in China.

FIGURE 1.20. Female Share at Primary Level



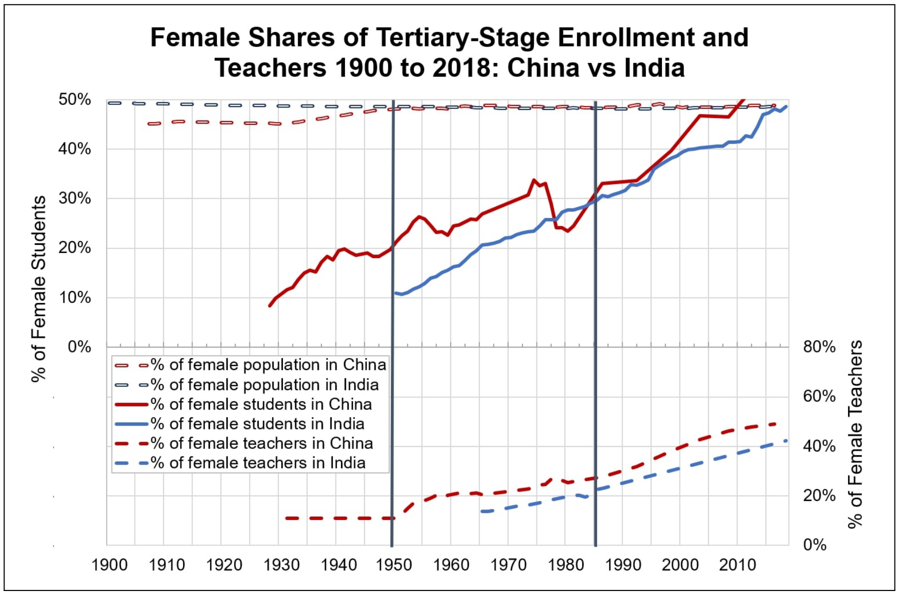
Notes: This figure plots the evolution of share of female enrollment (female enrollment/total enrollment) and female teachers (female teachers/total teachers) at primary stage along with share of female population in China and India from 1900-2018. Both countries have now bridged the gender gap in enrollment taking more than 100 years.

FIGURE 1.21. Female Share at Middle Level



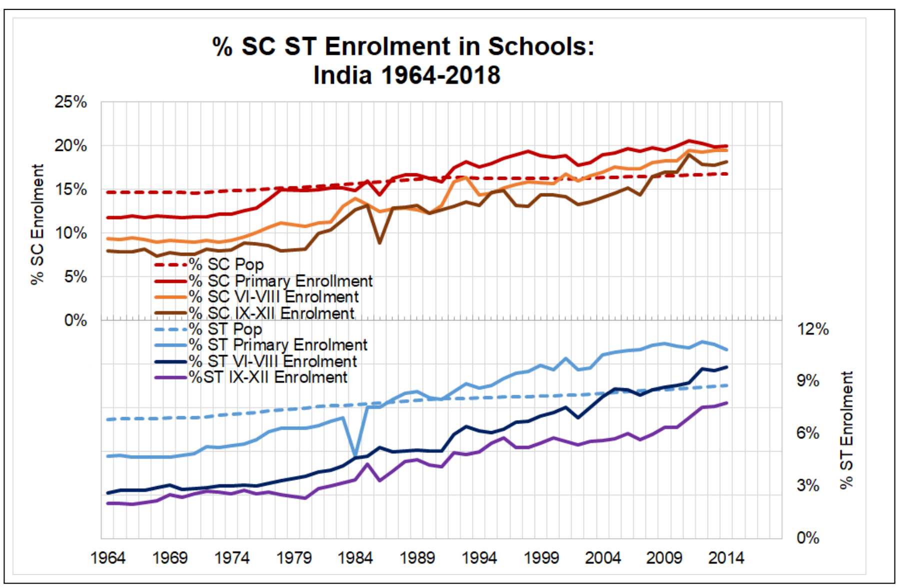
Notes: This figure plots the evolution of share of female enrollment (female enrollment/total enrollment) and female teachers (female teachers/total teachers) at middle stage along with share of female population in China and India from 1900-2018.

Figure 1.22. Female Share at Tertiary Level



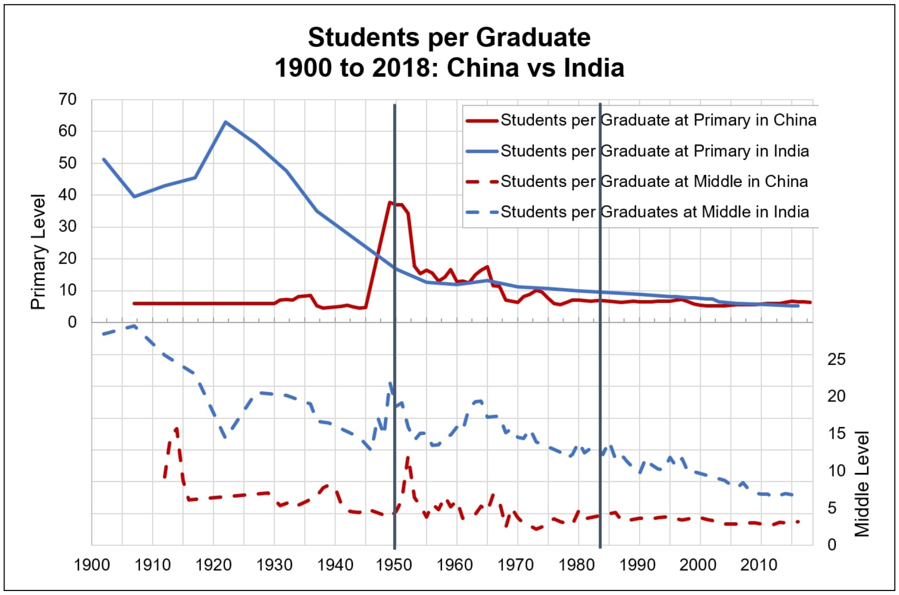
Notes: This figure plots the evolution of share of female enrollment and share of female teachers at tertiary stage along with share of female population in China and India from 1900-2018.

Figure 1.23. % Share of Enrollment of Scheduled Caste and Tribe



Notes: This figure plots the evolution of enrollment at school level along with population share of Scheduled Caste (top) and Scheduled Tribe (bottom). The caste-group gap in enrollment at school stage has been reduced for both the groups.

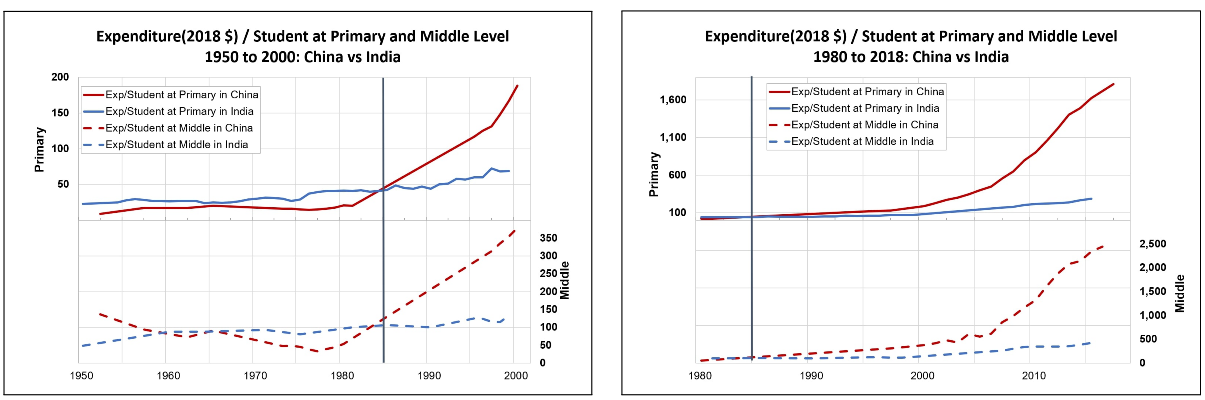
Figure 1.24. Students per Graduate in China and India



Notes: This figure plots the evolution of Students/Graduates in Primary and Middle stage of education in China and India from 1900-2018. China has better measure than India almost in all the years.

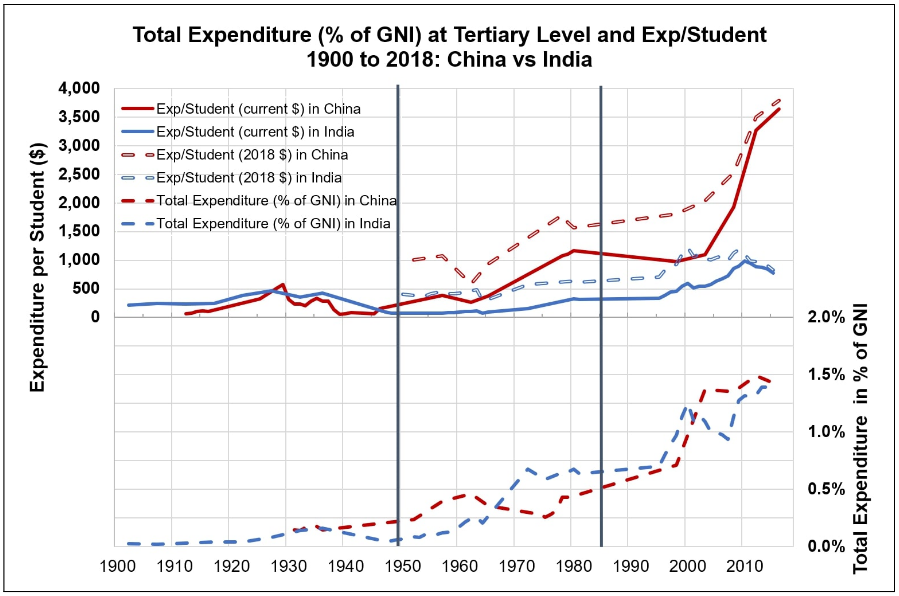
Figure 1.25. Expenditure per Student at Primary and Middle Level

1950-2000 1980-2018



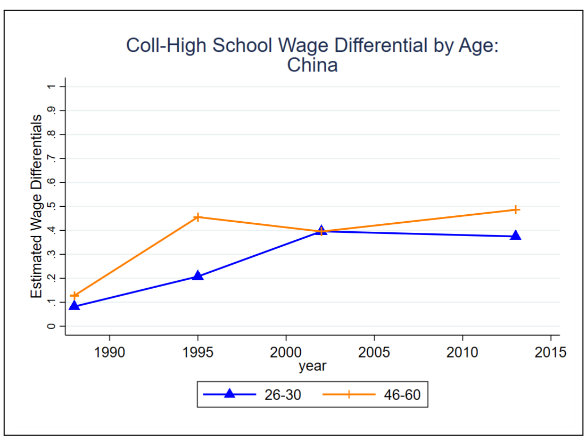
Notes: This figure plots the evolution of expenditure per student (in constant 2018 $) at primary and middle stage of education in China and India from 1900-2018. The measure is higher for India till late 1980's, after which it becomes higher for China.

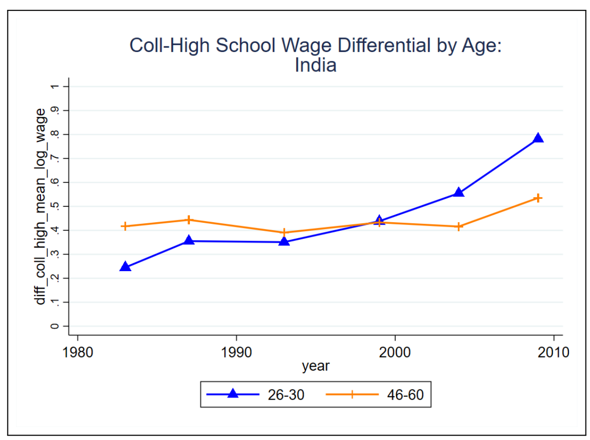
FIGURE 1.26. Expenditure per Student at Tertiary level



Notes: This figure (top) plots the evolution of expenditure per student (in current and constant $) at tertiary stage of education in China and India from 1900-2018. The values are always higher in China due to development of expensive streams. The bottom part shows the total expenditure at tertiary level as a share of gross national income.

FIGURE 1.27. Difference in College and Secondary Graduates Wages by Cohort





China India

Notes: The figure plots the average of the difference in log wage between Tertiary and Middle school graduates for two cohorts in China and India. The high wage effect in India comes mainly from the younger cohort.

B. Appendix Tables

TABLE 1.8. Share of Vernacular Middle Schools in India

% Vernacular Middle Schools				
Year	Total	*Within Public*	*Within Aided Private*	*Within Unaided Private*
1897	52%	65%	50%	32%
1902	49%	63%	49%	22%
1907	50%	66%	46%	20%
1912	47%	65%	46%	12%
1917	47%	73%	47%	5%
1922	53%	80%	44%	5%
1927	60%	88%	38%	5%
1932	60%	85%	33%	3%
1937	58%	83%	30%	4%

Notes: The table provides the share of the vernacular middle level schools - total, within public schools and private schools(aided i.e. receiving grant-in-aid from the government and non-aided). Rest of the schools are English medium schools. This is to show the contrast that the private initiatives were more geared towards English medium. The medium of instructions were native languages in Primary school, a mix of native and English in Middle level and English in Colleges.

TABLE 1.9. Average Daily Wages (2018 \$) by Educational Categories

	Primary		**Middle**		**Tertiary**		**Tertiary/Middle**		**Middle/Primary**	
Years	China (1)	India (2)	China (3)	India (4)	China (5)	India (6)	China (7)	India (8)	China (9)	India (10)
1983		2.5		4.47		7		1 .79		1.57
1987	2.88	3.6	2.77	6.27	3.23	10.46	0.96	1.74	1.2	1.67
1993/95	5.22	3.31	5.38	5.56	6.65	9.2	1.03	1.68	1.24	1.66
1999		4.11		6.83		11.86		1.66		1.74
2002/04	4.46	3.41	7.39	5.89	14.24	11.92	1.66	1.73	1.93	2.02
2009		3.95		6.82		14.63		1.73		2.15
2011/13	20.26	4.22	19.39	6.76	28.73	15.13	0.96	1.61	1.48	2.24
2018	22.75	5.75	26.05	8.68	44	17.27	1.15	1.51	1.69	1.99

Notes: The table presents the average daily wage (2018\$) for primary, middle and tertiary graduates from the labour force surveys (salaried, between 20-60 years old and having a positive income) for China and India. The daily wage has doubled in India for all three stages in 35 years, whereas in China, it has increased by ten times in 30 years. Col (7) and (8) are the ratio of wages of tertiary and middle graduates for China and India, respectively. Except in 2002, the ratio is close to 1 in China. In India, the ratio has decreased from 1.8 to 1.5. Col (9) and (10) are the ratio of wages of middle and primary graduates for China and India, respectively. It has increased from 1.2 to 1.7 in China, whereas in India, it has increased from 1.6 to 2.0.

TABLE 1.10. Earnings Inequality Measures

PANEL A: CHINA	(1)	(2)	(3)	(4)	(5)
Years	1988	1995	2002	2013	2018
Gini	.25	.33	.45	.35	.39
Variance of log wages	.20	.38	1.02	.64	.76
IQ90_10	2.8	4.2	13.2	5.6	7.1
IQ90_50	1.6	1.9	2.7	1.9	2.4
IQ50_10	1.7	2.2	4.9	2.9	3.0
Observations	18,411	12,091	15,784	21,055	20,200

PANEL B: INDIA	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years	1983	1987	1993	1999	2004	2009	2011	2018
Gini	.40	.38	.39	.44	.48	.49	.48	.43
Variance of log wages	.62	.73	.91	.73	.82	.82	.81	.62
IQ90_10	7.4	7.0	8.0	8.8	10.5	10.2	9.0	7.0
IQ90_50	2.08	2.2	2.3	2.6	3.6	3.6	3.6	3.2
IQ50_10	3.57	3.0	3.6	3.5	2.9	2.9	2.5	2.2
Observations	34,952	30,332	35,495	37,660	39,792	34,992	37,645	40,665

Notes: The table presents several wage inequality statistics (gini, variance of log wage, p90/p10, p90/50 and p50/p10) from the labor force surveys of the salaried class population between 20-60 years old and having a positive income. In China, inequality jumps in early 2000s (from a very low during the 1980s and 90s). The rapid increase in the tertiary supply possibly helped in reducing the inequality in later years (the maximum jump is in p90/p10 from 2002 to 2013). In India, inequality is always higher than China, though now it seems to be declining.

TABLE 1.11. Unconditional Partial Effects: China 1988 and 1995

			1988					**1995**		
Var	gini	variance	iq90_10	iq90_50	iq50_10	gini	variance	iq90_10	iq90_50	iq50_10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Primary	0.00384***	0.0269***	0.0248***	0.00947***	0.0135***	0.00920***	0.147***	0.0440***	0.0189***	0.0209*
	(0.000803)	(0.00982)	(0.00476)	(0.00274)	(0.00359)	(0.00228)	(0.0453)	(0.0130)	(0.00479)	(0.0108)
Higher	0.00232***	0.0266***	0.0132***	0.00359	0.00863***	-0.00110	-0.0146	0.00172	0.00467	-0.00324
	(0.000713)	(0.00879)	(0.00433)	(0.00315)	(0.00294)	(0.000932)	(0.0156)	(0.00617)	(0.00365)	(0.00461)
Observations	18,337	18,337	18,337	18,337	18,337	12,084	12,084	12,084	12,084	12,084
R-squared	0.107	0.070	0.095	0.055	0.062	0.094	0.084	0.070	0.065	0.043
Mean Dep Var	.036	.197	1.17	1.07	1.09	.047	.390	1.22	1.09	1.12
Fixed Effects: Industry, Occupation, Province; Standard Error: Robust; Controls: yes										

Notes: The table presents the results from RIF regression (Equation 5) for the years 1988 and 1995 China labor force surveys of the salaried class population between 20-60 years old and having a positive income. Col (1)-(5) is for 1988 and Col (6)-(10) for 1995. All the regressions use industry, occupation and province fixed effects. Controls include gender, rural, age and age square. Survey weights are used. In the 1980s and 90s the coefficients are too small, implying education was not a prominent factor in determining wage inequality.

Table 1.12. Unconditional Partial Effects: China 2002 and 2013

	2002					2013				
Var	gini (1)	variance (2)	iq90_10 (3)	iq90_50 (4)	iq50_10 (5)	gini (6)	variance (7)	iq90_10 (8)	iq90_50 (9)	iq50_10 (10)
Primary	0.0181***	0.443***	0.151***	0.0174***	0.114***	0.0119***	0.253***	0.0749***	0.0127***	0.0560***
	(0.00320)	(0.107)	(0.0288)	(0.00444)	(0.0249)	(0.00155)	(0.0406)	(0.0113)	(0.00239)	(0.0100)
Higher	0.0131***	0.414***	0.102***	0.0225***	0.0649***	0.00842***	0.215***	0.0521***	0.0237***	0.0231***
	(0.00130)	(0.0346)	(0.0110)	(0.00627)	(0.00785)	(0.00110)	(0.0251)	(0.00766)	(0.00316)	(0.00625)
Observations	15,464	15,464	15,464	15,464	15,464	21,049	21,049	21,049	21,049	21,049
R-squared	0.219	0.122	0.155	0.129	.096	0.057	0.042	0.044	0.029	0.037
Mean Dep Var	.08	1.33	1.46	1.13	1.28	.051	.636	1.25	1.09	1.15
Fixed Effects: Industry, Occupation, Province; Standard Error: Robust; Controls: yes										

Notes: The table presents the results from RIF regression (Equation 5) for the years 2002 and 2013 China labor force surveys of the salaried class population between 20-60 years old and having a positive income. Col (1)-(5) is for 2002 and Col (6)-(10) for 2013. All the regressions use industry, occupation and province fixed effects. Controls include gender, rural, age and age square. Survey weights are used.

Table 1.13. Unconditional Partial Effects: India 1983 and 1987

	1983					1987				
Var	gini (1)	variance (2)	iq90_10 (3)	iq90_50 (4)	iq50_10 (5)	gini (6)	variance (7)	iq90_10 (8)	iq90_50 (9)	iq50_10 (10)
Primary	0.0149***	0.183***	0.0572***	0.0487***	-0.00203	0.0158***	0.222***	0.0801***	0.0494***	0.0197***
	(0.000911)	(0.0195)	(0.00573)	(0.00278)	(0.00512)	(0.000887)	(0.0183)	(0.00633)	(0.00261)	(0.00560)
Tertiary	0.0166***	0.343***	0.124***	0.0619***	0.0440***	0.00887***	0.221***	0.0655***	0.0313***	0.0259***
	(0.00112)	(0.0235)	(0.00680)	(0.00477)	(0.00411)	(0.00102)	(0.0213)	(0.00716)	(0.00437)	(0.00479)
Observations	34,716	34,716	34,716	34,716	34,716	26,405	26,405	26,405	26,405	26,405
R-squared	0.164	0.086	0.100	0.094	0.068	0.131	0.087	0.103	0.073	0.074
Mean Dep Var	.064	.656	1.32	1.10	1.20	.057	.585	1.30	1.11	1.18
Fixed Effects: Industry, Occupation, Province; Standard Error: Robust; Controls: yes										

Notes: The table presents the results from RIF regression (Equation 5) for the years 1983 and 1987 India labor force surveys of the salaried class population between 20-60 years old and having a positive income. Col (1)-(5) is for 1983 and Col (6)-(10) for 1987. All the regressions use industry, occupation and province fixed effects. Controls include gender, rural, age and age square. Survey weights are used.

TABLE 1.14. Unconditional Partial Effects: India 1993 and 1999

	1993					**1999**				
Var	gini (1)	variance (2)	iq90_10 (3)	iq90_50 (4)	iq50_10 (5)	gini (6)	variance (7)	iq90_10 (8)	iq90_50 (9)	iq50_10 (10)
Primary	0.0182***	0.296***	0.0936***	0.0603***	0.0188***	0.0144***	0.229***	0.0772***	0.0659***	-0.00142
	(0.00133)	(0.0435)	(0.00702)	(0.00271)	(0.00623)	(0.00112)	(0.0248)	(0.00728)	(0.00402)	(0.00628)
Tertiary	0.00990***	0.280***	0.0684***	0.0281***	0.0312***	0.00529***	0.183***	0.0415***	0.0139***	0.0221***
	(0.00129)	(0.0375)	(0.00713)	(0.00397)	(0.00517)	(0.00136)	(0.0345)	(0.00720)	(0.00535)	(0.00540)
Observations	35,270	35,270	35,270	35,270	35,270	37,377	37,377	37,377	37,377	37,377
R-squared	0.073	0.026	0.106	0.101	0.084	0.139	0.110	0.113	0.138	0.124
MeanDepVar	.068	.88	1.34	1.11	1.21	.066	.770	1.34	1.12	1.20
Fixed Effects: Industry, Occupation, Province; Standard Error: Robust; Controls: yes										

Notes: The table presents the results from RIF regression (Equation 5) for the years 1993 and 1999 India labor force surveys of the salaried class population between 20-60 years old and having a positive income. Col (1)-(5) is for 1993 and Col (6)-(10) for 1999. All the regressions use industry, occupation and province fixed effects. Controls include gender, rural, age and age square. Survey weights are used.

TABLE 1.15. Unconditional Partial Effects: India 2004 and 2011

	2004					**2011**				
Var	gini (1)	variance (2)	iq90_10 (3)	iq90_50 (4)	iq50_10 (5)	gini (6)	variance (7)	iq90_10 (8)	iq90_50 (9)	iq50_10 (10)
Primary	0.0133***	0.249***	0.0732***	0.0746***	-0.0122*	0.0106***	0.211***	0.0471***	0.0586***	-0.0167***
	(0.000910)	(0.0204)	(0.00672)	(0.00457)	(0.00638)	(0.00109)	(0.0245)	(0.00593)	(0.00470)	(0.00614)
Tertiary	0.00477***	0.208***	0.0317***	-0.0260***	0.0531***	0.00534***	0.203***	0.0390***	-0.0351***	0.0675***
	(0.000896)	(0.0213)	(0.00695)	(0.00545)	(0.00546)	(0.000938)	(0.0222)	(0.00530)	(0.00560)	(0.00491)
Observations	39,602	39,602	39,602	39,602	39,602	37,488	37,488	37,488	37,488	37,488
R-squared	0.151	0.145	0.127	0.130	0.144	0.133	0.134	0.111	0.086	0.157
MeanDepVar	.070	.844	1.37	1.17	1.17	.066	.810	1.33	1.17	1.14
Fixed Effects: Industry, Occupation, Province; Standard Error: Robust; Controls: yes										

Notes: The table presents the results from RIF regression (Equation 5) for the years 2004 and 2011 India labor force surveys of the salaried class population between 20-60 years old and having a positive income. Col (1)-(5) is for 2004 and Col (6)-(10) for 2011. All the regressions use industry, occupation and province fixed effects. Controls include gender, rural, age and age square. Survey weights are used.

TABLE 1.16. Unconditional Partial Effects: China and India 2018

	China 2018					India 2018				
Var	gini	variance	iq90_10	iq90_50	iq50_10	gini	variance	iq90_10	iq90_50	iq50_10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Primary	0.0156***	0.375***	0.142***	0.0182***	0.110***	0.0102***	0.180***	0.0519***	0.0297***	0.0163***
	(0.00183)	(0.0567)	(0.0176)	(0.00244)	(0.0156)	(0.000989)	(0.0212)	(0.00614)	(0.00336)	(0.00539)
Higher	0.00506***	0.186***	0.0297***	0.0145***	0.0120	0.00595***	0.159***	0.0442***	0.00884**	0.0298***
	(0.000950)	(0.0259)	(0.00878)	(0.00272)	(0.00738)	(0.000733)	(0.0158)	(0.00508)	(0.00390)	(0.00368)
Observations	20,189	20,189	20,189	20,189	20,189	40,652	40,652	40,652	40,652	40,652
R-squared	0.058	0.041	0.055	0.027	0.051	0.184	0.149	0.156	0.062	0.095
Mean Dep Var	.053	.743	1.26	1.10	1.14	.058	.624	1.29	1.16	1.11
Fixed Effects: Industry, Occupation, Province; Standard Error: Robust; Controls: yes										

Notes: The table presents the results from RIF regression (Equation 5) for 2018 using the labor force surveys of the salaried class population between 20-60 years old and having a positive income. Col (1)-(5) is for China and Col (6)-(10) for India. All the regressions use industry, occupation and province fixed effects. Controls include gender, rural, age and age square. Survey weights are used.

TABLE 1.17. Ratio of Exp/Student in Nominal $

	Average value of Exp/Student in Nominal $						**Ratio of Exp/Student**		
	Primary		Middle		Higher		Primary	Middle	Higher
	CH	IN	CH	IN	CH	IN	CH/IN	CH/IN	CH/IN
1900-30	4	2	51	24	252	194	2.05	2.10	1.30
1930-50	2	4	23	20	186	117	0.45	1.19	1.60
1950-85	9	10	27	31	805	116	0.93	0.87	6.95
1986-10	230	54	413	135	1333	608	4.26	3.07	2.19
2010-20	1472	237	2013	350	3453	866	6.21	5.76	3.99

Notes: The table presents the average expenditure per student (nominal $) by period and ratio between China and India. At the primary stage, during the first 30 years of the century, exp/student was double in China than India, which reduced to half in the next 20 years. Source: Authors' calculations.

C. Appendix Notes

C.1. Other Results.

C.1.1. *Gender Gap in Education.* Gender disparity and discrimination is a well-known issue in both the countries. The Global Gender Gap 2020 report puts China and India at 106th and 112th position out of 153 countries [120]. Both countries are closely ranked in educational attainment and health gap sub-indices but differs along economic participation and political empowerment [121]. This section focuses upon the gender gaps in the educational outcomes by levels of education since 1900.

We describe gender gap in enrollment and teachers at all the three levels in detail here.

The upper portion of the Appendix Figures 1.20, 1.21 and 1.22 provide the evolution of share of female in total enrollment at primary, middle and tertiary stage of education respectively. Refer lower part of Appendix Figures 1.20, 1.21, 1.22) for female teachers share.

Primary stage: The first 50 years of the 20th century had huge gender gap (20-40 percentage point under-representation compared to overall population), though with considerable catching-up. In 1887, the female share at primary stage was a meagre 8.5% in India, which increased gradually to 28% by 1950. China also had a similar evolution - from 15% in 1931 to 28% in 1950. The evolution in two countries changes after 1950, due to extraordinarily rapid expansion of mass primary education in China reducing the gender gap faster. By 1985, female share was 45% in China and 40% in India. The final push leading to no gender gap came only in 21st century in India with several targeted measures under-taken to bring all the out-of-school kids to the school. Now, both countries have closed the gender gap at primary stage enrollment [122]. It took China-100 years and India-150 years to achieve this feet.

[120] The Index takes into account 4 dimensions economic participation and opportunity, educational attainment, health and political empowerment. Schwab et al., 2019

[121] A look by sub-indices of the index shows that for Educational attainment (CH-100th and IN-112th) and Health (CH-153rd and IN-150th) both countries are close. The major differences are seen in Economic participation (India is worse at 149th compared to China 91st) and Political empowerment (India ranks 18th compared to 95th position of China)

[122] Datta and Gandhi Kingdon, 2021 using NSS 71st 2014 round data shows no statistically significant for India for Age 5-9 years

Middle stage : Female share at middle stage [123] depicts similar pattern as before with one marked difference - higher level of female enrollment in China than India since 1925. Female share was stable at 3-5% in India during 1887-1930, even though the share of female enrollment at primary stage improved during this period - due to higher level of female drop-outs.[124] China on the other hand, saw female share in enrollment shooting up from 6% in 1925 to 15% in 1930. In a very short time period of 5 years - China and India diverged by 10 percentage points which remained for the next 50+ years. The next data point coming from liberated China puts the female share at 27% whereas it was 17% in independent India. By 1985 - China had 41% female enrolled in middle stage compared to 34% in India. The share of female at middle stage standard school is now close to the female share in the population in both the countries.

Higher Stage : The female share in higher education is always higher in China than India. But both countries had gender gap till last decade. In 1950, out of total enrollment at higher stage females were 20% in China and 10% in India. The difference continued for next 25 years, when at the end of cultural revolution there is a huge influx of male going back to school driving down the female share by from 33% in 1976 to 29% in 1977 and to 24% in 1978. India maintained it's gradual progress of narrowing down the gender gap till 2010. There is a big increase in female share after 2012, where it increases from 42% to 49% in 2018 closing the gender gap. In China, there is now over-representation of female in higher education at 52%.

C.1.2. *Students/Graduate.* Students per Graduate by the level of education gives an idea about the enrolled students finishing their studies (or dropouts). We find a higher dropout rate peculiar to the Indian education system. Only 1 out of 50-60 enrolled students finished their primary education in 1900-the 1920s. Over the years, there has been improvement in the situation at both the primary and middle levels. The change in the policy of free pass-through of the students up to Class VIII[125] can be seen in the declining ratio in India. China has a consistently better completion rate (except during the tumultuous revolution period of 1945-50) at all levels of education. (Refer Appendix Figure 1.24)

C.1.3. *Expenditure/Student.* This measure captures input resources per student. It is a simple qualitative measure but very data demanding for cross-country comparisons

[123]Female enrollment in standard school, as gender split for vocational education is not present.
[124]Class-wise enrollment figure reveals that more girls were dropping out than boys in primary stage.
[125]Right to Education Act 2008 in India allows kids to go to the next standard without an exam.

over a long period. It requires- expenditure in the same currency (exchange rate) at the constant price (price index) and a similar level of economy (GNI).

The table 1.17 provides the evolution of average values of exp/student in nominal dollars for both countries and the ratio between China and India. The ratio allows temporal comparison. Pre-1950, the most important thing to note is that from 1900-30 to 1930-50, the ratio for primary level gets reversed. During the first 30 years, exp/student was double in China than India, which reduces to half in the next 20 years. The other thing to note is that India's average exp/student doubled at the primary level. In contrast, it was reduced at the middle and higher level (in line with the adopted policy of slower expansion while maintaining quality).

We could convert all prices at 2018 ($) after 1950 but the divergence in GNI pc between China-India after 1980's, restricts the usefulness of this statistic. For the sake of completion we provide the comparison from 1950-2018 (Refer Appendix Figure 1.25, 1.25 and 1.26 at 2018 $). Between 1950-80, exp/student is higher in India at the primary stage. In real 2018 $ terms - the cost at the primary stage was around $25-30 at the primary stage in India compared to $17 in China in the 1950s. The exp/student increased by 30% in India by the 1980s but remained at the same level in China. In the middle stage, in general education, the expenditure/student is higher in China, except during the 1940s revolution and the 1960s cultural revolution. In recent years, the exp/student in China has been 5-6 times more than in India at both primary and middle stages. At the higher stage, after 1950, the exp/student has remained 1.5-3 times higher in China for all the years. It is not surprising as higher education in China has developed more vocational and professional courses (like engineering, teachers' training etc.), which are more expensive forms of education than social sciences. In real 2018 $ terms - the total expenditure at the Tertiary stage is $3500-4000 per student in China compared to $1000 in India in the late 2010s.

C.2. Data in Detail.

C.2.1. *China.* At the primary stage, we exclude adult education for all statistics other than expenditure. Thus, the enrollment is only the standard student enrollment. However, there are overaged kids at school; the ratio is around 20% from 1949-1981. Later the share became too small. The net enrollment rate data at the primary stage from 1951-2018 are reported values from the education yearbooks. In the education budget data containing revenue and expenditure at the primary stage, adult primary numbers are included, too, though it accounts for a minimal share $< 0.1\%$. All other measures like enrollment, net enrollment rate etc. do not include adults.

C.2.2. *India.* **Pre-indepence period** (i.e pre 1947), colonial government used to produce quinquinneal (i.e in every 5 years) reports titled "Progress of Education in India", which is primarily used here. The reports provide information on primary, secondary and higher education. The reports starts from 1887-88 and there are total 12 reports. There are two Volumes with Volume 2 containing statistical tables. The statistical tables are very extensive containing information on the aspects of enrollment, graduates, teachers and expenditures for all level of education.

Post-independence period, Indian government continued for some years the report on the same structure, however it starts becoming more and more complex. Further the division of administrative structure over the level of education made it difficult to assemble the information. There are frequent changes in the structure of the reports over the years. I enlist the important documents we use :
1) Education in India reports : Education in India reports are published annually by MHRD (Ministry of Human Resource and Development) [126]. It is a good first hand source of all-India collected data. Till 1986-87, the reports included all levels of education after which the responsibility of collecting information on affiliated universities and colleges was transferred to UGC (University Grants Commission).

Primary and Secondary education information comes from "Statistics of School Education". Higher education information comes from "Statistics of Higher and Technical information.

UGC reports have also been utilised to get detailed information for higher education.

Results of high school and higher secondary examination : is used to get the total graduates for secondary level of education. These reports provide important statistics of examination results of High School, Higher Secondary and Intermediate/Pre-University

[126] Ministry of Education and Social Welfare before. Precisely this is brought out by the Statistics and Information Division in the Department of Education

examinations conducted by various Boards of Secondary, Higher Secondary and Pre-University Education in the country.

C.3. Comparison with Other Datasets. Historical (1900-1970):
Mitchell : We compare the enrollment figures from 1900-1970 for Primary/Middle/Higher education with Mitchell, 1998. The different is less than 0.5% for China for entire duration. The difference between Mitchell, 1998 and Indian data is as expected since we have emphasized on carefully allocating students to their respective stage of education. Our numbers is higher at Primary level by 5-8% for different years, as the students at Primary stage but studying in secondary schools are allocated at Primary. On the other hand, our higher education numbers are lower because we take out the Intermediate level (Class XI-XII) students from Higher education and put them at Middle level.

UNESCO World Education Surveys: UNESCO, 1958 provides Primary level enrollment from 1930-58. The difference is close to zero for China. For India, our numbers are 11% higher in 1930 and decreases to 1-2% after 1950. UNESCO, 1961b and UNESCO, 1961a provides Secondary and Higher level enrollment. Since UNESCO method is also to allocate Intermediate students into Secondary, the higher level enrollment figures for India are very close.

Contemporary 1970:
UNESCO : UNESCO provides information on the variables from 1970 onwards. We compare our figures with UNESCO and highlight the contribution of our data.
First UNESCO does not provide information on following:

(1) enrollment by stage: enrollment figures are provided consistently post 1970. The Primary level enrollment figures differs by +/- 3% in comparison with UNESCO data, with more difference in the recent years for India. This is possibly on the account of frequent updates of past years by the Government of India on the estimated numbers.
(2) Discipline wise share: It is completely missing for China and for India the information is present only from 2013. We provide the discipline wise share of enrollment and graduates from very early 1900's.
(3) Expenditure split by Education: Once again the information is missing for China and for India, sparse data is present from 1999.

(4) Share of Private enrollment by stages: The information in UNESCO is present from 2000 for India and post 2005 for China for Primary level and only after 2013 for secondary and tertiary level education.
(5) Govt Exp as % of GDP: UNESCO provides the information for India from 1997-2013 and for China from 1971-1999.

C.4. Variable Creation. The important variables we create in this paper are defined below in detail.

(1) Total enrollment: is the total students enrolled (on roll) at a given date (31st March) of year in different stages of education. [127] It includes the non-attending students too (cite some paper or number highlighting difference). In India enrollment at middle stage include the intermediate students.

(2) Total Graduates : is the total students finishing a certain level of education in a given year. In Middle and Higher stage of education- vocational and non-vocational split is also provided (applies for enrollment too).
(3) Total Teachers : is the total teachers at a certain level of education. For India the numbers are imputed as for expenditure.
(4) Total Expenditure: is the total expenditure (public + private) at a certain level of education.
(5) enrollment/Graduates: is total students enrolled per student completing the level of education. This provides some sense of dropouts, but it is not perfect as increasing (due to expansion) or decreasing (due to contraction in population in certain age-cohort) trend in enrollment can lead to mis-interpretation.
(6) Students/Teacher: is total enrolled students per teacher at a certain level of education. It is one of the qualitative measures of education.
(7) Expenditure/Students : is total expenditure per student at a certain level of education.
(8) Gross/Net enrollment Ratio : is the usual definition, where Gross is total kids at a certain level of education over the total population of the kids in that respective

[127] India- Both recognised (course is prescribed/recognised by the Government/Board constituted by the law; open to inspection and eligible for admission to public examinations and tests held by Government) and unrecognised institutions are covered.

age group. Net enrollment ratio uses total kids of the respective age group in the numerator.

(9) Gender Ratio : We compute two measures to study gender differences (bias) in education system. The first measure is % Female enrollment which is total female enrollment divided by total enrollment by different stage of education. The second measure is % Female Teachers which is total female teachers divided by total teachers (stage-wise).

C.4.1. *Expenditure: India.* : There are mainly three types of sources which are utilised

(1) *Expenditure from Educational Statistics Report (upto 2000)*: is the first and the most important source. It provides income and expenditure receipts by type of institutions[128]. The income receipts are split by source type: Government Funds, Universities and Local Body Funds (all 3 forming the Public component); Fees, Endowment and Other sources (forming Private component)[129].
The expenditure is split by type of institutions (i.e Primary, Middle, Secondary) and not by the stage of education. A Secondary school in India usually also has primary (Grade I-V) and Upper Primary/Junior Low (Grade VI-VIII) classes. Similarly till 1960's, intermediate (IX-XII) were part of collegiate education and expenditure is reported under higher education.The computation of stage-wise expenditure is as follows:

   (a) Total primary stage expenditure = (Expenditure/kid in primary schools)*(Primary stage enrollment) i.e we use the expenditure per kid in primary schools (total cost in Primary schools/total enrollment in primary schools) provided in the reports and multiply with the total enrollment at primary stage to arrive at total expenditure at primary stage.
   (b) Total middle stage expenditure = Total Exp in secondary/higher secondary - (Total Primary stage Expenditure - Total Cost in Primary School) + Total Intermediate Stage Exp + Total Vocational/Professional Exp.

[128]It also splits into Recurring and Non-Recurring. Recurring expenditure is incurred every year by an educational institution e.g expenditure on salaries, Maintenance, scholarships, Direction/Inspection etc. No-recurring is other than recurring which mainly includes construction of buildings, equipment, libraries etc.

[129]It covers only recognized institutions. The surveys from post 1996 also capture unrecognized schools which have become important due to huge expansion of the unrecognized schools.

These reports stopped providing expenditure for higher education from 1986-87 and stopped completely after 1999-2000. Hence expenditure calculations after 1986-87 involves the use of Analysis of Budget Expenditure reports (annual; capturing public expenditure exponent) and NSS Education Surveys (1986, 1995, 2007, 2014 and 2018; capturing private expenditure).

(2) *Analysis of Budget Expenditure Reports 1951-2018*: are annual publications, which is compiled from the Demands for Grants made by Central and States governments[130]. There are three expenditure estimates - Budget(BE), Revised(RE) and Actual(AE). [131] We use Actual Estimates as they are the final estimates, and have gone through multiple rounds of vetting. The expenditure is split under Revenue and Non-Revenue(Capital and Loans & Advances Account). Non-Revenue portion is ~1-2% of the total expenditure, which goes into capital works. The expense is incurred not only by the Education Departments but also from Other Departments[132]. The share of Other Departments has increased a lot in recent years. It went up from 4% in 1950's to 7% in 1960's, to 13% in 1970's, remained below 20% upto 2012, but then after has increased consistently to 32% in 2015, 2016 and 2017. One of the limitation of these reports is that upto 2003, it was double counting the centrally sponsored schemes as it is entered both under Centre and State.

The stage wise analysis requires one extra step since the categories provided doesn't perfectly match with our stage definition. The categories provided in these reports are Elementary (Grade I-VIII), Secondary (Grade IX-XII), University & Higher Education, Adult Education, Technical Education and Others. We split the Elementary(Grade I-VIII) expenditure into two parts- Primary (I-V) and Upper Primary (VI-VIII). Upper Primary is included into Secondary to get complete Middle stage(Grade VI-XII) public expenditure. HOW DID WE DO IT! (WRITE)

[130]It provides Plan and Non-Plan Expenditure for various sub-sectors of Education

[131]Actual Estimate is the final expenditure coming with a delay of few years. The last Actual Estimate available is for the year 2015-16. Revised estimate is the pre-final estimate and last available for 2016-17. Budget estimate is the budgeted estimate, last present for 2017-18.

[132]Department of Arts, Culture, Agriculture, Health etc. also make provision towards education sector.

(3) *NSS Education Surveys* NSS started conducting "Participation in Education", all-India representative surveys to capture the expenditure details for currently enrolled students. These surveys are present for the years 1986, 1995, 2007, 2014 and 2018; the intermittent years are extrapolated. It captures broad range of expenses like tuition, examination, other fees , stationery, uniform, transport private coaching etc. The first three (i.e. only fees) are used to compute private expenditure, to make it comparable with previous years. The current level of enrollment is used to compute stage-wise average expenditure.

We compare the consistency between Total public expenditure component from Educational statistical reports and Budget Expenditure reports. Comparing Public component of expenditure from Education Statistical reports and Budget Expenditure Reports : The values from both the data sources are very close upto 1968, after which the discrepancy starts. It is worth noting that the budgetary data from 1951-52 to 1967-68 comes actually comes from "Combined Finance and Revenue Account" which was published by Comptroller and Auditor General (C&AG) of India. Possibly C&AG reports has tried solving this discrepancy. From 1968-69 onwards Ministry of Human Resources and Development (MHRD) started publishing annual reports. The values from Education statistical reports are usually ∼0.7-0.9 times of values from Budget data. Over the years, the discrepancy has increased. One of the possible reasons could be that in educational reports educational institutions under-report to get gain more government aid[133].

C.4.2. *Teachers: India.* : The main challenge is to get teachers by stage of education as the reports provide total teachers by school type (i.e teachers in primary, secondary schools etc.) and not by stage of education.

(1) Total primary stage Teachers = (Teachers/student in primary schools)*(Primary stage enrollment) i.e we use the teacher per student in primary schools (Teachers in Primary schools/Total enrollment in primary schools) multiplied with the total enrollment at primary stage to estimate total teachers at primary stage.
(2) Total Middle stage Teachers = Total teachers in secondary/higher secondary - (Total Primary stage Teachers - Total Teachers in Primary School) + Total Intermediate Stage Teachers + Total Vocational/Professional Teachers. After 1950,

[133]This is not completely implausible, as NSS 42nd report while comparing total enrollment figure with the educational statistical reports found educational statistics enrollments figure higher than surveys, and it provides exactly the same reason.

the reports started providing total teachers in Upper Primary schools (Grade VI-VIII) and Secondary/Sr Secondary schools (Grade IX-XIII). Further post 1990, teachers in Secondary (Grade IX-X) and Senior Secondary(XI-XII) school are present. Correspondingly we also estimate teachers at Upper Primary, Secondary and Senior Secondary level/stage.

C.5. Measuring Education Inequality. We compute all the measures at cohort-level following (Thomas, Wang, and Fan, 2001). We divide the population into 8 categories (Illiterate, <6 years, Primary, Lower Middle, Middle, Vocational, Bachelors, Masters and PhD) in China and 9 categories (Illiterate, Primary, Secondary, Senior Secondary, Vocational, Bachelors, Masters and PhD) in India. The reason to use different categories for China and India is the use of different standard exams to finish a certain level of education. For e.g exam after Junior Middle (or 9 years of schooling) is conducted is important exam in China whereas in India the standardized exam is conducted after 10 years of schooling called Matriculation exam. The categories are mutually exclusive and collectively inclusive.

C.5.1. *Average Years of Schooling.*

$$\mu = AYS = \sum_{i=1}^{n} p_i y_i$$

Here n is the number of level/categories. p_i is the probability of finishing a certain level of education which is computed simply as empirical ratio of number of graduates over total population. For the probability of finishing primary education (say 5 years of education) for a cohort born in 1960 is ratio of total students finishing primary stage divided by total population of Age 1 in 1960. Similarly probability of finishing middle stage (say 12 years of education) for the same cohort would be ratio of total students finishing 12 years of education divided by total population of Age 1 and so on. y_i is the years of schooling which is 0 for the population with no schooling.

C.5.2. *Education Gini.* The following formula provides an easy way to compute education overcoming the limitations in computing the traditional gini[134].

$$EducationGini = (1/\mu) \sum_{i=2}^{n} \sum_{j=1}^{i-1} p_i |y_i - y_j| p_j$$

[134]The limitations being discrete nature of the educational attainment with both lower(education-0; illiterate population) and upper boundary. In both China and India a big chunk of the population is illiterate.

p_i, p_j, y_i, y_j and n are the same as described above.

C.5.3. *Education Standard Deviation.* Education Gini computes the relative measure of inequality. The absolute measure of education dispersion is computed through the following formula of standard deviation of schooling (SDS):

$$\sigma = SDS = (1/\mu)\sqrt{\sum_{i=1}^{n} p_i(y_i - \mu)^2}$$

p_i, y_i and n are the same as described above.

CHAPTER 2

Wealth Inequality, Class and Caste in India, 1961-2018

[1] Abstract

This paper makes two main contributions. First, I combine data from wealth surveys (NSS-AIDIS) and millionaire lists to produce wealth inequality series for India over the 1961-2018 period. I find a strong rise in wealth concentration in recent decades, in line with recent research using income. For E.g. the top 10% wealth share rose from 45% in 1981 to 61% in 2018, while the top 1% share rose from 27% to 44%. Second, I gather information from censuses and household surveys (NSS- AIDIS and consumption, IHDS, NFHS) to explore the changing relationship between class and caste in India and the mechanisms behind rising inequality. Assortative matching is very high in India, both at the caste and education level (though not larger than in Western countries at the education level). I stress the limits of our knowledge and indicate possible lines of future research, particularly regarding the interplay between assortative matching and inequality dynamics.

[1] I would like to thank Abhijit Banerjee, Guilhem Cassan, Oliver Vanden Eynde, Catherine Guirkinger and Thomas Piketty for their comments and support. I am very thankful to Pulak Ghosh for providing the SECC dataset. I gratefully acknowledge the comments received during presentations at the Paris School of Economics, University of Namur, ISI Winter Conference (Delhi) and China-India Workshop (Warwick).

1. Introduction

The distribution of economic resources within society is at the core of economic science. Economic or wealth inequality, or the skewed distribution of these resources, is now possible to investigate empirically, given the increasing data availability over the past decades. There is relatively little work on wealth inequality compared to income and consumption inequality.[2] While such flow measures, although important, do not capture the full spectrum of resources in society. And distribution of resources matter: wealth inequality has been associated with tax evasion, political equilibrium and voting, and education provision- hence understanding the evolution of wealth inequality becomes essential. The general trend of increasing inequality worldwide further motivates country-specific analysis, as it may differ from country to country.

This paper estimates wealth inequality in India from 1961 to 2018 by combining household survey data and a millionaire's rich list. Then I study the dynamics of caste and inequality in detail, as I expect the hierarchical order imposed by castes to matter for the distribution of resources. Third, using census data, I examine village-level (agricultural) land inequality in the ten largest states of India.

The wealth concentration in India has increased to reach an extreme level, in line with the rising income inequality (Chancel and Piketty, 2017, A. Banerjee, 2005). The top 10% of Indian households in terms of wealth own 61% of the total household wealth in 2018, compared to 43% in 1961. [3] The wealth concentration within the top 10% population is very high: the top 5% own 50% of the total household wealth in 2018, compared to 31% in 1981, while the top 1% own 33% of total wealth in 2018, compared to 13% in 1981. Such wealth concentration is higher than in China but lower than in South Africa.[4]

The wealth basket of Indian households is dominated by physical assets, forming 90% of the wealth value. The land is the most valuable asset contributing 55-60% of

[2] E.g. World Inequality Report 2022, shows the presence of income data across the globe for the 20th century, but wealth inequality figures are available almost exclusively for the US and European countries.
[3] The top 10% population owned 55% of the total income in 2012.). The top 10% share from the consumption data has increased from 28% in 1983 to 32% in 2012.
[4] The wealth share of the top 1% in China is 29.8% and 49.9% in South Africa.

the total wealth, even in the top 10% wealthy households.[5] A closer look into the dynamics of land area and land price show that the rich class (top 10%) is moving away from agricultural land ownership and acquiring more expensive non-agricultural land (residential and others). The land premium for the rich class (compared to the rest of the population) for a given land type has also increased between 1991-2018.[6] E.g. the ratio of the land price of residential land owned by the top10% and bottom 50% was 8.8 in 1991 and 11.8 in 2018.

It almost becomes imperative to consider the caste, given land, which forms the most valuable asset, has been closely linked with the caste in the past. The gap between the higher and lower caste groups (Scheduled Caste, Scheduled Tribe and Other Backward Caste) for wealth ownership is higher than other economic indicators like income or consumption. [7] E.g. In 2011-12, the average income (consumption) of FC was 1.9 (1.7) times the average income of SC. For wealth, the ratio was 4.3.[8] There is a high level of ordinal polarization in society. [9] The FCs are clustered in the top decile (and middle 40%) while the SCs and STs are more present in the bottom 50%. And there is no sign of convergence between different caste groups during 1991-2018, the period of rapid economic development in India post-liberalization. However, within-caste inequality is also increasing, in line with the increasing within-caste consumption inequality studied in Borooah, 2005. The within-caste disparity is higher in the ST, Muslim and FC group (the rich class within each caste own more than 50% of the total wealth of the caste), followed

[5]One should see it with some caution because surveys do not capture well the wealthiest individuals.

[6]It could be due to the development of better infrastructure (like roads, sewage etc., in urban areas or irrigation facilities in rural areas) around the land of the rich class. The data does not allow us to pin down the reasons behind this price differential.

[7]The caste categorization detail is in the data section. Briefly, the three categories having reservation benefits - Scheduled Caste (SC), Scheduled Tribe (ST) and Other Backward Castes (OBC) - are referred to here as the lower-caste groups. Forward Caste (FC) group is higher castes comprising the rest of the population, excluding Muslims. Muslim category denotes Muslims who are non-SC, non-ST, and non-OBC

[8]Lower caste groups own lesser wealth than their population share, while the FC group has more wealth than its population share. The difference between SC-owned wealth share and its population share is -9 percentage points(pp), whereas the difference is +15pp for the FC group.

[9]The concept of ordinal polarization as developed in Jayadev and Reddy, 2011 refers to the presence of representational and sequence inequality. Representational inequality measures a degree of "segregation" between different social groups and any attribute space (here, wealth). Sequence Inequality measures the degree of "clustering" based on hierarchy among social groups.

by the OBC and SC.

The rising inequality could be due to increasing marriages between wealthy individuals (families). Though there is no data on the individual level of wealth, I use education and wages (a proxy for wealth) to explore this channel. I find an increasing level of education assortativity in the formation of couples. The correlation between husband and wife's education level was low (at 0.35) in the pre-1980 married couples, which has jumped to 0.6 among the 2006-12 married couples. Also, there are differences across caste groups. The education and economic assortativity are also the highest in the groups where within-caste wealth inequality is high (FC and ST). The correlation between husband and wife's wages is 0.75 for FC and ST, 0.6 for OBCs and 0.55 for SCs. It suggests that a high level of assortative matching could be the driver of increasing within-caste inequality.

Finally, I show that SC (Dalits) population share is a predictor of the agricultural land area inequality, controlling for institutional, demographic, geographic and climatic factors. A cent per cent rise in SC population share is associated to an increase in the the land inequality (gini) by 0.2, which is 16% of the mean (0.71). Even if not causal, the correlation is robust to different specifications or inequality measures, showing a strong pattern in line with the historical distribution of land.

The paper contributes to the growing wealth inequality literature in India (Jayadev, Motiram, and Vakulabharanam, 2007; Anand and Thampi, 2016; Zacharias and Vakulabharanam, 2011; Jayaraj and Subramanian, 2018; Anand and Kumar, 2022) and in other developing countries (South Africa- Chatterjee, Czajka, and Gethin, 2022; China -Piketty, Yang, and Zucman, 2019) It produces a long term wealth inequality series for India, combining two approaches - using household surveys of personal wealth and combining surveys with the list of wealthy individuals produced by different magazines like Forbes millionaire's list. [10] This paper uses Forbes list to correct the top shares from

[10] Other approaches, though it is not feasible in the Indian context. There are five ways to estimate the distribution of wealth (Alvaredo, Atkinson, and Morelli, 2016). Other than the two used in this paper, Kumar, 2016 uses inheritance tax returns with mortality tables data with an estate multiplier technique to produce the top .01% wealth share during 1961-85. Inheritance tax was active between 1953-85. It shows the limitation of this data in studying recent wealth inequality. The fourth method requires annual wealth data. Wealth tax prevailed in India between 1957-2016, though it exempted productive assets like shares,

the survey using Generalized Pareto Interpolation technique (Blanchet, Fournier, and Piketty, 2022).[11]

I also explore the puzzling decline in wealth inequality from 2018 survey compared to the previous surveys. Anand and Kumar, 2022 highlights the under-representation of the urban population in the survey, which could be the potential reason behind the decline. This paper corrects for another change in the adopted sampling strategy - the 2018 survey used consumption expenditure and indebtedness both to prepare second-stage strata. In contrast, the 2012 survey used only indebtedness. The correction (re-sampling and assigning new weights in a way to mimic the 2012 methodology) leads to only a marginal change (+0.4%) in the wealth share of the top 10% compared to the original survey, suggesting that it is not the reason behind the decline in the survey-based wealth inequality measures.

This paper adds to the existing literature on inequality studies making caste a central social stratification of the society. Several papers, such as Thorat, 2002; Borooah, 2005; Zacharias and Vakulabharanam, 2011, have highlighted the economic differences among castes which exactly matches the caste hierarchy present in the society. This paper extends the caste-economic analysis for wealth. The caste categories in the existing literature remain limited to SC, ST and Others. For the analysis, I used more caste categories, like Other Backward Castes. I also disentangle the broad higher-caste group into finer categories- Brahmin, Rajput, Bania and Kayasth and present their demographic and socioeconomic characteristics. The lack of caste census between 1931-2010 and the lack of caste information from SECC 2011 (Socio-Economic Caste Census) hinders performing any study at a finer caste level. It also highlights increasing inequality within caste groups, and especially within FC and ST group.

mutual funds and securities. The tax was abolished mainly due to the low collection and high cost of tax collection.

[11] Household surveys help provide information for the bulk of the population. However, they suffer from the limitation of not capturing the complete top distribution- either due to non-responses, under-reporting, or worse, exclusion of the rich Jayadev, Motiram, and Vakulabharanam, 2007; Deaton, 2005. To emphasize the issue, I find the combined wealth of the top 118 Forbes millionaires in 2018 is 6% of the total (net) wealth from a nationally representative wealth survey. Hence, I utilize the available Forbes list to correct the top series.

This paper also contributes to the literature on land-ownership inequality [12], while taking a micro-spatial approach, produces village-level agricultural land area inequality statistics for the universe of villages (374k) in ten states of India and highlights regional variation. It establishes the existence of a land-caste relationship through comprehensive village-level analysis. The existing studies using surveys are limited by the small sample, which does not allow to study at a more granular level than districts. Some studies have used a village-level approach but are limited by a few villages. Also, several papers use operational holding (Mohanty, 2001), whereas, in this paper, the emphasis is on land ownership.

I explore the channel of assortative matching (AM) potentially contributing to the rising inequality (Frémeaux and Lefranc, 2020). AM refers to the willingness to marry/cohabit with a person of similar attributes. Papers have studied it along several dimensions like education, employment and wages (Kremer, 1996; Cancian and Reed, 1999; Schwartz, 2010; Eika, Mogstad, and Zafar, 2014; Greenwood et al., 2014. This paper estimates the AM in Indian married couples for education, employment and wages. To my knowledge, there are no estimates for Indian society along those attributes. I find that the education and earnings assortativity is very high in India (and similar to the level in France), despite close to 95% marriages occurring within the caste (and religion). The higher level of education (and economic) assortativity in those caste groups where within-caste inequality is also very high suggests that this channel potentially contributes to the increasing inequality in general and rising within-caste inequality.

The paper is structured as follows. Section 2 outlines the context and data sets used in this paper. Section 3 presents the evolution of wealth distribution in India since 1961. Section 4 presents dynamics of caste and inequality. Section 5 studies land inequality in detail at the village level and explores factors explaining the inequality. Section 6 concludes.

[12] Bauluz, Govind, and Novokmet, 2020 in cross-country macro-approach highlights that the level of land inequality in India is the highest in the world.

2. Context and Data

2.1. Context. India's more than 1.3 billion population bonded together strongly with cultural thread. Religion, caste, language, and other social factors determine the personal identity and play an essential role in many economic decisions at the household and government levels. An analysis of economic inequality is almost incomplete without considering its societal structure. Among them, the thousands of years old caste system is possibly the most entrenched one, dividing society into thousands of groups (jatis).

Caste became part of the census in a full-fledged manner first time in the 1901 census. On the one hand, it started providing demographic and other characteristics for different castes; it also contributed to the rigidification of the caste system in society (Piketty, 2020). The independent Indian government decided to exclude caste from the census due to the fear of re-enforcing the caste identity. However, provisions for affirmative actions were made for Scheduled castes (Dalits) and Scheduled Tribes (ST) in the Constitution in 1951, and the census gathers information on SC and ST. Scheduled caste (SC) was not a single caste entity but a list of castes.[13] Later, in the 1990s, the Indian government extended the caste-based positive affirmative actions for Other Socio-Economic Backward classes of the society. These populations often come from Other Backward Castes (OBC) castes. Again, OBC is not a single caste but a list of hundreds of castes. Since past censuses did not collect information on OBC, it was estimated crudely from the 1931 census using population growth. Later, the Socio-Economic Caste Census was conducted in 2011 to better understand the population share of different castes, but the caste information has not been made public yet. In this paper, I use these administratively defined caste groups.

Historically, land distribution was entirely based on caste, where the upper castes of society possessed almost all the land, and lower castes predominantly formed the working class. The colonial period concreted the possessions of land by distributing land titles and land ownership. The land reforms adopted post-independence (partially to offset the historical prejudice against lower castes) were met only with limited success.

[13] They are also referred to as Dalits and were considered untouchables during that time. They formed the lowest rung in the caste hierarchy. In the varna system, they were the outcasts.

14

No all-India level administrative household wealth data or caste census data is publicly available to study the caste-wealth relationship systematically. Fortunately, there are many all-India-level comprehensive surveys to estimate statistics on wealth.

2.2. NSS- All India Debt and Investment Survey (AIDIS). The NSS-AIDIS are decennial surveys for 1961, 1971, 1981, 1991, 2002, 2012, and 2018. It is the primary data source for generating the wealth inequality series. [15] The surveys are conducted during the calendar year for the agricultural year of the country.[16] The asset and liability on a specific fixed date (mid-point of the agricultural period) are ascertained. Micro-individual survey files are available for the last four rounds of the survey- 1991-92, 2002-03, 2012-13 and 2018-19. For previous surveys, only annual reports are available. [17]

2.2.1. *Sampling Methodology and Sample Size.* Like all NSS surveys, these also have two-stage stratification (See Appendix F.1 for detailed information on the sampling methodology by survey years). Table 2.23 provides the sample size and rural-urban ratio. In pre-1991 surveys (second-stage), stratification to select households was using land possessed by the households, which was suspected to be the reason behind poor capture of the liabilities/debts.[18] All the surveys after 1991 (including 1991) use household indebtedness in addition to other variables (like assets or expenditure) to stratify. Hence, I always provide all the estimates using gross wealth too. [19]

[14]Going by the reports of the government - land re-distribution did not go well beyond certain ranks of the castes.

[15]RBI conducted the 1961 survey in rural areas only. NSSO (National Sample Survey Organisation) and RBI conducted the 1971-72 survey in rural and urban areas. The urban data was never published due to sampling issues. NSSO conducted all the later rounds of surveys.

[16]The agricultural year in India is from July to June. E.g. 1961-62 survey collected information from Jan-Dec 1962 for the reference year July 1961-June 1962. Similarly, the 1971-72 survey was conducted from Jan-Dec 1972 for the reference period July 1971-June 1972. And so on.

[17]The 1981-82 and 1971-72 reports are available at http://www.mospi.gov.in/download-reports. The 1961-62 report is not digitised and is available in hard copy format at the library of College of Agriculture Banking, Pune.

[18]Comparing the debts handed out by commercial banks, cooperatives and other lending agencies,Gothoskar, 1988 found under-estimation of 40% in 1971-72 round and 50% in 1981-82 round. The debt level from AIDIS is suspected to be unreliable due to a strong tendency of under-reporting of liability, issues related to sampling methodology, and relative increase in state sample compared to centre sample Narayanan, 1988; Chavan, 2008.

[19]The other reason to use total wealth is that for pre-1991 surveys, information is available in a tabulated form only which restricts estimation of the distribution of net wealth as the classification is by total asset ownership and not by net wealth status Subramanian and Jayaraj, 2008.

2.2.2. *Types of Assets and Liabilities.* The assets type captured in surveys has remained similar to a large extent, though there have been some changes over the years (Table 2.24). E.g. in 1961-62 (and in 1971, 1981), total assets included the value of 8 different types of assets- i) Ownership rights in land, ii) Special rights in land, iii) Buildings, iv) Livestock, v) Implements, Machinery, Transport Equipment etc. vi) Durable household assets (life more than a year and which can't be purchased at a nominal price) vii) Dues receivables on loans advanced in cash/kind. viii) Financial assets-Government securities, National Plan savings certificates, shares etc. In terms of liabilities, it captured cash loans. In 1991, there was an effort to collect information on cash. From 2002 onwards, bullions and ornaments started to be collected. The 2012 survey did not collect cash in hand. However, in 2012 (and 2019), household durables were not collected on the pretext of valuation concerns. To compare with previous years, I remove durable assets from the analysis. The surveys collected total 159 items under different heads in 1991-92 survey; 141 items in 2002-03; 86 in 2012-13 and 87 items in 2018-19 (See Appendix 2.25). [20]

2.2.3. *Valuation of Assets and Liabilities.* The surveys provide the value of assets and liabilities on a fixed reference date. The total value of the assets and liabilities is estimated as the total stock of assets on the survey date plus the net flow of the value of assets between the reference date and the survey date.

$$\begin{aligned} Asset\_Value_{i,ref\_date} = Asset\_Value_{i,date\_of\_survey} + Acquired\_Asset\_Value_{i,ref\_date+1 \text{ to } date\_of\_survey} \\ - Disposed\_Asset\_Value_{i,ref\_date+1 \text{ to } date\_of\_survey} \end{aligned} \tag{6}$$

where i is the type of asset (like land, building etc), ref_date is the fixed reference date 30th June (1961, 1981, 1991, 2001, 2018 depending on the survey). The valuation of assets is based on the market price remaining in the locality (reported by the respondents) and remained same from 1981-82 to 2002-03 survey rounds, which is detailed in Appendix F.2. Similarly liabilities (loans) is estimated. In 1971-72 and 2012 the value of assets and liabilities was directly collected for the fixed reference date of 30th June (1971 and 2012). Further in 2012-13 land and building values were recorded as per their

[20]Under Durable assets, 22 and 14 items were covered in the 1991-92 and 2002-03 survey, which are not included above.

normative/guideline values (instead of self-reporting). [21] In 2019, building values were recorded as per the market price prevailing in the locality.

2.3. Socio-Economic Caste Census (SECC) 2011. SECC was a unique census conducted in 2011 to get caste information. Its rural component captured household-owned agricultural land area. I use owned agricultural land area to study the village-level caste-land ownership structure. [22]. I received the rural level micro-level dataset for ten large states of India - Punjab, Uttar Pradesh, Bihar, Rajasthan, Madhya Pradesh, Maharashtra, Andhra Pradesh, Karnataka, Tamil Nadu, and Kerala.

2.4. Other Datasets. I utilize several other surveys to provide the population share of different castes in India.

2.4.1. *NSS- Consumption Surveys.* I use the well-known quinquennial NSS surveys of Consumption for the years: 1983, 1987, 1993, 1999, 2004 2009, for which micro-files are available.

2.4.2. *IHDS- Indian Human Development Survey.* IHDS - is a nationally representative panel survey which provides information on the "Brahmin" caste- highest in the caste-hierarchy. I use IHDS 2011-12 to produce general socio-economic characteristics of different castes.[23] I create caste groups based on religion and caste codes based on the following combination- a) Dalits (SC), Adivasis (ST) and OBCs (Other Backward Castes)are coded as they are, regardless of their religion; b) Next, Brahmins are coded as Brahmins. c) Next, all Hindus[24] who are not categorized above are coded as Forward Caste (FC); d) Next, all the Muslims who are not yet coded are coded as Muslims. e) All the rest of the population is grouped as Others. [25] from central and state governments, which directly impacts the educational and income outcomes.

[21]The enumerators had to consult Patwaris (or equivalent) in the rural areas and the Registrar's office in the urban areas to obtain them.

[22]It also collected data on house/dwelling and household amenities (refrigerator, telephone/mobile and motorized vehicle). It only collected whether the household owns them but not their values)

[23]I refer to IHDS 2011 in the whole document. The survey was conducted between Nov 11-Oct'12. The survey sampled 42,152 households with 204,568 individuals.

[24]There are five major religions in India-Hinduism, Islamism, Christianism, Buddhism, Sikhism

[25]The classification is different from the one given in IHDS. I chose this categorization since SC, ST, and OBC have some provisions of positive discrimination

Second, I use this survey to estimate the share of Inter-Caste Marriages (ICM) and assortative matching analysis. [26] ICM is identified based on the answer of eligible women respondents to the question- "Is your family the same caste as your natal family?" A negative response is categorized as ICM. The survey interviewed 39,523 ever-married women aged between 15-49 regarding health, education, employment and income. [27] Assortative Mating (AM) analysis is performed on 34,713 couples. The age of couples is restricted to 15-60 years. All the retired, disabled, and young persons who cannot work are excluded from the analysis. It is essential to highlight the different classifications of employment and education used in the analysis as they are ranked. The AM analysis uses ordered/ranked categories (Refer Appendix Notes F.7).

2.4.3. *World Bank Data.* Village-level geographic and climatic variables comes from Li et al., 2016. Three variables are used here:

1) Elevation (in metres): It is the area's average elevation. This indicator is constructed by averaging information from 1-km resolution global topographic grids.

2) Roughness (in metres): Surface roughness is the standard deviation of the area's elevation. The elevation is constructed by averaging information from 1-km resolution global topographic grids.

3) Precipitation (in mm): I use average decadal precipitation (2001-2011).[28]

---

[26] I create the categories slightly different for the analysis of marriage. Everything is the same except "Muslims OBC" are classified as Muslims instead of OBC. Since the focus here is on marriage, and there is strong adherence toward within-religion marriage, I use this categorization.

[27] One of the issues with studying ICM using this survey is that separate castes of husband and wife are not provided. There could be an underestimation of ICM as inter-caste marriages are usually not disclosed. National Family and Health Survey (NFHS) asks for separate castes from husbands and wives., but income information is missing. Since one of the essential objectives is to estimate income level assortative matching in society, I use IHDS.

[28] The primary source of this dataset is Climatic Research Unit Database Version 3.22 (CRU) University of East Anglia Climatic Research Unit; Climatic Research Unit. Jones, P. D., and I. Harris. 2008. "Climatic Research Unit (CRU) Time-Series Datasets of Variations in Climate with Variations in Other Phenomena." NCAS British Atmospheric Data Centre, 2015

3. Wealth Inequality Series:1961-2018

The real average net wealth per adult has been increasing since 1961, and the increase was the fastest in the decade of 2002-2012.[29] It was Rs. 62,592 in 1961 and Rs 554,126 in 2018.[30] The rural-urban gap during this period has also increased, though the major rise is again in the last decade. The ratio between urban and rural average wealth increased to 1.76 in 2018 from 1.25 in 1981. (Refer Table 2.1).

3.1. Methodology for estimating top wealth shares.

3.1.1. *2002, 2012 and 2018.* I estimate the top share of wealth combining survey data and Forbes millionaire list, post-2002, since Forbes data is present only after 2002. The main motivation of using a rich-list (in combination with survey) is because of non-capturing of wealthy individuals in the NSS-AIDIS survey (Subramanian and Jayaraj, 2008). The maximum wealth in the Forbes list to the maximum wealth observed in the survey is 3250, 16507, 3279 and 7163 times for 1991, 2002, 2012, and 2018 respectively.[31] The total (net) wealth from the Forbes list comes from 5, 46, and 117 individuals in 2002, 2012, and 2018 respectively owning 1.26%, 2.74% and 6.01% of the total survey wealth. The issue of under-representation of super-rich households in the survey is exacerbating over the years. The survey-based estimate will underestimate the wealth distribution.

First, using the millionaire list, the top end of the wealth distribution is estimated, assuming the Pareto distribution at the top end of the distribution. Though the millionaires' list provides a list of few individuals, a convenient property of Pareto distribution

[29] I convert household level wealth to per-adult individual (>20 years) level by equally splitting it among all the adult members of the household, following the literature. An equal split of household wealth is a big assumption. Women usually do not own wealth due to customary transfer of wealth from father to son, biased gender inheritance laws before 2005, and general gender discrimination. Though most physical wealth (e.g. land, building, transport) acts as public goods within a household, an equal split assumption is a more practical choice.

[30] Since for 1961 and 1971, the data is only available for rural areas, I estimate national level averages by simply taking the ratio from 1981. The evolution of average rural and urban area averages in 1981 and 1991 is very close. The constant 2010 level values are computed using the wholesale price index from the World Bank Dataset.

[31] The historical exchange rate is taken from the Foreign Exchange Dealers' Association of India. Exchange rate conversion from $ to Rs: 1991- 24.5, 2002- 48.4, 2012- 54.4 and 2018- 68.4. The Forbes wealth data is assumed at an individual level, though some of the names in the list suggest it to be household level wealth. e.g. Birla family. The unavailability of the household size restricts taking it to the individual level. The potential concern is that if the wealth is truly at the household level, the discrepancy between Forbes and the survey will be inflated.

is that the curve of log (wealth) and log (rank) follows a straight line, allowing to generate the distribution.[32] This newly generated top-wealth distribution is used to replace the top-wealth distribution generated from the surveys (Blanchet, 2016). This approach does not tell precisely the threshold above which the distribution will be a good approximation. Hence I show how the measure of wealth owned by the top 10%, 5%, 1%, 0.1% and 0.01% changes by using different cut-off thresholds (p=95 means I replace the top 5% survey distribution with Pareto generated distribution). The top wealth share post-correction is present in Table 2.27.

As a robustness check, I compute the top wealth share using an entirely different rich list provided by the Business standard for 2002 and 2012. The top wealth shares are very similar to the ones calculated using Forbes. E.g. Top 10% population own 55.3% and 62.5% in 2002 and 2012.

3.1.2. *1991.* For the year 1991, the survey's micro-dataset is available, but there was just one person on the Forbes rich list (and no other rich list), which is insufficient to compute the Pareto coefficient. Hence, the wealth inequality estimates are using only the survey.

3.1.3. *1981, 1971 and 1961.* For these years, the data is limited to the survey report tabulations providing average wealth and number of households into different wealth brackets. One could use the non-parametric generalized Pareto interpolation method to generate a continuous distribution of wealth from the tabulations. [33] However, there are two problems. The tabulations give the number of households (not the number of adult individuals, which is the basic unit of reference), and wealth brackets are on gross wealth (not net wealth). I estimate the average adult and household sizes at different percentiles, assuming its decadal rate of change remains unchanged from 1981-91 to 1991-2002. I corrected the estimates to keep the total predicted population the same as that of the survey, and the overall adult population share should be the same as in the

[32] A straightforward method is adding these few individuals to the dataset. This simple method ignores individuals between the poorest individual from the Forbes list and the wealthiest individual in the survey.

[33] Micro-data from the surveys in themselves provide the whole distribution. However, I test the generalized Pareto interpolation by estimating the entire distribution and comparing it with the survey distribution for the 1991-2012 surveys. The difference in decile level shares from the generalized percentiles method and survey shares is positive in lower deciles and negative in the top decile. The reduction in top decile share is 1.76 pp, 1.7 pp and 2.6 pp in 1991, 2002 and 2012, respectively, using a generalized percentile programme. It provides a robustness check that there is no overestimation.

census (See detail in Appendix (F.3).

Next, to generate net wealth distribution, I assume that the bracket will remain unchanged after taking out the debt from the gross wealth. This strong assumption is justified by the presence of large bracket sizes where less than half have any debt, implying that a large fraction of the population will remain in the same bracket. From the report, it is possible to calculate the net wealth for 1971 and 1981. But for 1961, the information is missing, and I used the debt ratio (total debt/ total gross wealth) by wealth brackets from 1971 to estimate the average net wealth in 1961.

The lack of rich list correction implies the inequality estimate will be a lower bound for 1961-91; hence the change in inequality from 1991 to 2002 should be seen with these lacunae in mind. However, for this period, the downward bias is most likely small for the following reasons. First, given the adopted socialist post-independence policies, the (land/building) market was less developed, population pressure was less, and urbanization levels were low. Second, the financial asset was less than 5% in pre-1981 which is an asset easiest to under-report. For E.g. the stock market was underdeveloped.

3.2. Concentration of Total Wealth. The wealth inequality increased slowly in between 1961-81 (pre-liberalization), thereafter it increased faster for the next three decades to reach a peak in 2012. Post that, the increasing trend seems to have stopped and in 2018, the level of inequality is lower than 2012. In the following discussion, I will use the estimates for net wealth, though the trend remains same using gross wealth.

The distribution's two ends (top 10% versus bottom 50%) are evident even in the pre-liberalization period of 1961-81. The top 10% population owned 45% of the total wealth, whereas the bottom 50% of the population owned just 11%. The middle 40% held close to 45% of the total wealth (See Table 2.2). The urban sector is more unequal than the rural. In the 1981 urban sector, the wealth share of the top 10%, middle 40% and bottom 50% were 50%, 42% and 7.4% respectively. In 1991, the year of economic crisis, the wealth share of the top 10% jumped to 52.5% with a corresponding decline of 4 percentage points in both middle 40% and bottom 50%. The decline in the bottom 50% was most prominent in the urban sector (wealth share declined from 7.4% in 1981 to 0.8% in 1991) (See Table 2.4 or Figures 2.1 - 2.5 for graphical representations).

The post-liberalization period (post-1990s) continued with the increasing wealth inequality. The wealth share of the top 10% increased to 55.6% in 2002 and 63% in 2012. The improved inequality estimates (using Forbes' rich list, overcoming the survey-based estimates) highlights that the wealth is heavily concentrated at the top. The wealth distribution is highly unequal within the top 10% of the population. Out of the total wealth held within the top 10% population, the top 5% own 77% (=42.9/55.6), 80% (=50.5/62.8) and 82% (=49.8/61) in 2002, 2012 and 2018. The top 1% of the population owned 24%, 31% and 33% of the total wealth in 2002, 2012 and 2018 respectively. The wealth share of the top .01% population also increased from 10% in 2002 to 12% in 2012 and 23% in 2018.

The last round of survey shows decreasing inequality (decreasing wealth share of top 10% and increasing wealth share for the rest 90% of the population). The survey-based estimates for the wealth share of the top 10% population show a decline of 7.8 pp (=59.9-52.1) between 2012-2018. There are two things to consider which could drive this decline- change in sampling methodology and increasing non-capturing of the wealthy individuals in 2018 compared to 2012. The 2018 survey did the second stage of stratification (to randomly select households in a selected village or urban block) based on the household's consumption and indebtedness compared to only household debt in 2012. I show that it is not the reason behind the declining inequality (See Appendix notes F.4 for details).

It leaves the issue of the non-capturing of wealthy individuals. In 2012, the survey did not capture the top 46 wealthy individuals (with 2.7% of total survey wealth) plus the population with wealth between the 46th richest individual (Rs 54,410M) and the wealthiest individual in the survey (Rs. 370M). In 2018, the non-capturing increased. The survey did not capture the top 117 wealthy individuals having 6% total survey wealth. Also, the gap between the lowest Forbes (i.e. 117th individual with Rs 68,411 M) and the wealthiest survey-based individual (Rs 383M) increased in 2018. Hence, the post-correction estimates for the top 10% wealth share compared to the survey-based estimates increased by 9pp in 2018, which was 3 pp in 2012. The reduction in the wealth inequality (measured by the top 10% wealth share) between 2012-2018 reduced to 1.8pp (from 7.8pp).

The wealth inequality in India (measured by the wealth share of the top 10%) was higher than in China but lower than in South Africa in 2002. From 2002-2018, the rise of wealth inequality in China was faster than in India. In China, the wealth share of the top 10% population was 68% compared to 61% in India (and 86% in South Africa). Compared to other developed countries, the wealth share of the Bottom 50% population in India is higher. (See Table 2.5)

3.3. Composition of Household Wealth. Physical assets dominate household wealth in both rural and urban areas. Among the physical assets, *land* is the most significant wealth, contributing 60% of the total wealth. In the rural sector, land is the most valued asset. The share of the wealth attributed to land has remained almost constant since 1961 at a level of approximately 67%. The land share is lower in the urban sector, but it has an increasing trend from 38% in 1981 to 48% in 2018. *Building* comes next, and together with the land, constitutes 90% of the total household wealth. Due to the more developed real-estate market and high population density in urban areas, the building share in urban areas (40%) is double that in the rural sector(20%). *Financial assets* comes third largest, and it increased to 10% in 2018 from 4% in 1981. It contrasts with a high level of financial assets in developed countries or even other developing countries. [34] (See Table 2.6)

For ease of discussion, I will use rich, middle and poor classes for the top 10%, middle 40% and bottom 50% population based on overall wealth.

The land is the most valuable asset throughout the distribution. It forms 64% of the total wealth for the rich class, 55% for the middle class and 40% for the poor class for all years (Figure 2.6). It has a vital role in shaping the overall wealth inequality in the country. Building contributes to 25-30% of the wealth of the rich and middle classes. In the poor class, the share of buildings is close to 40%. The percentage of other assets within each class category has also remained stable from 1991-2018.

[34]In France and the US, the share of financial assets stood at 30.9% and 48.3% in 1979 (Kessler and Wolff, 1991).

Over the years, the total land owned by households has declined from 308M acres in 1991 to 251M acres in 2018. The landless population share has remained close to 14%. Landlessness is highest among the poor class - nearly one-quarter of the population owns no land. It is higher among the urban poor class (bottom 50% living in urban), where more than half do not own any land. Within the rich class, those living in rural areas all own some piece of land though those living in urban areas close to 10% do not own any land. Regarding the share of the land area owned, the rich class owns 30% (all years), the middle class holds 50% (declined from 55% in 1991), and 19% by the poor class (14% in 1991) in 2018. The total land area owned slightly increased for the poor class from 43.4M acres in 1991 to 46.6M acres in 2018 (the significant gain occurred during 2002-2012). For the rest of the population, the total land area owned has declined. (See Table 2.7)

The predominance of the land value and decreasing land area suggests land price (value per unit area) plays a significant role. In 1991, the average real land price for the rich class was four times higher than the middle class and eight times higher than the poor class. In the next 27 years (from 1991-2018), the annual growth rate of land price was the highest for the middle class (9.2%), followed by 8.8% for the rich and 7.9% for the poor class. It reduced the gap slightly between rich and middle (3.7 times), though it increased between rich and poor (11.4 times). Between 2002-12 (a fast rise in inequality), the average annual growth rate of land price was 10.7% for the rich and 11.9% for the middle but only 6.2% for the poor class. (See Table 2.8)

One of the reasons behind increasing land premiums could be a change in land use. E.g. converting agricultural land for non-agricultural land purposes (for business or residential purposes) will increase the land price. The surveys have captured information on the different types of land use. I create three mutually exclusive categories of land - agricultural, residential and Others (and non-residential) .[35] In 1991, the agricultural

[35]The categories have changed slightly in different rounds of the survey. The survey in 2012 and 2018 captured the land ownership in rural and urban areas separately. Surveys in all years captured residential/house site land. Agricultural land covers 1) In 1991 and 2002, combining three types - irrigated land, un-irrigated land, orchards and plantation; 2) In 2012, combining four types of land in both rural and urban -irrigated land, un-irrigated land, orchards and plantation and forest 3) In 2018 - combining four types of land in the rural area (as in 2012) and crop area (irrigated/unirrigated) in the urban area. The rest of the land falls into Others category.

land area share was 94% which reduced to 90% in 2018. During this period, there was a 2pp gain in both residential and non-agricultural types land, both from 3% to 5% (See Table 2.9)

The portfolios of different land types in the wealth basket of rich, middle and poor classes show a similar declining agricultural land share pattern. The agricultural land share of rich, middle and poor classes declined by 7pp, 4pp and 2pp, respectively, from 1991 to 2018. In 1991, the agricultural land share was 96%, 95% and 90% for the rich, middle and bottom classes, respectively. The residential land share gain is 2 to 3 percentage points for all three classes. In 2018, within the land portfolio of rich, middle and poor classes, the residential land type was 4%, 5% and 9% respectively. There is an absolute rise in the land area dedicated for non-agricultural purposes for all classes, with the main difference that within the rich class, it is the urban rich driving the rise. The urban rich non-agricultural land area increased from 0.81M acres in 1991 to 5.32M acres in 2018 (for the rural rich, it increased from 2.94M acres to 3.08M acres). In contrast, among the poor class, the gain in the non-agricultural land is coming from the rural poor. The rural poor non-agricultural land area increased from 4.04M acres in 1991 to 5.04M acres in 2018 (for the urban poor, it increased from 0.30M acres to 0.55M acres). (See Table 2.10).

Lastly, the land price premium of the rich class is present within each land type, and it has increased between 1991-2018. The residential land of the rich class was 8.8 times more expensive (than the residential land of the poor class) in 1991, which increased to 11.8 times in 2018. One reason could be that the rich class might be over-represented in urban areas. To overcome this, I separately compare the land prices for each type of land within rural and urban areas. The price premium of the rich class in residential land type is present in both rural and urban areas, though it has increased only in the urban areas. The residential land of the urban rich class was 5.3 times more expensive in 1991, which increased to 6.3 times in 2018. In rural area, this premium has remained close to 3.2 (See Table 2.11). Similarly, there is a price premium for another non-agricultural land type for the rich class (in rural and urban areas). The land premium on non-agricultural land for the rich class living in urban areas could be due to the increasing availability of better infrastructure facilities (drinking water, sewage facilities, road connectivity etc.).

In conclusion, the analysis shows that rich class land wealth is increasing both from acquiring non-agricultural land (which are more expensive) and higher land price premium (possibly due to better infrastructure).

4. Economic-Social Inequality

"No collection of wealth must be made by a Sudra, even though he be able (to do it); for a Sudra who has acquired wealth, gives pain to Brahmanas." Manu Smriti 10.129

This section explores the wealth inequality along caste groups.

4.1. Population Share by Caste Groups. Census data shows that the SC population share has increased from 14.67% in 1961 to 16.6% in 2011. During the same period, the proportion of ST has increased from 6.23% to 8.6%. No census data for other caste groups (like OBC, Forward Caste) is available. The surveys (post-2000) can complement understanding other caste groups' population share.

The population share from the survey (NSS-AIDIS) also shows an increase in all the lower caste groups. The population share of SC was 19%, ST- 8%, OBC- 40%, FC-23% and Muslim 8%[36] of the total population in 2002. The population share has increased for ST by 2pp, SC by 1.2pp and OBC by 4.2pp in 2018 compared to 2002. Correspondingly, there is a 3.45pp decline in FC share and a 2.35pp decline in Muslims (See Table 2.12). Other NSS survey datasets provide similar trends - an increasing share of lower caste groups (SC/ST/OBC) and a decreasing share for FC/Muslims (See Appendix F.5). Apart from the high natural growth rate of lower caste groups, re-classification into these (from FC and Muslims) is potentially another reason. I come to this caveat in the last subsection.

IHDS and NFHS surveys allow to split the Forward Caste group further into Brahmins (5%), Rajputs (a proxy for Kshatriyas; 5%), Bania (merchant class; 2%), Kayasth (0.6%) and Others (9.3%). These are the highest castes(jatis) within the FC group. Interestingly, the population share of Brahmins and Rajput, the two topmost castes in the hierarchy, has remained similar compared to their population share in 1901 (or 1911). (See Appendix F.5 for further details)

[36]It is the share of Muslims who are not SC, ST or OBC, hence lower than the total Muslim share. As per Census 2001, the Muslim population share was 13.4% which increased to 14.2% in 2011.

The economic rank order of the caste group follows the caste hierarchy, making caste a relevant stratifying indicator. The average annual income in 2011 of the ST and SC group was 0.7 times and 0.8 times lower than the all-India average. OBC and Muslims had around 0.9 times the all-India average and Forward castes (FC) 1.5 times. Based on average income (or consumption), the groups are ranked as $ST < SC < Muslim < OBC <$ **OVERALL** $< FC(Non-Brahmin) < FC(Brahmin) < Others$. It remains the same using average net wealth, though the gap between SC (and ST) with others increases. The average net wealth of SC (and ST) is half the national average and has remained the same between 2002-2018. (See Appendix Table 2.30). [37]

4.2. Representation Inequality by Caste Groups. The difference between the (net) wealth share and population share by caste groups shows a caste gap in wealth. A negative value implies that the caste group owns less wealth than their population share. It is negative for SC, ST, OBC and Muslims for all years(See Table 2.13). SC suffers the worst; it owns only 9% of net wealth, which is 10 pp less than its population share. ST owns $4-5\%$ of net wealth, resulting in a negative 5pp gap. OBC owned $\sim 36\%$ of total net wealth in 2002, which increased to 40% in 2018; the gap remained similar at -4.6pp due to their increasing population share. Muslim group (after taking out SC/ST/OBC) gap is -1.29pp. For the FC group, the gap is positive at 15pp. The higher castes (FC) own most of the wealth. There has not been much change over the years, showing no convergence across caste groups.

Next, I check which part of the distribution different caste groups lie. I use the term representational inequality (RI) from Jayadev and Reddy, 2011 symbolizing segregation among social groups in a given attribute space (which here is wealth). The basic idea is to compare the population share of a given caste group in a given wealth decile with its

[37]Muslims who are almost closer to the all-India average in economic parameters fall behind the SC group in education. Other ranking remains the same. In terms of (adult) education in 2011, the ranking becomes $ST < Muslim < SC <$ **OVERALL** $< OBC < FC(Non-Brahmin) < FC(Brahmin) < Others$. The difference in average years of education between Brahmin and ST is 5.6 years, Brahmin and Muslim is 4.9 years, Brahmin and SC is 4.8 years, Brahmin and OBC is 3.7 years, and Brahmin and Other FC is 1.2 years. Using NFHS 2005, one can see the differences within the FC group. Kayasth with 12.3 years of education is the highest, followed by Brahmins (11.9years), Bania (10.3 years), Rest of FC(9.16yrs) and Rajput (9.05 years) (Refer Appendix Table 2.33). NFHS doesn't provide information on income or wealth.

share in the overall population. E.g. if the OBC population is 45%, perfect representational equality will imply the representation of 45% OBC population in all the deciles. Any deviation will lead to Representational Inequality. [38] The statistic is given by:

$$RI_j^{YCD} = \frac{Popshare_j^{YCD} - Popshare_j^{YC}}{Popshare_j^{YC}} \tag{7}$$

where $Popshare_j^{YCD}$ denotes the population share of caste C in decile D in year Y for sector $j \in (Rural, Urban, India)$. A positive value in the higher (lower) deciles for a caste group denotes it as a beneficiary(victim) of the inequality. Also, the higher the value, the more representational inequality.

The RI statistics are present in the Table 2.14 for three categories- Top 10%, Middle 40% and Bottom 50%. Within the top 10% decile, we see that only FC[39] have positive RI. Even in the Middle 40%, only FC has a positive value. SC, ST and Muslims have positive values in the Bottom 50%. The RI is close to zero only for the OBC group. The takeaway is that FC is disproportionately present in the top 10% and Middle 40% of the population, where almost 90% of the wealth resides. SC, ST and Muslims are disproportionately present in the Bottom 50%. The distribution of OBC is even across wealth deciles. Also, there was not much change between 2002-2018. An analysis of rural and urban areas separately also highlights the worst under-representation of the SC in the top wealth decile (or even in the top 50% decile) (Refer Appendix F.6).

The results show a very high level of sequential inequality (or clustering of social groups). The FC group is clustered more towards higher income/consumption/ wealth values. OBC and Muslims are in the middle and SC/ST cluster towards the lower end. This clustering of different groups is synonymous with the caste hierarchy, highlighting the imprint of the caste-based distribution of wealth in the past (and the lack of enough re-distribution through government policies to overcome this). In the next section, I explore it further using land area ownership, which is directly linked to the caste system. Before that, I look into the total wealth inequality within each caste group.

[38] The intuition is very simple: Suppose the wealth inequality is independent of social inequality, then the probability of falling into different wealth deciles is the same for all the caste groups. Any digression is due to the interaction of social and economic inequality.

[39] and a small group of "Other", which is a small, rich minority group

4.3. Total Wealth Inequality within Caste Group. Each broad social/caste group - SC, ST, OBC, FC - comprises hundreds of narrower versions of castes (jatis) in a hierarchical fashion. The standard deviation of annual income and consumption from IHDS 2011, capturing the variation present within each group, is highest for the FC, followed by SC, Muslim, OBC and ST (Refer Appendix Table 2.31). An additive decomposition of Theil's wealth inequality index into between and within components by caste groups highlights the importance of within components. The between component share capturing the "caste effect" is 12-14%. The within component explains a major share (86-88%) of wealth inequality. [40]

Fig. 2.7 shows the wealth share of top 10% (rich) and bottom 50% (poor) within each caste group. Since the within-components dominate, it is not surprising that the wealth shares of rich and poor classes within each caste group follow the general trend as the overall wealth shares (computed from surveys). E.g. The wealth share of the rich class within each caste group gained between 2002-12, though the levels differ between different groups.

The ST group was the least unequal, with the wealth share of rich ST at 40% and the poor ST owning 12% of the total ST-owned wealth. The wealth concentration in the next 30 years has consistently increased within ST. In 2018, the rich ST held close to 55% of the total wealth. The loss was most for the middle-class ST, as the drop in the wealth share of poor ST was close to 2pp during this period. The trajectory of wealth concentration in Muslims was similar to the ST group, though the poor Muslim class owns 8% of the total Muslim wealth. The rich SC used to own 46% of the total SC wealth in 1991, which has remained at a similar level in 2018. There is a gain in the bottom SC class, as their share increased from close to 8% to 12%.

The FC group's evolution shows a significant rise in wealth concentration from 2002-12. The rich FC class wealth share became 60% in 2012, from 47% in 2002, though it declined in 2018 to 50%. The trajectory of wealth concentration within OBC is similar to

[40] The values of Theil's index for 2002, 2012 and 2018 are 0.84, 1.31, 0.90. The between-component of caste groups is 0.12, 0.17, 0.11 which is 14%, 13% and 12% respectively.

the FC group.

4.4. Assortative Matching and Inequality. Motivated by the emerging literature on assortative marriages leading to an increase in inequality combined with the evidence of rising within-caste inequality, I explore this channel in detail.[41] The Indian context is interesting because caste and religion play the most dominant role in the marriage market and other attributes like education, income, and wealth are secondary. Close to 95% marriages occur within the caste, and close to 98% marriages occur within religion. (Ray, Chaudhuri, and Sahai, 2017, Das et al., 2011). I created a new dataset of married couples using the IHDS 2011 survey.[42] Appendix Notes F.9 provides the descriptive statistics of couples.

4.4.1. *Caste Assortativity.* The level of inter-caste marriage has increased marginally from 5% in the pre-1980 married cohort to 6% in the 2006-12 married cohort, i.e. 1 pp increase in 25 years (See Figure 2.8). It presents a very high level of persistent caste assortativity at 0.95. There is an increasing trend of ICMs in urban areas but remains below 10%. The postgraduates have a slightly higher level of ICM at 7.5% (See Appendix Figure 2.16). [43] It highlights strong caste homogamy in the society and own-caste preferences when it comes to marriages.[44]

4.4.2. *Education Assortativity.* I estimate correlation coefficients of the ordered categories of education.[45] The higher the magnitude of the coefficient, the higher the assortativity. The correlation between husband and wife's education in the full sample is 0.63 for Spearman (rural- 0.57 and urban - 0.66), and it is very similar for all the caste groups

[41]The debate on the impact on inequality is unsettled. Frémeaux and Lefranc, 2020 estimates a non-negligible effect of 3%-9% on the measured household earnings inequality in French working couples. The effect increases to 10%-20% on potential household earnings. On the other hand, Olivo-Villabrille, 2017 does not find the impact of AM alone in US data. It only acts as an amplifier of the underlying inequality in wages across educational groups. Greenwood et al., 2014 concludes that for AM to impact the household level of inequality, a married woman must work.

[42]The IHDS survey does not capture wealth information, but it captures income and education, which are a good proxy.

[43]Ray, Chaudhuri, and Sahai, 2017 also finds no statistically significant impact of education, urbanization and caste group on ICM rejecting modernization theory. It is in contrast to findings from the US, where studies have shown an increase in inter-faith marriages among educated cohorts Qian, 1997.

[44]Abhijit Banerjee et al., 2013 finds horizontal preference, i.e. preferences for marrying within caste in a rich, educated middle-aged cohort. Looking at patterns of ICM by education levels, we see an increasing trend with education levels.

[45]The details on the computation of Spearman, Polychoric and Pearson correlation coefficients are in Appendix Notes F.10

(Panel A of Table 2.15). Provided high illiteracy, 22% husbands and 40% wives have zero education, which could be driving up the correlation. I re-estimate the coefficients after removing couples with zero education years (both husband and wife with 0 education). The coefficient decreases to 0.5 (rural - 0.4 and urban- 0.6) at the all-India level. There is an increasing trend in education assortativity by marriage cohort (5-year moving average). The correlation was low at 0.35 in the pre-1980 married couples, which increased to 0.6 among the 2006-12 married couples (See Figure 2.8).[46]

The education assortativity differs among caste groups. The Spearman correlation for couple's education in decreasing order among caste groups are - FC (0.61), OBC (0.44) ST (0.39), SC(0.38), and Muslims(0.36) (Panel B of Table 2.15).

4.4.3. *Economic Assortativity.* The correlation between husband and wife's occupation types is 0.43 for Spearman (rural- 0.39 and urban - 0.54).[47]. The correlation is high at 0.7 (Pearson) for husband and wife's wages (with non-zero wages).[48] (See Table 2.16) The economic assortativity (wage or occupation) is the highest for FC and ST, followed by OBC (and Muslim) and SC. For E.g. the Pearson coefficient for wage assortativity is 0.75 for FC and ST, 0.6 for OBCs (and Muslims) and 0.55 for SCs. The order of the caste groups observed here is similar to the within-caste wealth concentration observed before. The rich class within FC and ST have the highest wealth share, followed by OBC and SC.

4.5. Reclassification of Other Caste Groups into OBC. One peculiar trend is the increasing OBC population share, which seems more than the natural growth rate. Over the past years, the demand from several castes(jatis) groups to be included in the OBC list has been accepted, while there has been no exclusion from the list in the last 17 years (Annual Reports of NCBC). It is possibly one of the reasons behind the increase in the population share of OBC. It makes the OBC group less stable in composition compared

[46]Appendix Figure 2.17 plots the correlation coefficients for the full sample (including the zero education couple), where also the increasing trend in the Spearman coefficient is present.

[47]The economic assortativity by marriage cohort will be very noisy (due to the small sample) as occupation types and wages are available for only a fraction of the population.

[48]The Spearman correlation is at 0.52. I refer to the Pearson statistic for the continuous variable because it is better than Spearman. The Pearson coefficient is also closer to the qualitative evidence using one of the survey's questions. It asks women to compare the economic status of a natal family to a husband's family- 74% women feel that they marry to the same economic status family. 16.5% feel that their natal family was economically better off, and the rest of 9.4% feel that their husband's family is better off.

to the more rigid SC and ST group. [49] This issue has been politicized now with increasing demand from FC group to be categorized as OBC to avail the benefits of the reservation (positive affirmation). E.g. "Jat" and Patidar community demand. [50]. [51].

Also, there is an increasing representation of OBC in the top 10% wealth decile. Figure 2.19 depicts the change in the population share of OBC in different deciles between 2002-12. There is a positive change in the population share of the OBC group across all the deciles. Correspondingly there is a decline in the population share of FC and Muslims. The difference is that the decrease in the FC population is in the higher deciles, whereas the decline in the Muslim population is more concentrated in the lower deciles. It suggests that relatively rich FC castes are categorized in the OBC category.

[49] Mandal Commission in 1980 charted out the criteria based on which OBC status is conferred to a caste for receiving the reservation benefits. A score is calculated based on social, educational and economic criteria. Castes which score above a fixed point get OBC status. Higher weight is assigned to social criteria. It ensures that socially backward castes have a higher chance of getting the status. The National Council of Backward Commission (NCBC)- a statutory body, hears the petition from different castes to request inclusion into the OBC list.

[50] "Jat" an agricultural community from North India, are demanding the OBC status in the Central Government list of OBC. There are two lists: The Central level OBC list, which makes a caste eligible for reservation benefits in central universities and central government jobs, and a corresponding state-level OBC list. https://en.wikipedia.org/wiki/Jat_reservation_agitation

[51] Patidar (people with well-known surnames Patel) group in 2015 started agitation for similar demand, which became the central issue in the 2017 state election.; https://en.wikipedia.org/wiki/Patidar_reservation_agitation

5. Village-Level Agricultural Land Inequality

The share of land owned by households to the total land in India has declined from 40% in 1961 to 31% in 2018. [52] The average area owned by households is also on the decline. It has decreased from 2.56 acres in 1981 to 0.96 acres in 2018. Further, in rural areas average area owned per household decreased from $\sim$ 4.8 in 1953-54 (3.16 acres in 1981) to 1.3 acres in 2018.

Table 2.17 provides the representational inequality in *land area distribution* by caste groups. Only the SC group have a lesser land area among the lower caste groups than their population share. They own 9pp less land area share than their population (close to 19%). The total land area with ST was 2 pp more than their population share in 1991 and increased slightly to 2.51pp in 2018. ST group own more land as most of them live closely with nature in forests with large land areas.[53]. The OBC group also own approximately 1pp more land share than their population.[54]

The surveys are an excellent source to establish macro-level land inequality among the SC caste group. However, SCs are spread across millions of villages dependent on agriculture and work as agricultural labourers. The SECC-2011 data provides a unique opportunity to study the relationship between caste and agricultural land ownership at a village level- the lowest level administrative unit - which is impossible through the surveys due to their small sample size. The SECC data is limited to agricultural land ownership, which forms close to 95% of the household-owned land in rural areas. Table 2.9 gives a snapshot of the coverage of the population in the available SECC dataset. I use it to build land inequality measures at the village level in ten large states of India. I combine it with village-level characteristics from two other sources - Census 2011

[52]The pace of decline increases over time and is highest in the recent decade. The reduction can be generally devoted to the increased pace of development post-1991 liberalization, where the land is used for developmental purposes- roads, commercial buildings or other infrastructure development.

[53]The increase could be due to the implementation of the Forest Rights Act (2006). It grants legal recognition to the rights of traditional forest-dwelling communities. Though the Act is not only restricted to ST, the condition that the last four generations should have operated the forest land will be more the case for ST. Land title rights were one of the rights. As per Aggarwal, 2012, 1.1169 million claim covering 3% of the forest area was recognized till 30th Apr 2011.

[54]Though looking at the *land value distribution*, we go back to the same level of inequality as we observed in total wealth. The total land value share is 2 pp less for ST and Muslim, 11 pp less for SC, 7.9 pp less for OBC, and 14.7 pp more for FC. It suggests that FC owns high-valued land, and ST/OBC owns low-valued land.

(area, demographic, literacy and employment characteristics) and World Bank Dataset (geographic and climatic), and look into the factors affecting land inequality.

5.1. Village Level Characteristics. I first present the average village level characteristics using the Census 2011.

Number of Households: On average, villages in MP (220), Rajasthan (225), UP (274), Punjab (274), and Karnataka (298) have less than 300 households. West Bengal (356) and Maharashtra's (327) average villages have between 300-400 households. Bihar (455) and Tamil Nadu (635) villages have more than 400 households and 2500 population. Kerala has a different definition of a village where villages have, on average, 4000 households.

Population density (number of persons/sq. km) : Bihar(1420) and UP(1219) have very high population density, followed by West Bengal(957) and Kerala (918). MP, Rajasthan and Maharashtra have the lowest rural population density (less than 300).

Scheduled Caste/Tribe share: The population share of SC is close to 20% in rural India. In rural areas, we see variations across states. Punjab has the highest SC share at 35.8%, followed by Tamil Nadu (28%) and West Bengal (27.5%). Kerala (10.5%), Maharashtra(11.3%), and MP (14.6%) have a lower share of the SC population. However, Maharashtra (20%) and MP (31.2%) have a high ST share compared to the all-India average of 8-9%.

Literacy Rate : Literacy rate in Bihar (49%), Rajasthan (50%), MP (53%) and UP (56%) are less than the national average of 63%. All Southern states have a higher literacy rate, with Kerala leading the chart.

Working Population Share (total working population/total population): The working population share is less than 35% in the villages of Bihar, UP, Kerala and Punjab. West Bengal has a 41% working population share. The rest of the states have villages with more than 50% working population share.

Agricultural Population Share (total working population in agriculture/total working population): Kerala (22%) has exceptionally the least agriculturally dependent working population in its village. More than half of the working population is engaged in agriculture in other states, and it is >70% in UP, Bihar, MP and Rajasthan.

5.2. Village-Level Land Inequality Measures.

5.2.1. *Landlessness and Land Dependency Ratio.* Agricultural landlessness varies from 35% to 75% in an average village in different states. Among the four states where the dependency on agriculture is very high, Rajasthan (35%) and UP (42%) have a relatively lower level of landlessness than MP (53%) and Bihar (61%). Punjab has a very high level of landlessness at 76%. (See Appendix Table 2.45)

Considering different shares of the population engaged in agriculture in different states, I create a statistic- landless households per landowning household. It provides an idea of (agricultural) dependency in a village. (Table 2.18)) E.g. In an average village in Punjab with 275 households, 140 households are dependent on agriculture (52% working population in agriculture), and 66 households are land-owners; so, on average, 1.12 landless households (=(140-66)/66) are dependent on one landowning family. Among other states, Bihar and Tamil Nadu have a high dependency. UP differentiates itself from Bihar here, with a much lower dependency of landless households on the landowning population. A negative value implies the presence of non-agriculturally dependent households owning land. Only Kerala has a negative value.

5.2.2. *Gini and Top decile share.* I create two well-known statistics, namely Gini and Top decile share (total land area share owned by the top 10% of the households in a village) at different units - household level, equal split among all members of households and equal split among adults members only. [55]

Both statistics highlight a very high level of land concentration. In all the states, the level of the Gini coefficient is more than 0.6, increasing to 0.8-0.9 in Punjab, Bihar (and Kerala). It is not surprising that around 50% of the population in rural areas is landless. The level of agricultural land inequality is comparable to the high land inequality often

[55] The household-level inequality ignores the family size. Two families with the same land but different household size is treated similarly. An equal split among all household members divides the land area equally among all the household members. An equal split among all household members divides the land area equally among all the adult members (> 20 years) of a household only.

found in Latin American countries. Top 10% of the households own 54% of the total land in Rajasthan, and it increases to 88% in Kerala. Punjab, Bihar and Tamil Nadu have an extreme land concentration in rural areas. The values are very close to 80%.

5.2.3. *Land Share of the Top 1, Top 2 and Top 3 HHs.* The granularity of data allows estimating the share of land owned by the top 1, 2 and 3 households in a village. It is an interesting and critical statistic which captures the stronghold of the richest households in a village. It could potentially be used to study village-level voting patterns.[56] Appendix Table 2.46 provides the average values of the statistic. The share of land owned by the top 3 households in an average village is lowest in Maharashtra at 20%, and the highest is in Punjab at 40%.

5.3. Factors affecting land inequality. In this subsection, I try to understand the factors explaining the land inequality at the village level. The main factors in land inequality literature can be categorized under three heads. First, institutional setup usually has the most significant effect on the land distribution in a country. In the Indian setup, the most critical institutional features impacting land concentration are- the historical caste system, colonial land revenue system and post-independence land reforms. Historically, upper-caste (and lower castes in a few areas) are associated with more land ownership. In the medieval period, kings started paying priests (brahmins) and military leaders (Kshatriyas) inland for state services. The British colonial government started awarding land titles during their regime, thereby concretizing land ownership. They introduced different land revenue systems in different parts of India with the primary objective to increase land revenue and rents but without reforms targeting reforms. It was only after independence that India adopted land reforms in three waves: Zamindari abolition, Tenancy laws and Land ceiling, which helped reduce the land concentration to some extent. However, due to various reasons, the success in redistributing land was partial.[57] Second, demographic factors like population density. Third, geographic and climatic factors.

[56]Work by Andres Siegfred on rural setup in France found that villages with a high land concentration were voting more for the right-wing.

[57]Besley et al., 2016 notes that tenancy reforms benefited richer and more productive middle-caste tenants, but reduced land access for poorer low-caste tenants.

In this section, I perform a simple correlational exercise to understand potential factors explaining village-level land inequality. I combine land inequality dataset (produced from SECC) with census data (for demographic and geographic controls) and World Bank dataset (for climate controls like- temperature, precipitation, roughness and elevation). I use district (or subdistrict) fixed effects to absorb the different institutional factors (like historical colonial land revenue system, land reforms post independence etc).

The coefficient on SC population share remain positive (0.12) and significant (at 1% level) in all specifications (Table 2.21). Col 1 is without any controls. From Col 2 to 4, I keep adding demographic, geographic and climatic controls, and the coefficient on SC population share remain stable. In Col 6, instead of District FE, I introduce Sub-District (or Tehsil) FE which is a much smaller geographical administrative unit in India. The coefficient still remains 0.12. The results remain unchanged with other measures of Gini coefficients. (See Appendix Table and 2.22 and 2.22) I run the same regression in each state separately and find the coefficient to remain stable at 0.12 (Appendix Table 2.47)

6. Conclusion

This paper shows that in India the fast pace of economic development, especially after 2000s, has come at the cost of fast rising level of wealth inequality. Further, the rise of wealth is not This increasing wealth inequality is not orthogonal to the caste system. Forward Caste group has captured a larger share of the wealth creation, but the group also has the highest level of within-caste inequality. Scheduled Caste group has the worst condition in terms of owning wealth, and there is negligible convergence. Their plight is mainly due to lack of land ownership, which forms the most important asset in the household wealth basket. The wealth distribution estimates could be improved by national wealth accounting, which is currently not feasible due to lack of national wealth balance sheet. Surveys are not the best source as they miss out more on the financial assets, due to under-reporting and missing the wealthy individuals.

Bibliography

Aggarwal, Ashish (2012). “Implementation of Forest Rights Act, changing forest landscape, and “politics of REDD+” in India”. In: p. 18.

Alvaredo, Facundo, Anthony B Atkinson, and Salvatore Morelli (Dec. 2016). “Top wealth shares in the UK over more than a century”. In: *Working Paper*. WID.World 2017/2, p. 111.

Anand, Ishan and Rishabh Kumar (Feb. 22, 2022). *The sky and the stratosphere: concentrated wealth in India during the ‘lost decade’*.

Anand, Ishan and Anjana Thampi (Dec. 10, 2016). “Recent Trends in Wealth Inequality in India”. In: *Economic and Political Weekly* 51.50.

Banerjee, A. (May 18, 2005). “Top Indian Incomes, 1922-2000”. In: *The World Bank Economic Review* 19.1, pp. 1–20.

Banerjee, Abhijit, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune (May 2013). “Marry for What? Caste and Mate Selection in Modern India”. In: *American Economic Journal: Microeconomics* 5.2, pp. 33–72.

Bauluz, Luis, Yajna Govind, and Filip Novokmet (June 2020). *Global Land Inequality*. halshs-03022360. Publication Title: PSE Working Papers. HAL.

Besley, Timothy, Jessica Leight, Rohini Pande, and Vijayendra Rao (Jan. 1, 2016). “Long-run impacts of land regulation: Evidence from tenancy reform in India”. In: *Journal of Development Economics* 118, pp. 72–87.

Blanchet, Thomas (2016). “Wealth inequality in Europe and in the United States: estimations from surveys, national accounts and wealth rankings”. In: p. 96.

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty (2022). “Generalized Pareto Curves: Theory and Applications”. In: *Review of Income and Wealth* 68.1, pp. 263–288.

Borooah, Vani K. (Aug. 1, 2005). “Caste, Inequality, and Poverty in India”. In: *Review of Development Economics* 9.3, pp. 399–414.

Cancian, Maria and Deborah Reed (1999). “The Impact of Wives’ Earnings on Income Inequality: Issues and Estimates”. In: *Demography* 36.2, pp. 173–184.

Chancel, Lucas and Thomas Piketty (July 1, 2017). "Indian income inequality, 1922-2015: From British Raj to Billionaire Raj?" In: *wid.world* 2017/11, p. 71.

Chatterjee, Aroop, Léo Czajka, and Amory Gethin (Feb. 2, 2022). "Wealth Inequality in South Africa, 1993–2017". In: *The World Bank Economic Review* 36.1, pp. 19–36.

Chavan, Pallavi (2008). *RAS | Debt of Rural Households in India: A Note on the All-India Debt and Investment Survey*. (Visited on 04/09/2018).

Das, Kumudin, K C Das, T K Roy, and P K Tripathy (2011). "Dynamics of inter-religious and inter-caste marriages in India". In: *Princeton Papers*, p. 14.

Deaton, A. (Aug. 11, 2005). "Data and Dogma: The Great Indian Poverty Debate". In: *The World Bank Research Observer* 20.2, pp. 177–199.

Eika, Lasse, Magne Mogstad, and Basit Zafar (July 2014). *Educational Assortative Mating and Household Income Inequality*. Working Paper 20271. National Bureau of Economic Research.

Frémeaux, Nicolas and Arnaud Lefranc (2020). "Assortative Mating and Earnings Inequality in France". In: *Review of Income and Wealth* 66.4, pp. 757–783.

Gothoskar, S. P. (Dec. 1988). "On Some Estimates of Rural Indebtedness". In: *Reserve Bank of India Occasional Papers* 9.4, pp. 299–325.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos (2014). "Marry Your Like: Assortative Mating and Income Inequality". In: *NBER Working Paper*, p. 26.

Jayadev, Arjun, Sripad Motiram, and Vamsi Vakulabharanam (2007). "Patterns of Wealth Disparities in India during the Liberalisation Era". In: *Economic and Political Weekly* 42.38, pp. 3853–3863.

Jayadev, Arjun and Sanjay G. Reddy (2011). "Inequalities and Identities". In: *SSRN Electronic Journal*.

Jayaraj, D. and S. Subramanian (Aug. 1, 2018). "The Distribution of Household Assets in India: 1991–1992 to 2012–2013". In: *Indian Journal of Human Development* 12.2. Publisher: SAGE Publications India, pp. 181–203.

Kessler, Denis and Edward N. Wolff (Sept. 1991). "A Ccomparative Analysis of Household Wealth Patterns in France and the United States". In: *Review of Income and Wealth* 37.3, pp. 249–266.

Kremer, Michael (May 1996). *How Much Does Sorting Increase Inequality?* Working Paper 5566. National Bureau of Economic Research.

Kumar, Rishabh (2016). “Capital and the Hindu Rate of Growth: Top Indian Wealth Holders 1961-1986”. In: *SSRN Electronic Journal*.

Li, Yoe, Martin Rama, Virgilio Galdo, and Florencia M. Pinto (2016). “A Spatial Database for South Asia”. In: World Bank.

Mohanty, B. B. (2001). “Land Distribution among Scheduled Castes and Tribes”. In: *Economic and Political Weekly* 36.40. Publisher: Economic and Political Weekly, pp. 3857–3868.

Narayanan, M. P. (1988). “Debt Versus Equity under Asymmetric Information”. In: *The Journal of Financial and Quantitative Analysis* 23.1, pp. 39–51.

Olivo-Villabrille, Miguel (Oct. 2017). “Assortative Marriages and Household Income Inequality”. In: p. 57.

Piketty, Thomas (2020). *Capital and ideology*. OCLC: 1119745744.

Piketty, Thomas, Li Yang, and Gabriel Zucman (July 2019). “Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015”. In: *American Economic Review* 109.7, pp. 2469–2496.

Qian, Zhenchao (1997). “Breaking the Racial Barriers: Variations in Interracial Marriage Between 1980 and 1990”. In: *Demography* 34.2, pp. 263–276.

Ray, Tridip, Arka Roy Chaudhuri, and Komal Sahai (Sept. 18, 2017). “Whose Education Matters an Analysis of Inter Caste Marriages in India”. In: *Discussion Papers in Economics*, p. 49.

Schwartz, Christine R. (Aug. 2010). “Pathways to Educational Homogamy in Marital and Cohabiting Unions”. In: *Demography* 47.3, pp. 735–753.

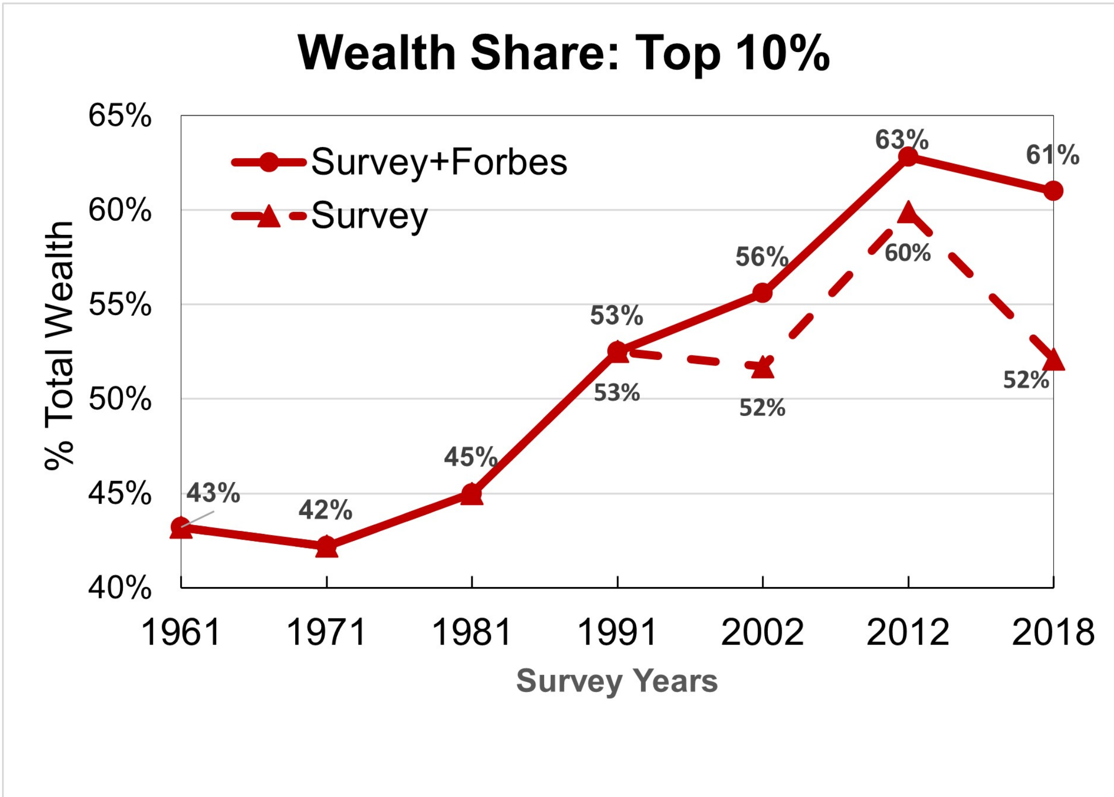
Subramanian, S. and D. Jayaraj (Oct. 16, 2008). “The Distribution of Household Wealth in India”. In: *Personal Wealth from a Global Perspective*. Oxford University Press, pp. 112–133.

Thorat, Sukhadeo (2002). “Oppression and Denial: Dalit Discrimination in the 1990s”. In: *Economic and Political Weekly* 37.6, pp. 572–578.

Zacharias, Ajit and Vamsi Vakulabharanam (Oct. 1, 2011). “Caste Stratification and Wealth Inequality in India”. In: *World Development* 39.10, pp. 1820–1833.

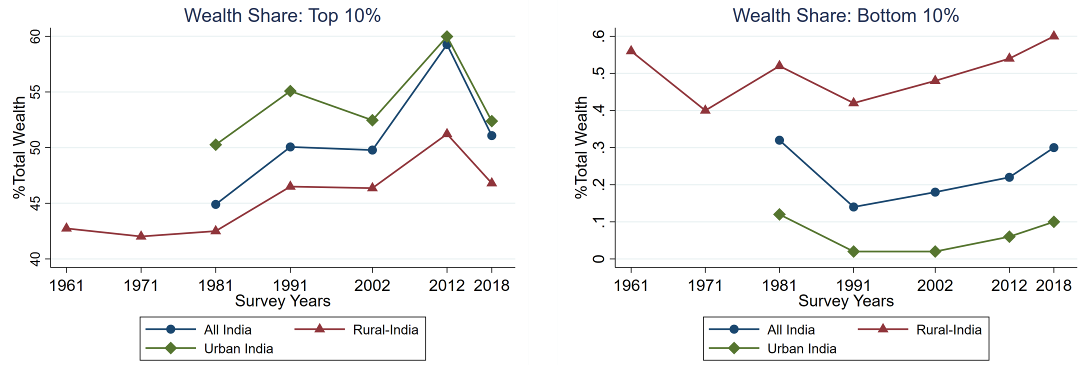
7. Figures

FIGURE 2.1. Share of (Net) Wealth owned by Top 10%



Notes: The figure presents the share of the *net wealth* owned by the Top 10% and Bottom 10% of the population, using NSS-AIDIS surveys and correcting the top distribution of wealth by combining the millionaires' list with surveys from 1961 to 2018. In 2018, the top 10% owned 61% of the total wealth after correcting the survey. The gap between survey and post-correction (survey plus Forbes) has increased in the latest survey round, highlighting the increased non-capturing of the wealthy individuals.

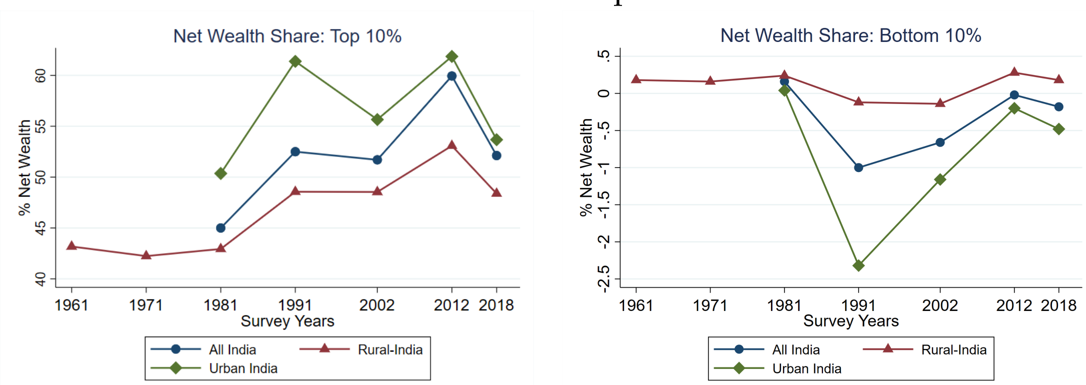
FIGURE 2.2. Gross Wealth Share:Top 10% and Bottom 10%



Top 10% Bottom 10%

Notes: The figure presents the share of the *gross wealth* owned by the top 10% and bottom 10% of the population, using NSS-AIDIS surveys only, from 1961 to 2018 within rural and urban areas separately. The inequality within urban areas is higher than in rural areas. In 2018, the top 10% within urban (rural) locations owned 52% (47%) of the total wealth, in contrast to the bottom 10% within urban (rural) areas holding just 0.1% (0.6%) of the total wealth.

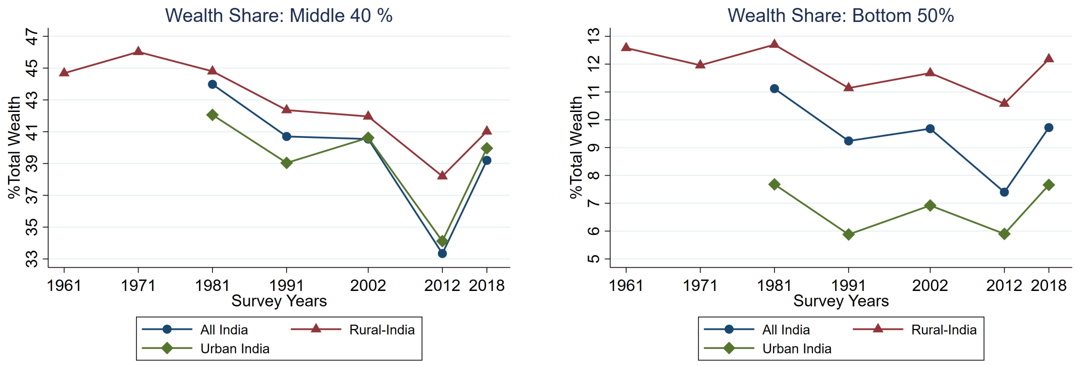
FIGURE 2.3. Net Wealth Share:Top 10% and Bottom 10%



Top 10% Bottom 10%

Notes: The figure presents the share of the *net wealth* owned by the top 10% and bottom 10% of the population, using NSS-AIDIS surveys only, from 1961 to 2018 within rural and urban areas separately. The inequality within urban areas is higher than in rural areas. In 2018, the top 10% within urban (rural) locations owned 53.4% (48.4%) of the net wealth, in contrast to the bottom 10% within urban (rural) areas holding -0.5% (0.2%) of the total (net) wealth.

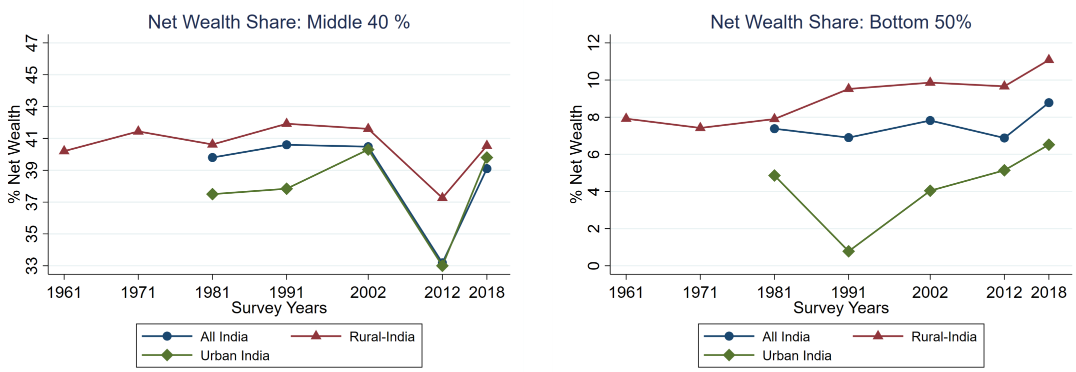
FIGURE 2.4. Gross Wealth Share:Middle 40% and Bottom 50%



Middle 40% Bottom 50%

Notes: The figure presents the share of the *gross wealth* owned by the middle 40% and bottom 50% of the population, using NSS-AIDIS surveys only, from 1961 to 2018 within rural and urban areas separately. In 2018, the middle 40% within urban (rural) locations owned 40% (41%) of the gross wealth, and the bottom 50% within urban (rural) areas owned 8% (12%) of the total (gross) wealth.
The share is on the decline for both groups (i.e. 90% of the population) after 1981.

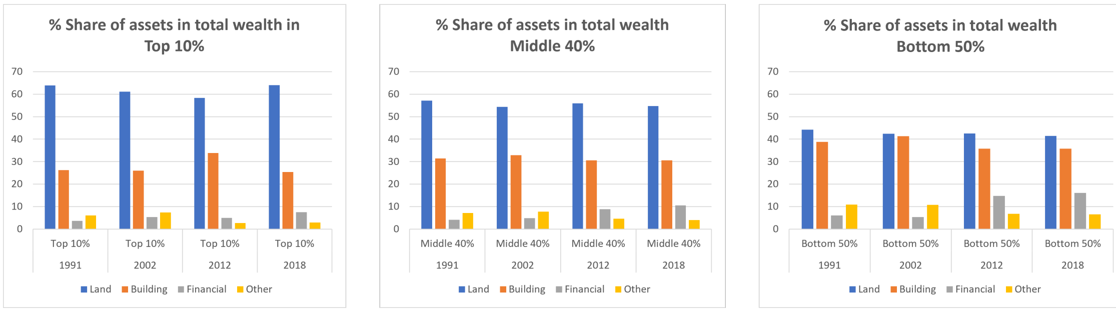
FIGURE 2.5. Net Wealth Share:Middle 40% and Bottom 50%



Middle 40% Bottom 50%

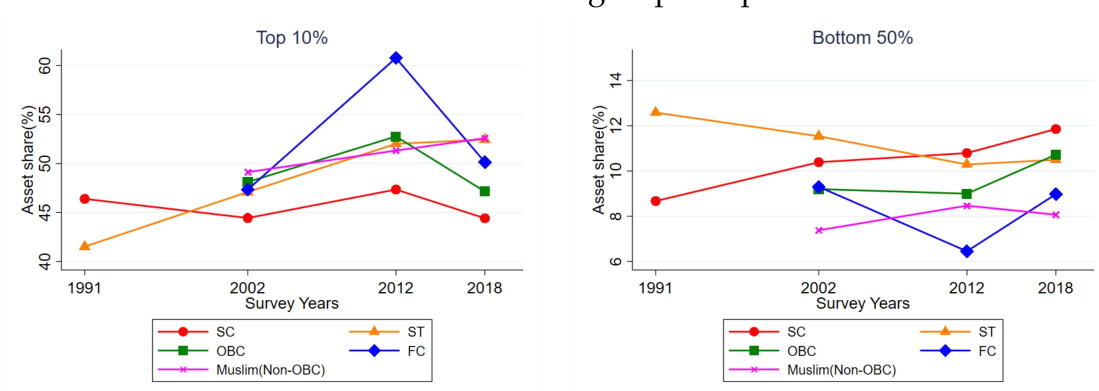
Notes: The figure presents the share of the *net wealth* owned by the middle 40% and bottom 50% of the population, using NSS-AIDIS surveys only, from 1961 to 2018 within rural and urban areas separately. In 2018, the middle 40% within urban (rural) locations owned 39.8% (40.5%) of the net wealth, and the bottom 50% within urban (rural) areas owned 6.5% (11.1%) of the total (net) wealth.

FIGURE 2.6. Land Value Share in Top 10%, Middle 40% and Bottom 50%



Notes: The figure presents the contribution of different assets and their evolution in the wealth basket of the top 10%, middle 40% and bottom 50% of the population, using NSS-AIDIS micro-datasets. Land contributes to 60% of the total wealth for the top 10% population, 55% for the middle 40% population and 40% for the bottom 50% population.

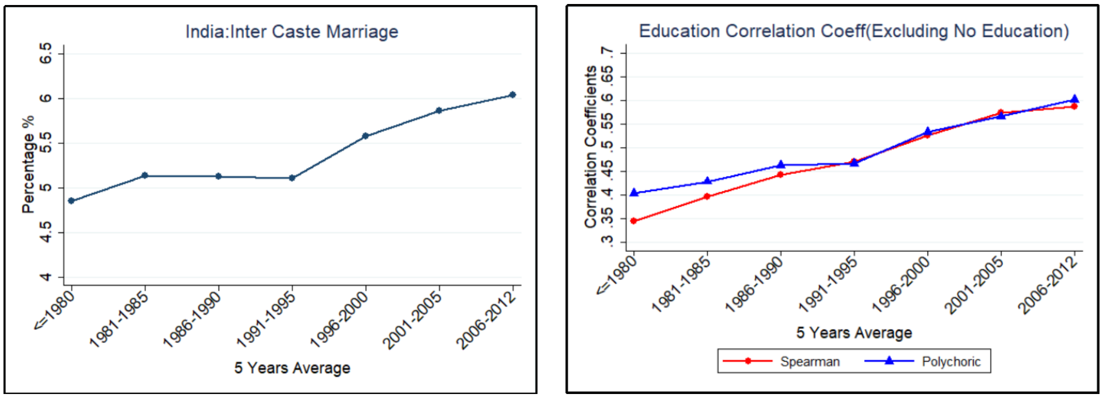
FIGURE 2.7. Wealth Share within caste groups: Top 10% and Bottom 50%



Top 10% Bottom 50%

Notes: The figure presents net wealth share by the top 10% and bottom 50%, using NSS-AIDIS surveys only, from 2002 to 2018. The deciles here are created within each caste group. Within-caste inequality has increased in each caste group, with an increasing (decreasing) wealth share of the top 10% (bottom 50%).

FIGURE 2.8. Caste and Education Assortativity in India



Caste Assortativity Education Assortativity

Notes: The figure presents the share of inter-caste marriages (caste assortativity) on the left and the correlation between husband and wife's education (education assortativity) using the couples' dataset prepared using IHDS 2011 datasets. The x-axis is the marriage cohort. The level of caste assortativity is very high as 95% of all marriages are within caste (the share of inter-caste marriage is 5-6%). It is very persistent, too, as the percentage of inter-caste marriage has increased by 1pp from the pre-1980 married cohort to the 2006-12 married cohort. The correlation between husband and wife's education (excluding couples with zero education) has increased over the years. It has grown from 0.35 in pre-1980 married couples to 0.6 in 2006-12 married couples.

FIGURE 2.9. Rural Data Coverage from Socio-Economic Caste Census 2011

Rural Areas					
Region	**State**	**Total Villages**	**Total HH**	**Total Population**	**Total Adult Population**
South India	Tamil Nadu	15 352	10 084 510	39 632 360	27 387 743
South India	Karnatka	27 623	8 041 410	38 147 745	24 862 517
South India	Kerala	1 459	6 310 072	27 214 909	19 120 587
Central India	Maharashtra	41 899	13 835 228	63 936 212	41 046 798
Central India	Madhya Pradesh	52 994	11 266 931	51 646 466	29 332 367
Central India	Rajasthan	44 168	10 213 736	54 391 054	29 311 620
North India	UP	100 820	25 750 198	162 623 342	88 807 913
North India	Bihar	39 221	17 683 790	98 461 997	51 204 574
North India	Punjab	12 292	3 267 045	16 853 568	10 890 143
East India	West Bengal	38 384	15 633 607	70 745 486	44 367 630
Total		**374 212**	**122 086 527**	**623 653 139**	**366 331 891**

Notes: The table provides the list of the states (with villages, households and population coverage) in the available Socio-Economic Caste Census-2011. I use these ten states covering 623 million population, which is 75% of the total rural population of India. The states are from India's South, Central, North and East parts.

8. Tables

TABLE 2.1. Average Net Wealth per adult (2010 Rs) : 1961-2018

Real Average Net Wealth Per Adult Rs				
	India	**Rural**	**Urban**	**Ratio=Ur/Rur**
1961		59,331		
1971		67,318		
1981	82,307	76,447	95,573	1.25
1991	126,001	107,038	129,047	1.21
2002	172,465	139,081	200,035	1.44
2012	479,207	320,353	732,174	2.29
2018	554,126	422,856	730,198	1.73

Notes: The table presents the real average net wealth per adult from 1961 to 2018, using NSS AIDIS. Separate values for rural and urban areas are also provided. Wholesale Price Index (from World Bank Data) is used to bring the prices from nominal to the 2010 price level.

TABLE 2.2. Top Net Wealth Share

PANEL A: Post Correction					(5)	(6)	(7)
Years					2002	2012	2018
Top 10%					55.6	62.8	61
Top 5%					42.9	50.5	49.8
Top 1%					24.4	30.7	33.3
Top .1%					14.3	18.3	23.1
Top .01%					9.8	12.4	17.5
Middle 40%					36.2	33.7	31.8
Bottom 50%					8.2	6.4	7.2
Bottom 10%					-.05	-.03	-.15

PANEL B: Survey Only	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years	1961	1971	1981	1991	2002	2012	2018
Top 10%	43.2	42.2	45	52.5	51.7	59.9	52.1
Top 5%	30.2	29.3	31.4	38.2	37.3	46.7	38.3
Top 1%	11.8	11.2	12.5	17.0	16	25.4	18.1
Top .1%	2.8	2.4	2.7	4.7	4.3	12	5.5
Top .01%	.6	.5	.5	.9	1	6.9	1.3
Middle 40%	44.5	46.0	44.1	40.6	40.5	33.2	39.1
Bottom 50%	12.3	11.8	10.9	6.9	7.8	6.9	8.8
Bottom 10%	.5	.4	.3	-1.0	-.6	-.03	-.2
Observations	Rural	Rural	All-India	All-India	All-India	All-India	All-India

Notes: The table presents the top *net* wealth share computed from the methodology explained in the Section 3.1. In Panel A, I keep the most conservative estimate with p=.999 for adding the top wealth distribution (from Forbes) to the surveys. Panel B provides the estimate using only NSS-AIDIS surveys. The correction of surveys is possible only for 2002, 2012 and 2018.

TABLE 2.3. Top Gross Wealth Share

PANEL A: Post Correction	(5)	(6)	(7)
Years	2002	2012	2018
Top 10%	55.0	62.1	59.8
Top 5%	42.3	49.8	48.6
Top 1%	23.9	30.1	32.2
Top .1%	13.9	17.8	22.1
Top .01%	9.6	12.0	16.7
Middle 40%	36.4	31.0	32.2
Bottom 50%	8.7	6.9	8.0
Bottom 10%	.16	.21	.24

PANEL B: Survey Only	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years	1961	1971	1981	1991	2002	2012	2018
Top 10%	42.7	42	44.9	50.1	49.7	59.3	51.1
Top 5%	30.1	29.1	31.6	36.1	35.7	46.0	37.4
Top 1%	12	11.2	12.4	15.9	15.1	24.8	17.5
Top .1%	2.8	2.4	2.8	4.3	4.0	11.6	5.2
Top .01%	.7	.5	.5	.8	1.0	6.7	1.3
Middle 40%	44.7	46	44	40.7	40.6	33.3	39.2
Bottom 50%	12.6	12.0	11.1	9.2	9.7	7.4	9.7
Bottom 10%	.5	.4	.3	.2	.2	.02	.3
Observations	Rural	Rural	All-India	All-India	All-India	All-India	All-India

Notes: The table presents the top *gross* wealth share computed from the methodology explained in the Section 3.1. In Panel A, I keep the most conservative estimate with p=.999 for adding the top wealth distribution (from Forbes) to the surveys. Panel B provides the estimate using only NSS-AIDIS surveys. The correction of surveys is possible only for 2002, 2012 and 2018.

TABLE 2.4. Top Net Wealth Share (Survey-based estimates)

PANEL A: Rural	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years	1961	1971	1981	1991	2002	2012	2018
Top 10%	43.2	42.2	42.9	48.6	48.5	53.1	48.4
Top 5%	30.2	29.3	29.6	34.5	34.5	39.8	34.9
Top 1%	11.9	11.2	11.4	14.3	14.2	18.1	15.4
Top .1%	2.8	2.4	2.3	3.5	3.3	4.8	5.0
Top .01%	.9	.5	.4	.6	.7	1.2	.9
Middle 40%	44.5	46.0	45.2	41.9	41.6	37.2	40.5
Bottom 50%	12.3	11.8	11.9	9.5	9.9	9.7	11.1
Bottom 10%	.5	.4	.5	-.1	-.1	.3	.2

PANEL B: Urban	(3)	(4)	(5)	(6)	(7)
Years	1981	1991	2002	2012	2018
Top 10%	50.3	61.4	55.7	61.2	53.4
Top 5%	35.6	46.2	40.5	49.1	39.7
Top 1%	15.0	22.2	18.3	29.6	19.1
Top .1%	3.5	6.3	5.0	15.9	5.6
Top .01%	.7	.9	1.0	9.2	1.9
Middle 40%	42.2	37.9	40.3	33.0	39.8
Bottom 50%	7.4	.8	4.1	5.1	6.5
Bottom 10%	.08	-2.3	-1.1	-.2	-.5

Notes: The table presents the *net* wealth share inequality using surveys (Panel A: Rural; Panel B: Urban). In 1971 and 1961, the survey was conducted only in rural areas. The inequality was stable from 1961-1981. There was a big jump in 1991 when the top 10% wealth share in rural and urban sectors increased by 6pp and 11pp respectively. In 2002, there was s decline in urban inequality. The inequality increased from 2002 to 2012—the 2018 survey hints toward a declining inequality in both rural and urban sectors. The decline in between 2012-18 is partly due to increasingly poor capturing of the rich households.

Table 2.5. Comparison of Wealth Inequality in India with other countries

PANEL A: 2002	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country	India	China	South Africa	Germany	France	USA	UK
Top 1%	24.4	20.6	49.9	26.2	26.3	31	20
Top 10%	55.6	49.2	82.8	57.0	57.9	67.3	55.5
Bottom 50%	8.2	13.6	-.4	3.7	6.0	1.9	5

PANEL B: 2018	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country	India	China	South Africa	Germany	France	USA	UK
Top 1%	33.3	29.8	55.1	29.7	26.0	34.9	21.3
Top 10%	61	67.5	85.7	59.6	58.9	70.7	57.1
Bottom 50%	7.2	6.4	-2.4	3.4	5.0	1.5	4.6

Notes: The table compares the *net* wealth inequality (Panel A: 2002; Panel B: 2018) in India with other countries. The data for other countries comes from the World Inequality Database (wid.world). In 2002, the wealth inequality was higher in India, with more wealth accumulation in the top 10% and top 1%. In both countries, the wealth accumulated by the top 10% population increased, though the rise was higher in China. However, the wealth share of the top 1% population continued to remain higher in India. Among developing countries, South Africa and developed countries, the USA have a higher level of inequality than India.

Table 2.6. Contribution of different assets in total wealth

Contribution of Different Assets in Wealth (excluding Durable HH's)																	
	Rural							**Urban**					**All-India**				
Type of Assets/Year	1961	1971	1981	1991	2002	2012	2018	1981	1991	2002	2012	2018	1981	1991	2002	2012	2018
Land	66.49	69.92	66.92	68.12	65.00	69.84	67.06	38.21	40.04	41.62	45.26	47.93	60.06	59.33	56.59	56.36	58.19
Building	22.52	18.76	22.31	22.69	24.28	20.33	21.59	41.98	44.34	40.92	43.24	36.41	27.00	29.47	30.26	32.89	28.46
Livestock	8.00	6.81	5.39	3.58	4.35	1.55	1.22	0.94	0.48	0.47	0.10	0.10	4.33	2.61	2.95	0.75	0.70
Agri Machinery, Equipments and Transp	2.99	2.83	3.99	3.99	3.99	2.70	2.59	5.90	5.36	6.76	3.17	3.63	4.44	4.42	4.99	2.96	3.07
Financial Assets	-	1.68	1.40	1.62	2.38	5.58	7.55	12.97	9.78	10.23	8.24	11.92	4.16	4.17	5.20	7.04	9.57

Notes: The table presents the contribution of different assets to household wealth using NSS AIDIS surveys. The values are provided for rural (1961-2018), urban(1981-2018) and all-India (1981-2018) separately. Land is the most valuable asset in the household wealth basket, making up 55-60% of the total wealth.

TABLE 2.7. Landlessness and Total Land Area

	% Landless				Total Land Area (mill acres)			
PANEL A: India	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years	1991	2002	2012	2018	1991	2002	2012	2018
Total	13.9	11.2	12.9	13.8	307.76	284.42	257.93	250.61
Top 10%	2.6	1.7	8.5	4.7	95.66	92.52	74.08	78.72
					(31%)	*(33%)*	*(29%)*	*(31%)*
Middle 40%	3.4	2.1	4.3	3.7	168.74	150.49	136.31	125.29
					(55%)	*(53%)*	*(53%)*	*(50%)*
Bottom 50%	24.6	20.3	20.6	23.7	43.36	41.42	47.54	46.6
					(14%)	*(15%)*	*(18%)*	*(19%)*
PANEL B: Rural								
Years	1991	2002	2012	2018	1991	2002	2012	2018
Total	5.6	4.7	4.1	5.9	285.88	262.76	218.37	217.09
Top 10%	.05	.09	.05	.13	84.59	79.27	51.64	58.67
					(30%)	*(30%)*	*(24%)*	*(27%)*
Middle 40%	.9	.4	.4	.8	159.17	143.36	121.78	113.53
					(56%)	*(55%)*	*(56%)*	*(52%)*
Bottom 50%	10.5	8.7	7.1	10.5	42.11	40.13	44.95	44.89
					(15%)	*(15%)*	*(21%)*	*(21%)*
PANEL C: Urban								
Years	1991	2002	2012	2018	1991	2002	2012	2018
Total	38.0	28.6	29.6	29.6	21.88	21.67	39.56	33.52
Top 10%	7.6	4.0	14.2	8.8	11.06	13.25	22.44	20.05
					(51%)	*(61%)*	*(57%)*	*(60%)*
Middle 40%	11.9	6.9	11.4	9.2	9.57	7.13	14.53	11.75
					(44%)	*(33%)*	*(37%)*	*(35%)*
Bottom 50%	64.3	54.6	55.0	57.3	1.25	1.29	2.59	1.71
					(6%)	*(6%)*	*(7%)*	*(5%)*

Notes: The table presents the share of the landless population from Col (1)-(4) and the total land area owned from Col (5)-(8) for the years 1991-2018. The top 10% (rich), middle 40% (middle) and bottom 50% (poor) categorization are based on total wealth at the all-India level. Panel A is for all-India estimates, Panel B for those living in rural areas and Panel C for those living in urban areas.

Table 2.8. Average Real Land Price and Growth Rate

	Average Real Land Price (2010 Rs) in thousands				Annual growth rate Land price			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A: India	1991	2002	2012	2018	1991-2018	1991-2002	2002-12	2012-18
Total	1,793	3,640	9,735	17,646	8.8%	7.3%	10.3%	10.4%
Top 10%	6,506	12,124	33,494	65,183	8.9%	6.4%	10.7%	11.7%
Middle 40%	1,616	3,557	11,029	17,950	9.3%	8.2%	12.0%	8.5%
Bottom 50%	759	1628	2906	5,442	7.6%	7.9%	6.0%	11.0%
PANEL B: Rural	1991	2002	2012	2018	1991-2018	1991-2002	2002-12	2012-18
Total	356	842	1,757	3,266	8.6%	9.0%	7.6 %	10.9%
Top 10%	368	778	2,955	4,858	10.0%	7.8%	14.3%	8.6%
Middle 40%	347	771	1,915	3,488	8.9%	8.3%	9.5%	10.5%
Bottom 50%	363	914	1,492	2,853	7.9%	9.7%	5.0%	11.4%
PANEL C: Urban	1991	2002	2012	2018	1991-2018	1991-2002	2002-12	2012-18
Total	8,115	13,645	29,930	56,793	7.5%	5.3%	8.2%	11.3%
Top 10%	19,283	28,885	56,676	123,596	7.1%	4.1%	7.0%	13.9%
Middle 40%	6,388	11,977	28,635	48,659	7.8%	6.5%	9.1%	9.2%
Bottom 50%	3,561	5,899	10,313	19,364	6.5%	5.2%	5.7%	11.1%

Notes: The table presents the average real land price (2010 Rs) in thousands from Col (1)-(4) and the annual growth rate of the land price from Col (5)-(8). The top 10% (rich), middle 40% (middle) and bottom 50% (poor) categorization are based on total wealth at the all-India level. Panel A is for all-India estimates, Panel B for those living in rural areas and Panel C for those living in urban areas.

TABLE 2.9. Percentage Share of Land Area by Land Type

	Agricultural Land Area (%)				Residential(%)				Other(%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PANEL A: India	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	94.5%	93.5%	88.6%	89.7%	3%	4.1%	4.9%	5.4%	2.8%	2.1%	6.5%	4.8%
Top 10%	96.2%	94.2%	85.9%	89.3%	1.6%	3.4%	3.3%	3.7%	2.3%	1.9%	10.8%	7%
Middle 40%	94.7%	94.4%	90.5%	90.7%	2.8%	3.5%	4.3%	5.2%	2.8%	1.9%	5.2%	4.1%
Bottom 50%	90%	88.9%	87.3%	88%	6.3%	8%	9.1%	9%	3.7%	3%	3.6%	3%
PANEL B: Rural	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	95.1%	94.5%	92.2%	92.5%	2.5%	3.3%	4.4%	4.7%	2.6%	2%	3.5%	2.8%
Top 10%	96.6%	95.8%	93.1%	94.7%	1.4%	2%	2.2%	2%	2.1%	1.6%	4.7%	3.3%
Middle 40%	95.5%	95%	93.1%	92.8%	2.2%	3%	3.6%	4.5%	2.6%	1.9%	3.3%	2.7%
Bottom 50%	90.4%	90.1%	88.5%	88.8%	6.1%	6.9%	8.9%	8.7%	3.5%	2.9%	2.6%	2.6%
PANEL C Urban	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	86.5%	82%	69%	72%	8.5%	14.2%	7.7%	10%	5%	3%	23%	18%
Top 10%	92.8%	84.4%	69.6%	73%	3%	11.6%	5.8%	9%	4%	4%	25%	18%
Middle 40%	80.5%	83.1%	68.5%	70.4%	14.2%	14%	10%	12%	6%	3%	22%	17%
Bottom 50%	75.9%	52%	66%	66%	12%	43.3%	13.3%	17%	12%	5%	21%	16%

Notes: The table presents the share of the total land area owned for three categories of land - agricultural (Col (1)-(4)) , residential (Col (5)-(8)) and Other (Col (9)-(12)). The top 10% (rich), middle 40% (middle) and bottom 50% (poor) categorization are based on total wealth at the all-India level. Panel A is for all-India, Panel B for those living in rural areas and Panel C for those living in urban areas.

TABLE 2.10. Land Area by Land Type

	Agricultural Land Area (Mill Acres)				**Residential (Mill Acres)**				**Other (Mill Acres)**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PANEL A: India	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	290.8	266.1	228.5	224.9	9.1	11.7	12.6	13.6	8.5	5.9	16.8	12
Top 10%	92	87.1	63.7	70.3	1.5	3.1	2.4	2.9	2.2	1.8	8	5.5
Middle 40%	159.8	142.1	123.3	113.6	4.8	5.2	5.8	6.5	4.7	2.9	7.1	5.1
Bottom 50%	39	36.8	41.5	41	2.7	3.3	4.3	4.2	1.6	1.2	1.7	1.4
PANEL B: Rural	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	271.9	248.3	201.2	200.8	7.2	8.6	9.5	10.2	7.3	5.1	7.6	6.2
Top 10%	81.7	75.9	48.1	55.6	1.2	1.6	1.1	1.2	1.7	1.3	2.5	1.9
Middle 40%	152.1	136.2	113.4	105.4	3.4	4.2	4.4	5.1	4.1	2.7	4	3.1
Bottom 50%	38.1	36.2	39.8	39.9	2.6	2.8	4	3.9	1.5	1.2	1.2	1.2
PANEL C: Urban	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	18.9	17.8	27.3	24.1	1.9	3.1	3.1	3.4	1.2	0.7	9.2	5.9
Top 10%	10.3	11.2	15.6	14.7	0.3	1.5	1.3	1.8	0.5	0.5	5.5	3.6
Middle 40%	7.7	5.9	10	8.3	1.4	1	1.4	1.4	0.6	0.2	3.2	2
Bottom 50%	0.9	0.7	1.7	1.1	0.1	0.6	0.3	0.3	0.2	0.1	0.5	0.3

Notes: The table presents the land area (in Million Acres) owned for three categories of land - agricultural (Col (1)-(4)) , residential (Col (5)-(8)) and Other (Col (9)-(12)). The top 10% (rich), middle 40% (middle) and bottom 50% (poor) categorization are based on total wealth at the all-India level. Panel A is for all-India, Panel B for those living in rural areas and Panel C for those living in urban areas.

TABLE 2.11. Average Real Land Price (2010 Rs) by Land Type

	Agricultural Land (thousands per acre)				Residential (thousands per acre)				Other (thousands per acre)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PANEL A: India	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	137	248	1477	1368	2361	4504	15017	24070	1100	4671	7000	7596
Top 10%	284	750	5200	3305	8239	14345	51010	90352	3430	12457	15836	15942
Middle 40%	134	203	1254	1278	2308	4620	17533	23800	725	3405	5699	6252
Bottom 50%	85	123	451	757	932	1958	4075	7655	379	926	2249	3343
top/bottom	3.3	6.1	11.5	4.4	8.8	7.3	12.5	11.8	9.1	13.5	7	4.8
mid/bottom	1.6	1.7	2.8	1.7	2.5	2.4	4.3	3.1	1.9	3.7	2.5	1.9
PANEL B: Rural	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	123	191	595	924	856	1659	3264	6095	338	1159	2236	3157
Top 10%	250	428	1998	2483	1735	2552	7935	13244	772	2359	5932	7124
Middle 40%	123	185	590	940	1016	1930	3948	7207	317	1121	1743	3054
Bottom 50%	81	118	276	463	549	1270	2168	4131	179	509	1317	1604
top/bottom	3.1	3.6	7.2	5.4	3.2	2	3.7	3.2	4.3	4.6	4.5	4.4
mid/bottom	1.5	1.6	2.1	2	1.9	1.5	1.8	1.7	1.8	2.2	1.3	1.9
PANEL C: Urban	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
Total	329	1226	5591	4088	9198	14800	44739	72783	4439	20286	12696	13234
Top 10%	511	2781	9767	5065	20963	31959	84394	164809	8732	33786	21782	21180
Middle 40%	282	524	4355	3392	7233	12897	43884	59298	2674	15610	10548	10497
Bottom 50%	206	298	2177	4327	3932	6093	13925	26183	1819	3879	4155	7189
top/bottom	2.5	9.3	4.5	1.2	5.3	5.2	6.1	6.3	4.8	8.7	5.2	2.9
mid/bottom	1.4	1.8	2	0.8	1.8	2.1	3.2	2.3	1.5	4	2.5	1.5

Notes: The table presents the real land price (in Rs thousands per Acres) owned for three categories of land - Agricultural (Col (1)-(4)), Residential (Col (5)-(8)) and Other (Col (9)-(12)). The top 10% (rich), middle 40% (middle) and bottom 50% (poor) categorization are based on total wealth at the all-India level. Panel A is for all-India, Panel B for those living in rural areas and Panel C for those living in urban areas. The price premium for the top 10% over the bottom 50% is present across all years for all types of land (ratio of top/bottom is greater than 1), and it has increased between 1991-2018.

TABLE 2.12. Population Share by Caste Groups

	Total Population Share				Adult Population Share (> 20yrs)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Caste Groups	1991	2002	2012	2018	1991	2002	2012	2018
ST	8.74	8.01	9.29	10.1	8.42	7.6	8.78	9.43
SC	18.4	19.71	18.78	19.62	18.01	18.81	17.89	18.7
OBC		40.28	43.57	44.51		39.72	42.85	44.06
FC		22.56	20.48	19.11		24.87	22.8	21.29
Muslim		7.56	6.26	5.21		6.75	5.77	4.8
Non-Hindu-Muslim		1.89	1.62	1.45		2.24	1.91	1.72
Others	72.86				73.57			
Total Pop (mill)	810	986	1,058	1118	427	546	638	716

Notes: The table presents the total population share (Col (1)-(4)) and adult (> 20 years) population share (Col (5)-(8)) where an adult is above 20 years old using NSS-AIDIS surveys (1991, 2002, 2012 and 2018). The population share of SC/ST/OBC is increasing, whereas other caste groups are declining.

TABLE 2.13. Difference between Wealth Share and Population Share

	India				Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Caste	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
ST	-4.45	-3.9	-5.52	-4.88	-5.13	-4.34	-5.95	-5.7	-1.52	-1.16	-1.63	-0.89
SC	-10.07	-10.96	-11.34	-9.81	-10.47	-11.12	-10.36	-9.35	-7.84	-8.97	-9.76	-8.65
OBC		-4.2	-7.16	-4.61		-1.06	0.36	0.52		-8.26	-12.12	-10.07
FC		15.78	20.65	16.77		12.49	11.02	11.41		18.01	22.52	18.71
Muslim		-2.87	-2.52	-1.29		-2.26	-1.97	-0.92		-5.01	-3.71	-2.21
Others		6.13	5.91	3.81		6.29	6.91	4.06		5.39	4.69	3.11

Notes: The table presents the difference between wealth owned share and population share by caste groups, using NSS-AIDIS surveys (1991, 2002, 2012 and 2018) for rural, urban and all-India separately. A positive value for a caste group implies that it owns more wealth share than its population size. E.g. In 2018, FC held 16.77pp more wealth than its population share. SC, ST, and OBC's wealth shares are lower than their population share.

TABLE 2.14. Representational Inequality

	2002			2012			2018		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Caste	bottom50	middle40	top10	bottom50	middle40	top10	bottom50	middle40	top10
ST	0.31	-0.21	-0.71	0.39	-0.30	-0.74	0.40	-0.34	-0.61
SC	0.36	-0.26	-0.75	0.31	-0.21	-0.74	0.28	-0.18	-0.68
OBC	-0.03	0.07	-0.15	-0.05	0.09	-0.15	-0.05	0.09	-0.14
FC	-0.32	0.20	0.82	-0.29	0.14	0.92	-0.32	0.14	1.01
Muslim	0.17	-0.12	-0.40	0.11	-0.06	-0.30	0.12	-0.11	-0.18
Others	-0.59	-0.15	3.55	-0.57	-0.17	3.54	-0.50	0.04	2.33

Notes: The table presents the representational inequality by different caste groups. The numbers are comparable across caste and years as they are standardised. A positive value implies more population share in that class group than their population share in overall population.

TABLE 2.15. Assortative Matching: Education

	All couples								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A	All-India	Rural	Urban	Brahmin (FC1)	Other FC FC2	OBC	Dalit (SC)	Adivasi (ST)	Muslim
Polychoric	0.64	0.58	0.66	0.61	0.65	0.60	0.58	0.60	0.6
C.I.	[.64,.64]	[.58,.59]	[.66,.66]	[.60,.61]	[.65,.66]	[.59,.60]	[.57,.58]	[.59,.60]	[.59,.60]
Spearman	0.63	0.57	0.66	0.63	0.65	0.57	0.56	0.59	0.56
C.I.	[.62,.64]	[.56,.58]	[.65,.67]	[.60,.66]	[.64,.67]	[.56,.59]	[.54,.57]	[.57,.61]	[.54,.58]
Obs	34174	22,766	11,408	1,704	5,741	11,511	7,359	2,826	4,095

	Excluding Cases when both husband and wife have 0 education								
Panel B	All-India	Rural	Urban	Brahmin (FC1)	Other FC FC2	OBC	Dalit (SC)	Adivasi (ST)	Muslim
Polychoric	0.51	0.40	0.60	0.61	0.61	0.46	0.39	0.37	0.38
C.I.	[.51,.51]	[.39,.41]	[.60,.61]	[.60,.63]	[.61,.62]	[.45,.47]	[.38,.40]	[.35,.39]	[.36,.40]
Spearman	0.50	0.40	0.6	0.61	0.60	0.44	0.38	0.39	0.36
C.I.	[.495,.512]	[.39,.41]	[.59,.61]	[.58,.64]	[.59,.62]	[.42,.46]	[.36,.4]	[.35,.43]	[.33,.39]
Obs	28,541	17,992	10,549	1,669	5412	9,772	5,727	1,953	3,111

Notes: The table presents Polychoric and Spearman correlation coefficients (C.I. with 95% confidence interval) for husband and wife education (Panel A: full sample; Panel B; excluding zero education cases) from the IHDS 2011 dataset. The last row of each Panel is the number of observations used in the computation. Col (1) is for all-India, and Col(2) and Col (3) present for rural and urban households separately. Col(4)-Col(9) is for different caste groups. The correlation coefficients in Panel A for different caste groups are very close (0.6) due to several couples with zero education. In Panel B, the coefficients are highest for the F.C. (0.6), followed by OBC (0.46). For SC, ST and Muslims, it is close to 0.38.

TABLE 2.16. Assortative Matching: Occupation and Wage Earnings

	Occupation								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A	All -India	Rural	Urban	Brahmin (FC1)	Other FC FC2	OBC	Dalit (SC)	Adivasi (ST)	Muslim
Polychoric	0.43	0.39	0.54	0.46	0.49	0.44	0.32	0.49	0.39
C.I.	[.42,.44]	[.38,.40]	[.52,.56]	[.40,.51]	[.47,.52]	[.42,.45]	[.30,.34]	[.47,.52]	[.35,.43]
Spearman	0.35	0.33	0.45	0.45	0.41	0.33	0.29	0.41	0.35
C.I.	[.34,.37]	[.31,.35]	[.41,.48]	[.34,.54]	[.36,.45]	[.31,.36]	[.25,.32]	[.36,.45]	[.28,.41]
Obs.	10,843	8,732	2,111	258	1,208	3,911	3,023	1,592	655

	Annual Wage Earnings								
Panel B	All -India	Rural	Urban	Brahmin (FC1)	Other FC FC2	OBC	Dalit (SC)	Adivasi (ST)	Muslim
Pearson	0.70	0.63	0.69	0.76	0.78	0.60	0.55	0.75	0.59
C.I.	[.70,.71]	[.62,.64]	[.69,.70]	[.76,.77]	[.77,.78]	[.59,.61]	[.54,.56]	[.74,.75]	[.58,.60]
Spearman	0.52	0.45	0.55	0.76	0.56	0.52	0.42	0.63	0.34
C.I.	[.51,.54]	[.43,.472]	[.51,.58]	[.69,.820]	[.51,.61]	[.5,.55]	[.38,.45]	[.6,.66]	[.26,.42]
Obs	8,006	6,555	1,451	153	685	2,772	2,545	1,275	442

Notes: The table presents Polychoric and Spearman correlation coefficients (C.I. with 95% confidence interval) for husband and wife's occupation (Panel A) and annual wage earnings (Panel B, both partners have non-zero wages.) from the IHDS 2011 dataset. The last row of each Panel is the number of observations used in the computation. Col (1) is for all-India, and Col(2) and Col (3) present for rural and urban households separately. Col(4)-Col(9) is for different caste groups.

TABLE 2.17. Difference between Land Area Share and Population Share

	India				Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Caste	1991	2002	2012	2018	1991	2002	2012	2018	1991	2002	2012	2018
ST	2.05	2.26	2.78	2.51	0.81	1.12	1.33	0.74	-0.44	-0.16	1.56	2.04
SC	-8.38	-10.77	-9.24	-8.71	-9.86	-12.04	-10.39	-9.54	-7.57	-9.68	-9.13	-9.95
OBC		1.62	-0.14	1.15		0.78	0.18	1.21		-3.73	-5.19	-1.32
FC		9.98	8.46	6.98		12.16	10.53	8.68		17.12	12.83	13.07
Muslim		-3.89	-3.38	-2.84		-3.2	-2.99	-2.33		-3.61	-3.63	-4.1
Others		0.79	1.53	0.91		1.17	1.34	1.25		0.06	3.56	0.27

Notes: The table presents the representational land area inequality by caste groups, using NSS-AIDIS surveys (1991, 2002, 2012 and 2018) for rural, urban and all-India separately. A positive value for a caste group implies that it owns more land (area) share than its population share. E.g. In 2018, FC owned 6.98pp more land area than its population share and SC owned 8.71pp less land area than its population share.

TABLE 2.18. Village-Level Average of Households Dependency on Landowning Class

States	(1) Total HH	(2) Share of Agriculture Dependence	(3) # of HH's Dependent on Agriculture	(4) Landlessness (%)	(5) Landowning	(6) Dependency Ratio
TN	635	0.532	338	69	197	0.72
Karnatka	298	0.626	187	41	176	0.06
Kerala	4090	0.223	912	71	1186	-0.23
Maharashtra	327	0.682	223	50	164	0.36
Rajasthan	224	0.729	163	36	143	0.14
UP	274	0.706	193	43	156	0.24
Bihar	455	0.738	336	61	178	0.89
Punjab	274	0.518	142	76	66	1.16
West Bengal	355	0.653	232	58	149	0.55
MP	220	0.778	171	53	103	0.66

Notes: The table presents average of village-level characteristics computed from the SECC-2011. Col(1) is average number of HHs in village. Col(2) is the share of HHs dependent on agriculture. Col(3) is the total number of HHs dependent on agriculture. Col(4) is the share of population without agricultural land. Col(5) is the number of households owning land. Col(6) is the dependency ratio, which is ratio of HHs without land (dependent on agriculture) per land-owning household.

TABLE 2.19. Village-Level Average Gini Coefficient and Agricultural Land Area Share

	Gini			Top Share		
States	(1) HH Level	(2) Individual Level	(3) Adult Level	(4) Top 10%	(5) Top 5%	(6) Top 1%
Rajasthan	0,618	0,614	0,596	53.72	38.92	19.46
Karnatka	0,644	0,634	0,626	56.06	41.05	20.65
UP	0,673	0,675	0,665	60.71	45.40	23.43
Maharashtra	0,71	0,703	0,688	60.22	43.84	20.49
MP	0,717	0,723	0,702	63.07	46.70	23.48
West Bengal	0,789	0,778	0,772	72.01	57.52	33.14
TN	0,826	0,822	0,813	77.23	61.06	31.56
Bihar	0,827	0,825	0,803	78.83	65.44	40.14
Punjab	0,844	0,84	0,829	80.13	68.04	40.61
Kerala	0,924	0,924	0,921	88.37	79.00	57.01

Notes: The table presents average of village-level land inequality measure (Gini and top share) computed from the SECC-2011. The states are arranged in ascending order of the land inequality measure. Col(1)-Col(3) is gini coefficient computed at HH, individual and adult level, respectively. Col(4)-Col(6) is the land share owned by the top 10%, top 5% and top 1% households.

TABLE 2.20. Village-Level Average Characteristics (2011)

States	(1) Tamil Nadu	(2) Karnatka	(3) Kerala	(4) Maha- -rashtra	(5) Raja- -sthan	(6) Uttar Pradesh	(7) Bihar Bihar	(8) Punjab	(9) West Bengal	(10) Madhya Pradesh
# of HHs	635.0	298.2	4,090	327.0	224.9	274.2	455.9	274.6	355.5	220.1
Population	2,481	1,403	17,222	1,522	1,219	1,661	2,497	1,419	1,608	1,041
Pop density	482.7	359.7	918.4	289.2	258.6	1,203	1,420	441.4	957.7	255.1
Sex ratio	0.499	0.495	0.518	0.491	0.481	0.480	0.481	0.477	0.489	0.483
SC share	0.281	0.209	0.105	0.113	0.175	0.238	0.186	0.358	0.275	0.146
ST share	0.028	0.087	0.031	0.20	0.20	0.007	0.025	0.00	0.151	0.312
Literacy	0.651	0.615	0.830	0.661	0.501	0.557	0.494	0.654	0.628	0.529
Working sh	0.519	0.521	0.374	0.524	0.495	0.339	0.347	0.351	0.415	0.481
Agri share	0.532	0.626	0.223	0.682	0.729	0.706	0.738	0.518	0.653	0.778
Area	1,705	1,640	7,620	1,731	1,896	584.2	580.4	957.2	501.5	1,225

Notes: The table presents average of village-level demographic characteristics computed from the Census-2011.

TABLE 2.21. Land Inequality (Gini HH) - SC and ST population share

VARIABLES	(1) gini hh	(2) gini hh	(3) gini hh	(4) gini hh	(5) gini hh	(6) gini hh
SC pop share	0.131***	0.122***	0.119***	0.119***	0.119***	0.118***
	(0.0127)	(0.00778)	(0.00647)	(0.00652)	(0.00651)	(0.00545)
ST pop share	-0.0245	-0.00969	0.00232	0.00454	0.00710	0.0106
	(0.0165)	(0.00870)	(0.00772)	(0.00766)	(0.00768)	(0.00649)
Total Population			1.44e-05***	1.42e-05***	1.42e-05***	1.40e-05***
			(7.92e-07)	(7.91e-07)	(7.86e-07)	(7.56e-07)
Distance from Town (in km)				-0.000257***	-0.000257***	-0.000253***
				(4.80e-05)	(4.57e-05)	(3.21e-05)
Observations	352,633	352,633	352,196	343,482	343,438	343,437
R-squared	0.029	0.311	0.379	0.384	0.386	0.435
Mean	0.71	0.71	0.71	0.71	0.71	0.71
District FE	No	Yes	Yes	Yes	Yes	No
Sub District FE	No	No	No	No	No	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes
Geographic Controls	No	No	No	Yes	Yes	Yes
Climatic Controls	No	No	No	No	Yes	Yes
Num_cluster	337	337	337	337	337	337

Notes: The table reports OLS estimations based on a village-level sample for the ten states. Robust standard errors in parentheses are clustered at the district level. The dependent variable is the Gini coefficient at the village level computed based on the HH owned agricultural land. All estimations include SC and ST population share in the village. Column 1 is without controls; Column 2 adds District fixed effect; Column 3 further adds Demographic controls (total population, population density, literacy rate, working population share, and agricultural population share); Column 4 further adds Geographic controls (distance from the nearest statutory town in km, total village land, and forest density); Column 5 further adds Climatic controls (elevation, roughness, decadal average temperature 2001-11 and decadal average precipitation 2001-11); Column 6 replaces District FE with Sub-district FE while keeping all the controls. Demographic and Geographic controls come from the Census 2011.

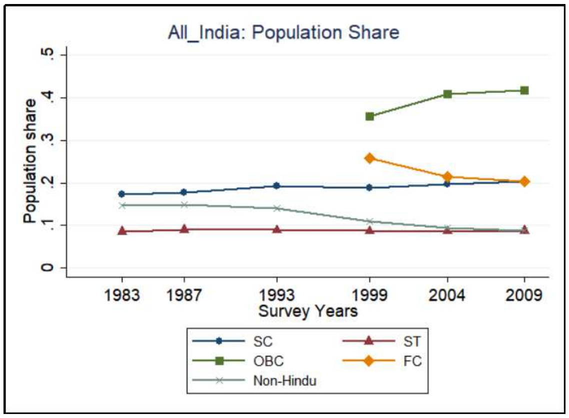
TABLE 2.22. Land Inequality (Gini Coeff Adult Pop) - SC and ST population share

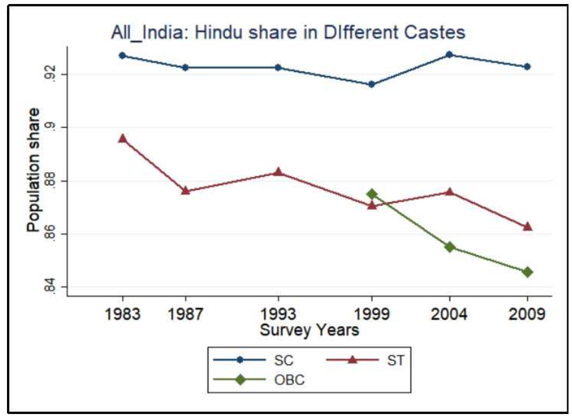
VARIABLES	(1) gini adult	(2) gini adult	(3) gini adult	(4) gini adult	(5) gini adult	(6) gini adult
sc_ratio_2011	0.136***	0.116***	0.113***	0.112***	0.113***	0.114***
	(0.0132)	(0.00798)	(0.00639)	(0.00649)	(0.00648)	(0.00537)
st_ratio_2011	-0.00670	0.00974	0.0180**	0.0188***	0.0206***	0.0209***
	(0.0153)	(0.00826)	(0.00721)	(0.00706)	(0.00702)	(0.00599)
TOT_P			1.45e-05***	1.43e-05***	1.43e-05***	1.42e-05***
			(7.58e-07)	(7.57e-07)	(7.51e-07)	(7.34e-07)
DIST_STAT_TOWN_2011				-0.000196***	-0.000200***	-0.000216***
				(4.60e-05)	(4.36e-05)	(2.93e-05)
Constant	0.667***	0.669***	0.867***	0.872***	0.935***	0.828***
	(0.00739)	(0.00207)	(0.0104)	(0.0106)	(0.0833)	(0.0725)
Observations	322,525	322,525	322,120	314,043	313,999	313,998
R-squared	0.028	0.315	0.390	0.395	0.396	0.441
Mean	0.69	0.69	0.69	0.69	0.69	0.69
District FE	No	Yes	Yes	Yes	Yes	No
SubDistrict FE	No	No	No	No	No	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes
Geographical Controls	No	No	No	Yes	Yes	Yes
Climatic Controls	No	No	No	No	Yes	Yes
Num_cluster	314	314	314	314	314	314

Notes: The table reports OLS estimations based on a village-level sample for the ten states. Robust standard errors in parentheses are clustered at the district level. The dependent variable is the Gini coefficient at the village level computed based on the HH owned agricultural land equally split within household adult members (age>20 years). All estimations include SC and ST population share in the village. Column 1 is without controls; Column 2 adds District fixed effect; Column 3 further adds Demographic controls (total population, population density, literacy rate, working population share, and agricultural population share); Column 4 further adds Geographic controls (distance from the nearest statutory town in km, total village land, and forest density); Column 5 further adds Climatic controls (elevation, roughness, decadal average temperature 2001-11 and decadal average precipitation 2001-11); Column 6 replaces District FE with Sub-district FE while keeping all the controls. Demographic and Geographic controls come from the Census 2011.

D. Appendix Figures

FIGURE 2.10. Population Share of Caste Groups 1983-2009

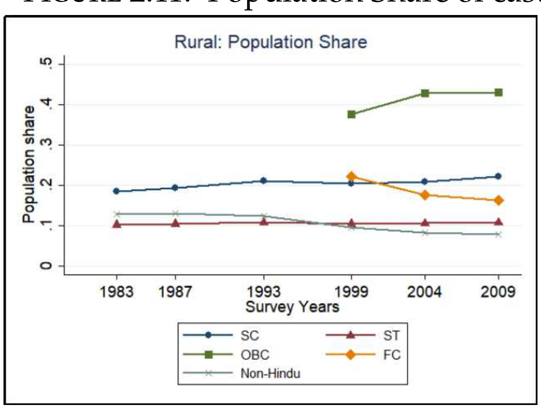


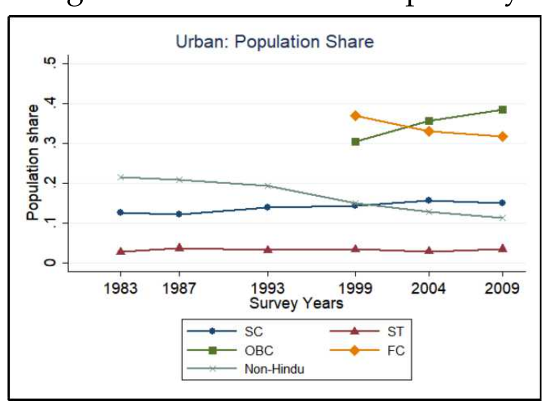


Proportion of castes Share of Hindus

Notes: The left figure presents the population share of different caste groups using NSS consumption surveys (1983, 87, 93, 99, 2004 and 2009). There is an increasing trend in SC, ST and OBC caste groups and a declining trend in OBC and FC caste groups. The right figure shows the population share of Hindus within SC, ST and OBC groups, which indicates the presence of other religions within each caste category. The Hindu share is declining in ST and OBC due to the inclusion of other religions.

FIGURE 2.11. Population Share of caste categories in Rural Urban separately

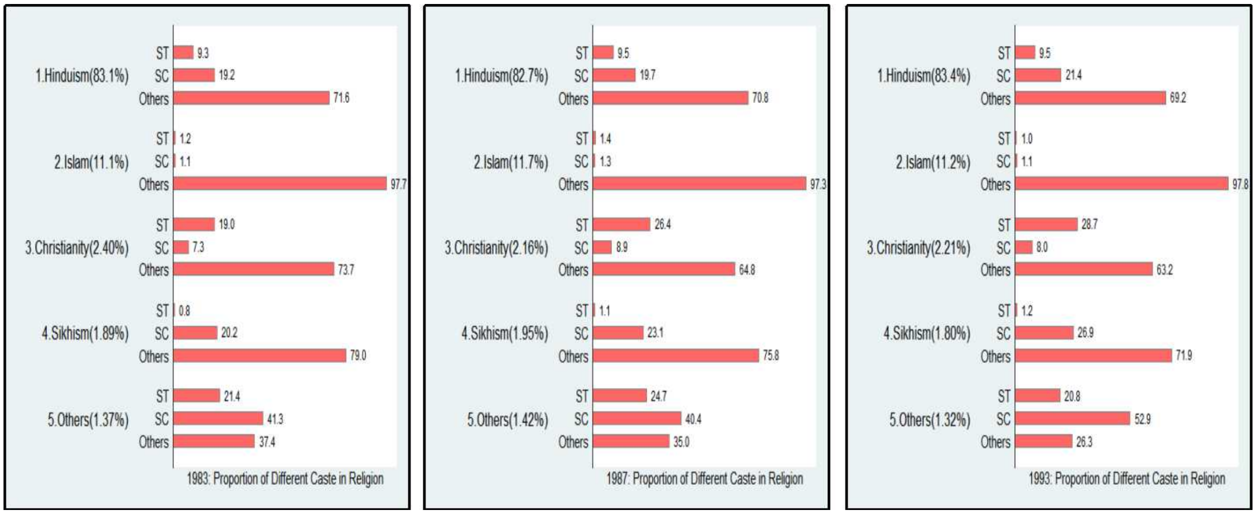




Rural: SC/ST/OBC/FC Urban: SC/ST/OBC/FC

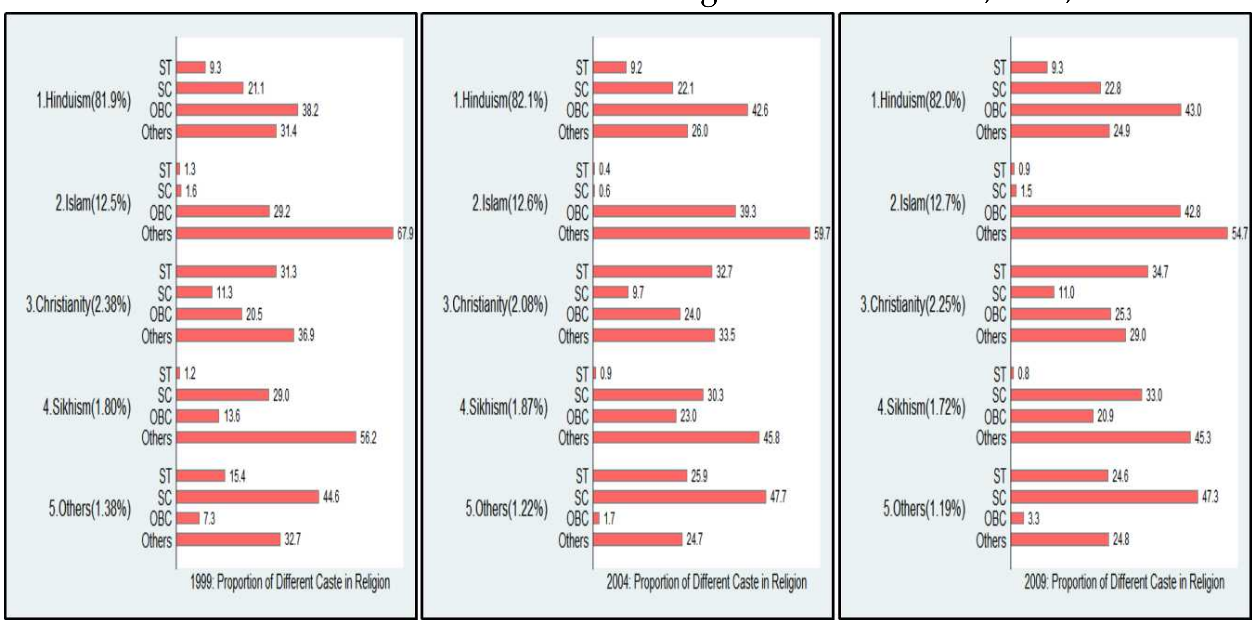
Notes: The left figure presents the population share of different caste groups - SC, ST, OBC, FC and Non-Hindu from NSS consumption surveys (1983, 1987, 1993, 2004 and 2009) living in rural areas. The right figure shows the same for an urban area. In urban areas, FC presence is higher than their population share, whereas ST presence is lower than their population share.

FIGURE 2.12. Caste Share in Different Religion from NSS-1883 1887 1993



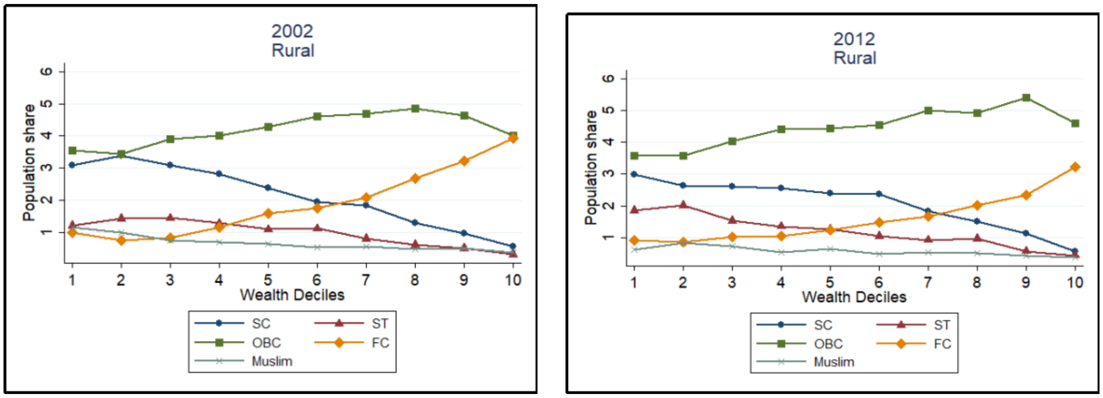
Notes: The figure presents the population share of different caste groups - SC, ST, and others from the NSS-Consumption 1983, 1987 and 1993 survey datasets.

FIGURE 2.13. Caste Share in Different Religion from NSS-1999, 2004, 2009



Notes: The figure presents the population share of different caste groups - SC, ST, and others from the NSS-Consumption 1999, 2004 and 2009 survey datasets.

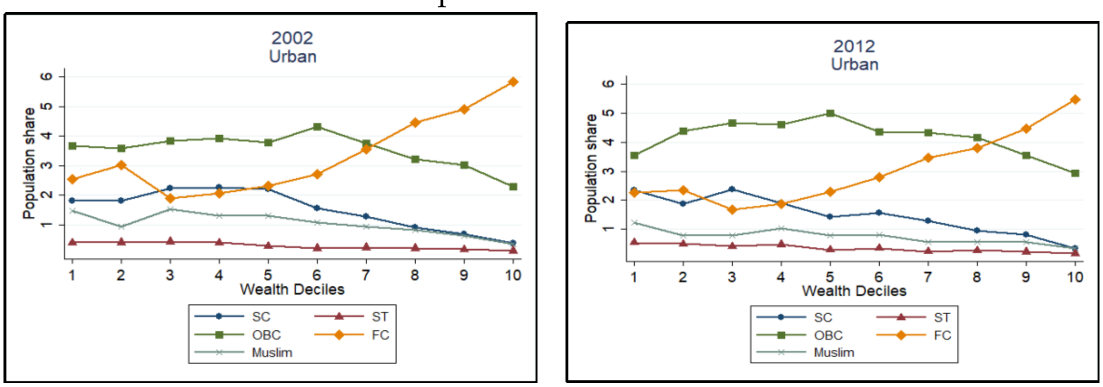
FIGURE 2.14. Proportion of different castes-Rural



2002 2012

Notes: The figure presents the the proportion of population in different (gross) wealth deciles using NSS-AIDIS datasets in rural areas by different caste groups. "Other" population is not presented here as their population share is very small.

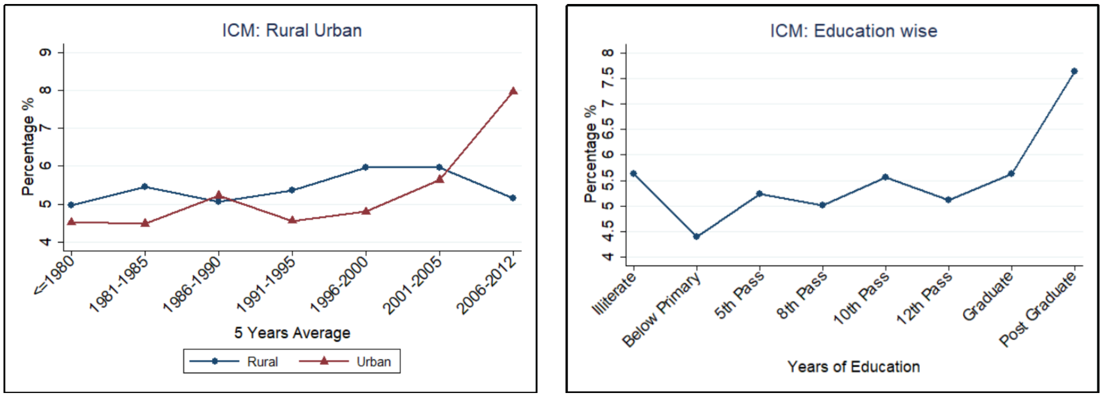
FIGURE 2.15. Proportion of different castes-Urban



2002 2012

Notes: The figure presents the the proportion of population in different (gross) wealth deciles using NSS-AIDIS datasets in urban areas by different caste groups. "Other" population is not presented here as their population share is very small.

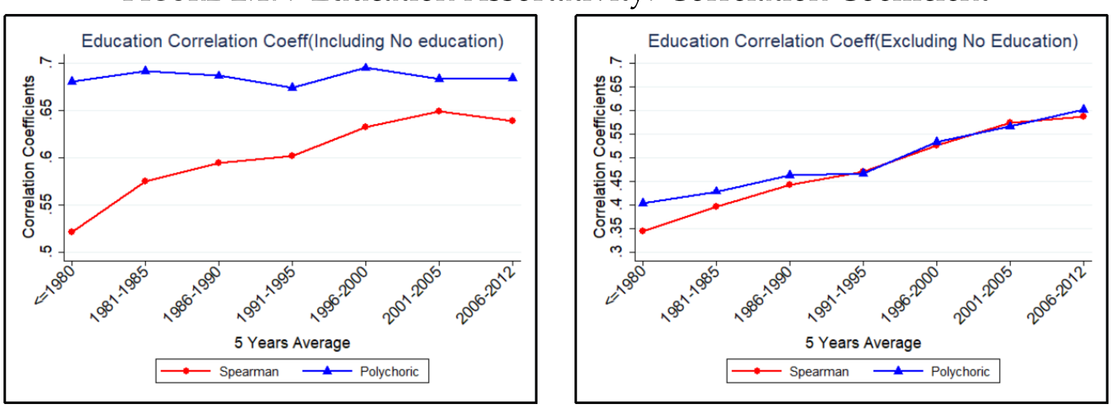
FIGURE 2.16. Share of Inter-Caste Marriages by cohort



ICM: Rural Urban ICM: Education of women

Notes: The figure plots the share of inter-caste marriages in rural and urban areas by marriage cohort (left part) and by the years of education (right part) using the couples' dataset prepared using IHDS 2011 datasets. ICM is derived from the (survey) question asking eligible ever-married women- "Is your family the same caste as your natal family?" A low share of inter-caste marriages highlights a high level of caste endogamy in society.

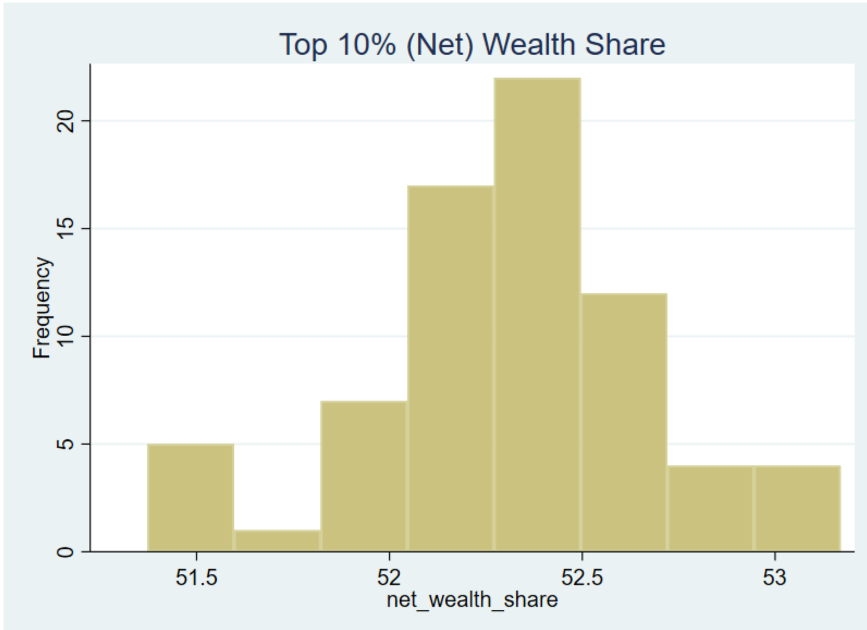
FIGURE 2.17. Education Assortativity: Correlation Coefficient



Including zero education Excluding zero education

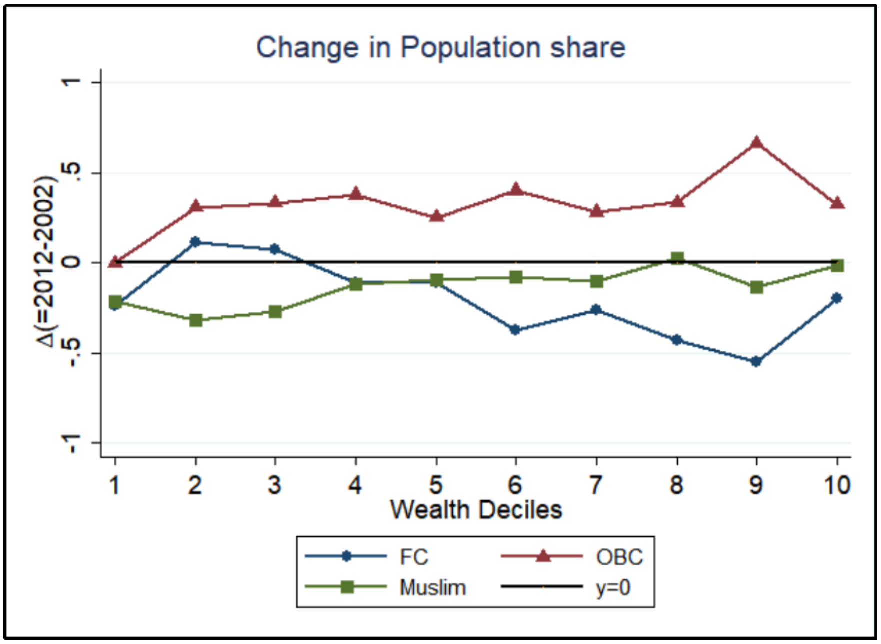
Notes: The figure plots the Spearman and Polychoric correlation coefficients by marriage cohort for the total sample (left part) and excluding the couples with no education (right part) using the couples' dataset prepared using IHDS 2011 datasets. The marriage cohort cut-off is to keep enough and similar sample size in each created category. The level of education assortativity is increasing in society. The Spearman coefficient increased from 0.35 in pre-1980 married couples to 0.6 in the 2006-12 married couples (having non-zero education).

FIGURE 2.18. Top 10% (Net) Wealth Share Distribution in 2018 (using 2012 sampling methodology)



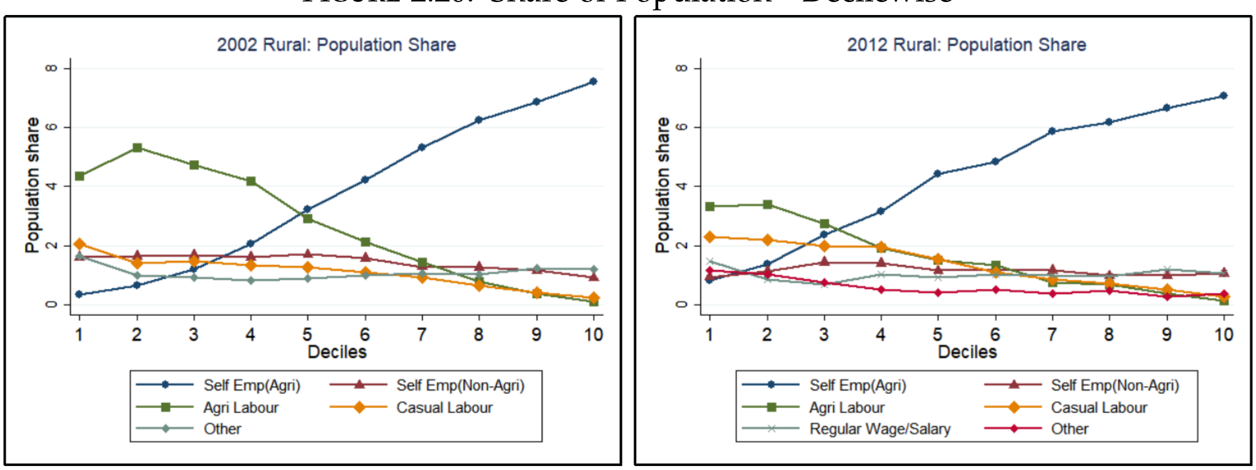
Notes: This figure presents the distribution of the top 10% net wealth share for 2018, following the 2012 sampling strategy. The mean of the wealth share is 52.33% (s.d.=36)

FIGURE 2.19. Change in population share between 2002 and 2012



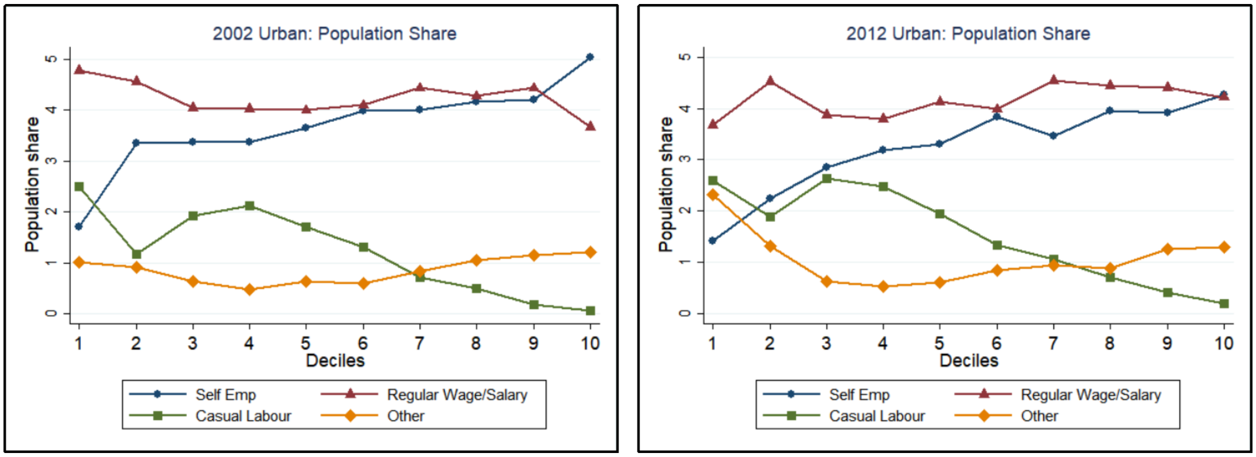
Notes: This figure presents a change in the population share of FC, OBC and Muslim groups (y-axis) in percentage points between 2002-2012 and across wealth deciles (x-axis) using NSS AIDIS surveys. The OBC population has seen a positive change in population vis-a-vis a drop in population share in FC and Muslim groups.

FIGURE 2.20. Share of Population - Decilewise



Notes: The figure presents the population share of different employment types in different wealth deciles for the rural sector, using NSS-AIDIS surveys for 2002 and 2012. In rural areas, the population declaring to be Self Employed in Agriculture is concentrated in higher deciles, and Agricultural Labours are concentrated in the lower deciles.

FIGURE 2.21. Share of Population - Decilewise



Notes: The figure presents the population share of different employment types in different wealth deciles for the urban sector, using NSS-AIDIS surveys for 2002 and 2012. In urban areas, the population declaring to be Self Employed in Agriculture is concentrated in higher deciles, and Agricultural Labours are concentrated in the lower deciles.

E. Appendix Tables

TABLE 2.23. Sample Size in NSS-AIDIS Surveys

	FSU's Surveyed			Households (HH's) Surveyed			HH/FSU			Rural/Urban
Year	Total	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	(HH)
1961										
1971	12452	12452	-	99616	99616	-	8	8	-	-
1981	12887	7718	5169	92122	61157	30965	8	8		2.0
1991	6650	4231	2419	57031	36425	20606	9	9	9	1.8
2002	10309	6552	3757	143285	91192	52093	14	14	14	1.8
2012	8036	4529	3507	110800	62135	48665	14	14	14	1.3
2018	9935	5940	3995	116461	69455	47006	12	12	12	1.5

Notes: The table presents the total FSU's (First Stage Units) and Households (HH's) surveyed in rural and urban sectors separately. Households per FSU shows how many households were sampled within each FSU. The last column shows how many rural households were sampled on each urban household. The survey years from 1991 onwards only provides the information on the sample collected by the NSSO central office. The micro files do not provide information on the sample collected by the state agencies.

TABLE 2.24. Assets and Liabilities Covered in NSS-AIDIS

<table>
<tr><th>Assets/Liabilities</th><th>1961</th><th>1971</th><th>1981</th><th>1991</th><th>2002</th><th>2012</th><th>2018</th></tr>
<tr><td>Physical Assets</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Land (including vacant house site)</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Building</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Livestock</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Agricultural Implements
and Machinery</td><td rowspan="2">✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Non-Farm
Business Equipment</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Transport Equipment</td><td></td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Durable Assets</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>×</td><td>×</td></tr>
<tr><td>Financial Assets</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>All kinds of Shares/Debentures/
Mutual Funds etc</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Deposits (in companies,
banks, post-offices etc)</td><td></td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Dues Receivable Cash</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Dues Receivable Kind</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Liabilities</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Dues Payable Cash</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Dues Payable Kind</td><td></td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
</table>

Notes: The table presents the broad categories of assets and liabilities captured in different NSS AIDIS survey rounds. The last two surveys did not capture durable assets due to the concern about their valuation.

TABLE 2.25. Items covered under Inventory of Assets in 2018-19

Land (5)	**Livestock (15)**	**Agricultural Implements and Machinery (12)**	**Building (9)**	**Non-Farm Business Equipments (18)**	**Transport Equipments (7)**	**Ownership of Shares, Debentures (4)**	**Other Financial Assets and Loans (17)**
Crop Area Irrigated and unirrigated	Cattle Cross bred and non-descript	Power tiller	Residential Building (used as dwelling)	Handloom (semi-automatic and power looms), Gins pressing and balling equipment	Cart (hand/ animal driven)	Cooperative Society	Cash in Hand
Other area for agricultural farm business	Buffalow	Crop Harvester (power-driven) /Combined Harvester	Other residential building within village/town	Reeds, bobbins and other accessories used in spinning and weaving and other tailoring equipment	Bicycles	Company	Amount in Current Bank Account
Non-Farm Business Area	Other Large Animals (Horses, Donkeys, Camels, Elephants etc.)	Thresher, Other power driven machinery and Equipment	Other residential building outside village/town	Mills (Ghanies, Oil-Mills power-driven, rice-milling, flour milling)	Rickshaw/e-rikshaw	mutual fund	Deposits in Bank (savings)
Residential Area including housesite	Sheep, Goat, Pig and Rabbit	Laser Land leveler	Animal shed	Equipment used in beauty salon/spa	Motor Cars/Jeep/Van	Debentures/Bonds in Companies	Deposits- Fixed Term, Recurrent
Other Areas	Poultry (Chicken, Duck and other poultry birds)	Manually driven machinery and equipment	Farm house/Barn (Cola)/Others	Instruments used in Gyms	Tractor (all-types)		Savings in Post-office bank
	Others	Diesel Pumps	Building for non-farm business purposes (Factory and Workshop, shop)	Electric Motors, Generators, Pump sets, inverters etc	motor cycles/scooters/ auto-rickshaw		Other Fixed income deposits (NSC, KVP, saving bonds, etc)
		Electrip Pumps	Building for other purposes (charitable, recreational etc)	casting, melting and welding equipments / furnace, bellows/kilns	Others (including- Truck, Light Commercial Vehicle, Buses etc)		Deposits in Coperative Banks
		Drip Sprinkler	Work-in-progress (structure under construction)	scales, weights and measures			Deposit in non-banking Finance company
		Other machineries for irrigation	Other constructions (well,borewell, tubewell, etc)	saws: all types/ carpentry tools, drilling machines			Deposits with coop credit society, micro-finance insti, self-help groups
		Capital work-in-progress (machinery under installation)		xerox machine, printing press, computer, duplicating machine, fax			Insurance premium
		Furniture and Fixtures		Tools for Mobile repairing, Computer repairing etc			Provident fund
		Others		X-ray machine, medical equiments, Ultra sound, ECG machines etc			Contribution to Pension Fund,NPS
				Lathes, Other Machinery tools and Appliances			Interest free loans to friends/relatives
				Intangible assets like software, artistic orignals, manuscripts etc			Business Loans given to others
				Capital work-in-prrogress (under installation)			Personal loans given to others
				Other non-farm business equipment			Bullion and Ornaments
				Furnitutre and Fixtures			Paintings and Artistic originals

Notes: The table presents the entire list of items covered under different broad asset heads covered in the 2018-19 NSS AIDIS survey.

TABLE 2.26. Average Gross Wealth per adult (2010 Rs): 1961-2018

Real Average Gross Wealth Per Adult Rs				
	India	**Rural**	**Urban**	**Ratio=Ur/Rur**
1961		64,196		
1971		70,442		
1981	82,307	77,871	98,060	1.26
1991	126,001	116,506	153,420	1.32
2002	172,465	151,742	227,902	1.50
2012	479,207	330,884	759,603	2.30
2018	554,126	444,439	775,216	1.74

Notes: The table presents the real average gross wealth per adult from 1961 to 2018, using NSS AIDIS. Separate values for rural and urban areas are also provided. Wholesale Price Index (from World Bank Data) is used to bring the prices from nominal to the 2010 price level.

TABLE 2.27. Top Wealth Share (Net Wealth)

	Top (Net) Wealth Share														
	2002					**2012**					**2018**				
	10%	5%	1%	0.10%	0.01%	10%	5%	1%	0.10%	0.01%	10%	5%	1%	0.10%	0.01%
Raw Survey Shares	51.7	37.3	16	4.3	1	59.9	46.7	25.4	12	6.9	52.1	38.3	18.1	5.5	1.3
Threshold p for Pareto	**Post Correction Shares**														
90	66.7	56.9	39.2	22.9	13.2	69.5	59.4	41.3	24.4	14.2	71	62.2	45.6	29	18.2
91	66.6	56.8	39.2	22.9	13.2	69.7	59.6	41.4	24.4	14.2	70.7	62	45.6	29.1	18.2
92	66.4	56.6	39.1	22.9	13.2	69.7	59.6	41.4	24.4	14.2	70.5	61.8	45.5	29.1	18.3
93	66.1	56.4	39	22.9	13.2	69.7	59.6	41.4	24.4	14.2	70.4	61.7	45.5	29.1	18.3
94	65.9	56.1	38.9	22.8	13.2	69.5	59.5	41.3	24.4	14.2	69.9	61.2	45.3	29.1	18.4
95	65.3	55.4	38.5	22.7	13.2	69.3	59.2	41.2	24.4	14.2	69.6	60.8	45.1	29.2	18.5
96	64.8	54.8	38.2	22.6	13.2	69.2	59.1	41.1	24.4	14.2	68.9	60	44.7	29.1	18.7
97	64.1	53.8	37.6	22.4	13.2	68.8	58.5	40.8	24.3	14.3	68.1	58.9	44	29	18.8
97.5	63.6	53.2	37	22.2	13.2	68.4	57.9	40.4	24.2	14.3	67.6	58.2	43.5	28.9	18.8
98	62.9	52.3	36.3	22	13.1	67.9	57.3	39.8	24	14.3	66.9	57.3	42.7	28.7	18.9
98.5	62.1	51.3	35.3	21.6	13	67.3	56.5	39	23.8	14.2	66	56.2	41.6	28.3	18.9
99	60.8	49.7	33.3	20.7	12.7	66.6	55.6	37.8	23.3	14.2	64.7	54.6	39.7	27.6	18.9
99.5	58.8	47	29.8	18.9	12	65.2	53.7	35.2	22.1	13.8	63.4	52.8	37.4	26.6	18.7
99.9	55.6	42.9	24.4	14.3	9.8	62.8	50.5	30.7	18.3	12.4	61	49.8	33.3	23.1	17.5

Notes: The table presents the top (net) wealth shares after using surveys (NSS-AIDIS) and the Forbes millionaire. The threshold p is the cutoff up to which distribution from the survey is assumed to be valid. Beyond that threshold, Pareto distribution is created from the millionaire list. E.g. the threshold of p=.995 shows that the wealth share of the top 10% population increased to 69.6% (which was 52.1% from the raw survey) in 2018.

TABLE 2.28. Population Share (Brahmin) - IHDS 2011

	Percentage(%) of Population across Caste group							
	SC	**ST**	**OBC**	**FC(Brahmin)**	**FC(Non-Brahmin)**	**Muslim**	**Others**	**Total**
Total	21.80	8.00	42.78	4.86	14.90	6.19	1.47	100
0-14 Boys	23.12	8.21	43.95	4.27	12.41	6.96	1.09	100
0-14 Girls	23.44	8.59	44.13	4.08	11.56	7.28	0.92	100
15-20 Boys	22.60	7.75	43.10	4.72	13.93	6.55	1.36	100
15-20 Girls	23.42	7.62	43.18	4.59	13.00	7.09	1.09	100
>21 yrs Men	20.91	8.01	41.79	5.18	16.71	5.69	1.71	100
>21 yrs Women	20.82	7.74	42.43	5.25	16.44	5.55	1.77	100
>60yrs Men	18.93	6.89	43.17	6.06	17.99	4.80	2.16	100
>60yrs Women	19.71	5.92	43.76	6.22	17.87	4.27	2.23	100

Notes: The table presents the population share of different caste groups by age and gender, using IHDS 2011. The caste categories are as explained in the data section (See 2.4.2). The most important category to look for here is FC(Brahmins), which is for the Brahmin caste within the Forward caste. The rest of the Forward castes are in the FC(Non-Brahmin) column.

Table 2.29. Population Share FC split - NFHS 2005

		Population %											
	Overall	Young (0-14yrs) %			Adolescent(15-20yrs) %			Adult(>21 yrs) %			Old (>60yrs) %		
Caste Group		M	F	T	M	F	T	M	F	T	M	F	T
ST	8.397	9.07	9.44	9.25	8.38	8.81	8.6	7.86	7.73	7.79	6.46	6.74	6.6
SC	19.08	19.87	20.39	20.12	19.96	19.79	19.87	18.36	18.1	18.23	16.83	16.3	16.57
OBC	40.13	41.55	41.09	41.33	39.31	40.21	39.78	39.05	39.8	39.43	40.06	39.79	39.93
FC(Brahmin)	4.648	3.87	3.67	3.77	4.68	4.07	4.36	5.37	5.2	5.28	6.93	6.59	6.76
FC(Rajput)	4.902	4.37	4.07	4.23	5.03	4.68	4.85	5.39	5.32	5.36	5.83	5.48	5.66
FC(Baniya)	2.008	1.77	1.75	1.76	1.96	1.8	1.88	2.25	2.16	2.2	2.35	2.41	2.38
FC(Kayasth)	0.6344	0.41	0.41	0.41	0.46	0.5	0.48	0.81	0.82	0.82	0.97	1.18	1.07
FC(Other)	9.315	7.4	7.59	7.49	8.86	8.1	8.46	10.82	10.59	10.7	11.13	11.7	11.41
Muslim	8.691	9.98	10.07	10.03	9.25	10.31	9.81	7.49	7.63	7.57	6.18	6.28	6.23
Other	2.187	1.71	1.5	1.61	2.12	1.75	1.92	2.62	2.64	2.63	3.26	3.53	3.39

Notes: The table presents the population share of different caste groups by age and gender, using NFHS 2005. The caste categories are as explained in the data section (See 2.4.2). The most important category to look for here is the split of the Forward Caste group into Brahmin, Rajput, Baniya, Kayasth and Others.

Table 2.30. Economic and Education Indicators by Caste groups

		Ratio wrt National Averages						
	All-India Average	**SC**	**ST**	**OBC**	**FC (Brahmin)**	**FC (Non-Brahmin)**	**Muslim**	**Other**
Per Capita Annual Income (2011)	23,798	0.8	0.7	0.9	1.5	1.5	0.8	2.4
Per Capita Annual Consumption (2011)	22,956	0.8	0.7	1.0	1.3	1.4	0.9	1.8
Highest Adult Education (2011)	8.0	-1.3	-2.1	-0.2	3.5	2.3	-1.4	3.6
Highest Male Education (2011)	7.6	-1.3	-2	-0.1	3.7	2.3	-1.6	2.9
Highest Female Education (2011)	5.3	-1.4	-2	-0.3	3.3	2.5	-0.7	4.7
Net Wealth 2002	98,183	0.5	0.5	0.9	1.5		0.7	3.6
Net Wealth 2012	543,696	0.4	0.4	0.9	1.8		0.7	4.0
Net Wealth 2018	695,809	0.5	0.5	0.9	1.7		0.8	3.1

Notes: The table presents the ratio of the average values for economic and education indicators for different caste groups with respect to national averages, using IHDS 2011 and NSS-AIDIS (2002, 2012 and 2018) datasets. The first column provide the national averages. The last column, "Others", contains the rest of the population, which belongs to the Non-Hindu, Non-Muslim religion group and does not fall under SC/ST/OBC. It is a rich minority cluster (with only $\sim$ 1.5% population) regarding all the economic and educational parameters. For Education the indicator is between 0-16 with 0 denoting no education, 12- Higher Secondary 15- Bachelors, 16 above Bachelors.

TABLE 2.31. Wealth,Income and Consumption Standard Deviation-IHDS 2011

	SC	ST	OBC	FC(Brahmin)	FC(Non-Brahmin)	Muslim	Others	OVERALL
Annual Income of HH (in Rs)	190,165	106,633	156,237	211,666	277,773	162,160	411,715	196,567
Per Capita Annual Income (in Rs)	40,503	23,252	32,338	44,742	60,841	30,801	95,077	41,316
Annual Consumption of HH (in Rs)	77,577	73,477	108,385	139,527	136,120	98,475	163,451	109,792
Per Capita Annual Consumption (in Rs)	16,523	16,022	22,433	29,493	29,815	18,704	37,745	23,077
ASSETS	6.1	5.7	6.4	6.4	6.2	6.2	4.7	6.7
Highest Adult Education	5.7	5.3	6.0	5.9	5.7	5.8	4.4	6.2
Highest Male Education	5.1	5.0	5.1	4.1	4.7	5.1	3.8	5.2
Highest Female Education	5.0	4.9	5.0	4.0	4.6	5.1	3.9	5.1

Notes: The table presents the standard deviation of economic and education indicators, using IHDS 2011. Design weights are used to estimate these values. Assets is an average of 33 different household durable goods' (like TV, Air Conditioning, 4-wheeler etc.) dummy (=1). It ranges from 0-to 33 and is computed at the household level. The second last column, "Others", contains the rest of the population, which belongs to the Non-Hindu, Non-Muslim religion group and does not fall under SC/ST/OBC. It is a rich minority cluster (with only $\sim$ 1.5% population) regarding all the economic and educational parameters. The last column shows the all-India statistics.

TABLE 2.32. Population share by wealth index-NFHS 2005

	Wealth Index				
	Poorest	Poorer	Middle	Richer	Richest
Overall	20.63	19.82	19.86	19.6	20.09
ST	51	23.09	12.87	7.78	5.26
SC	28.47	24.8	21.19	16.08	9.46
OBC	18.87	21.66	22.94	20.7	15.83
FC(Brahman)	4.62	9.7	13.86	21.9	49.91
FC(Rajput)	7.27	13.78	21.9	25.89	31.15
FC(Bania)	5.8	11.86	16.52	22.17	43.66
FC(Kayasth)	2.17	5.25	10.89	24.67	57.02
FC(Other)	9.75	13.42	17.13	24.45	35.26
Muslim	20.91	21.19	19.11	21.8	16.99
Other	2.45	4.08	9.45	22.2	61.81

Notes: The table presents the population share of different caste groups in five quintiles (based on the wealth index), using NFHS 2005. The lowest quintile is labelled Poorest, and the highest quintile is Richest.

TABLE 2.33. Highest Adult Education Level-NFHS 2005

	Highest Adult Education in HH		
Caste Group	Overall	Male	Female
ST	4.34	3.92	1.8
SC	5.69	5.11	2.61
OBC	6.81	6.1	3.62
FC(Brahman)	11.88	10.87	8.28
FC(Rajput)	9.05	8.23	5.71
FC(Bania)	10.33	9.57	6.81
FC(Kayasth)	12.33	11.04	9.86
FC(Other)	9.16	8.15	6.35
Muslim	5.84	4.97	3.35
Other	10.83	9.22	8.51

Notes: The table presents the average education level of the adult members using NFHS-2005. The caste categories are as explained in the data section (See 2.4.2). The most important category to look for here is the split of the Forward Caste group into Brahmin, Rajput, Baniya, Kayasth and Others. The education levels are between 0-16 with 0 denoting no education, 12- Higher Secondary 15- Bachelors, 16 above Bachelors.

TABLE 2.34. Within SC/ST/OBC Relative to all-India: IHDS 2011

	Scheduled Tribe (ST)				Scheduled Caste (SC)			Other Backward Caste		
	Hindu	**Muslim**	**Christian**	**Others**	**Hindu**	**Muslim**	**Others**	**Hindu**	**Muslim**	**Others**
Annual Mean Income of HH	0.58	0.97	1.6	0.54	0.78	0.81	0.9	0.91	0.88	1.38
Per Capita Mean Annual Income	0.58	0.86	1.69	0.67	0.78	0.75	0.97	0.92	0.74	1.51
Annual Mean Consumption of HH	0.63	1.27	1.29	0.44	0.8	0.99	0.86	0.97	1.09	1.26
Per Capita Mean Annual Consumption	0.64	1.12	1.37	0.55	0.8	0.92	0.93	0.98	0.92	1.39
Assets 2011	-4.9	-1.21	1.71	-5.82	-2.2	-1.08	0.9	-0.08	0.43	4.9
Assets 2005	-4.66	-1.07	1.39	-5.46	-2.1	-0.88	0.85	-0.09	0.46	4.48
Highest Adult Education	-2.36	-1.41	1.8	-2.99	-1.41	-2.13	0.25	-0.08	-1.13	2.05
Highest Male Education	-2.19	-0.48	1.68	-2.98	-1.35	-2.26	-0.22	0.07	-1.2	1.48
Highest Female Education	-2.3	-1.24	2.4	-2.86	-1.51	-1.18	0.73	-0.28	-1	3.07

Notes: The table presents economic and education indicators (relative to all-India values) for different religions within SC, ST and OBC groups, using IHDS 2011 data. The relative values are the ratio for the first four rows (e.g. Hindu ST/All-India) and the difference (Hindu ST-All-India) for the last four rows.

TABLE 2.35. Percentage of Employment Type in Castes(2002)

	Employment type (2002)										
	Rural						**Urban**				
Caste Group	Self-Employed Agriculture	Self-Employed Non-Agriculture	Agricultural Labour	Casual Labour	Other	**Caste Total**	Self- -Employed	Regular Wage Salary	Casual Labour	Other Other	**Caste Total**
ST	42.75	6.16	33.95	11.56	5.58	**100.00**	21.86	45.62	21.44	11.08	**100.00**
SC	22.26	12.53	42.40	13.96	8.85	**100.00**	29.50	42.25	23.88	4.37	**100.00**
OBC	39.40	16.55	23.03	10.80	10.23	**100.00**	40.43	37.01	14.68	7.88	**100.00**
FC	50.23	13.44	13.11	6.65	16.57	**100.00**	34.34	49.62	4.98	11.05	**100.00**
Muslim	30.48	24.01	23.67	12.12	9.72	**100.00**	47.86	34.07	12.57	5.50	**100.00**
Other	49.81	13.59	9.17	10.69	16.74	**100.00**	43.26	38.76	4.58	13.40	**100.00**
Employment Total	37.63	14.46	26.31	10.86	10.75	**100.00**	36.87	42.34	12.21	8.58	**100.00**

Notes: The table presents the percentage share of the population engaged in different employment types - in rural and urban areas- in 2002. 56% of SCs in rural areas are involved in agricultural/casual labour, compared to 46% of STs, 34% of OBCs and 20% FC. Similarly, the share of SCs engaged in casual labour is also the highest in urban areas.

Table 2.36. Employment Type in different Caste(2012)

Employment Type (2012)	Rural							Urban				
Caste Group	Self-Employed Agriculture	Self-Employed Non-Agriculture	Agricultural Labour	Regular Wage Salary	Casual Labour	Other Other	**Caste Total**	Self- -Employed	Regular Wage Salary	Casual Labour	Other	**Caste Total**
ST	51.11	4.59	17.46	8.91	15.28	2.65	**100**	20.03	45.45	23.47	11.04	**100**
SC	28.90	10.41	24.76	10.29	20.21	5.43	**100**	23.44	43.80	25.51	7.26	**100**
OBC	44.76	12.56	15.05	9.43	12.43	5.76	**100**	36.17	35.86	17.65	10.32	**100**
FC	51.82	11.25	9.21	13.17	6.25	8.31	**100**	30.92	49.06	6.89	13.13	**100**
Muslim	31.84	21.04	13.89	8.85	16.44	7.94	**100**	42.46	32.02	19.24	6.28	**100**
Other	49.29	13.24	7.49	15.78	6.10	8.08	**100**	33.55	44.15	8.11	14.19	**100**
Tot Employment	42.76	11.45	16.19	10.23	13.47	5.91	**100.00**	32.44	41.59	15.28	10.68	**100**

Notes: The table presents the percentage share of the population engaged in different employment types - in rural and urban areas- in 2012. 45% of SCs in rural areas are involved in agricultural/casual labour, compared to 32% of STs, 27% of OBCs and 15% FC. Similarly, the share of SCs engaged in casual labour is also the highest in urban areas.

Table 2.37. Proportion of caste in Employment types(2002)

Proportion of caste in Employment types (2002)	Rural						Urban				
Caste Group	Self-Employed Agriculture	Self-Employed Non-Agriculture	Agricultural Labour	Casual Labour	Other	**Caste Total**	Self- -Employed	Regular Wage Salary	Casual Labour	Other Other	**Caste Total**
ST	11.19	4.19	12.71	10.49	5.11	**9.85**	1.74	3.17	5.16	3.80	**2.94**
SC	13.04	19.12	35.53	28.35	18.16	**22.05**	11.61	14.48	28.37	7.40	**14.51**
OBC	43.22	47.25	36.15	41.06	39.28	**41.28**	37.84	30.16	41.48	31.69	**34.50**
FC	25.54	17.78	9.54	11.72	29.49	**19.13**	33.39	42.00	14.63	46.18	**35.84**
Muslim	4.99	10.22	5.54	6.87	5.57	**6.16**	11.46	7.11	9.09	5.67	**8.83**
Others	2.02	1.44	0.53	1.51	2.38	**1.53**	3.96	3.09	1.27	5.27	**3.37**
Tot Employment	100	100	100	100	100	**100**	100	100	100	100	**100**

Notes: The table presents the percentage share of the different caste groups by employment types - in rural and urban areas- in 2002.

Table 2.38. Proportion of caste in Employment types(2012)

Employment Type (2012)	Rural							Urban				
Caste Group	Self-Employed Agriculture	Self-Employed Non-Agriculture	Agricultural Labour	Regular Wage Salary	Casual Labour		**Caste Total**	Self- -Employed	Regular Wage Salary	Casual Labour	Other Other	**Caste Total**
ST	13.90	4.66	12.55	10.13	13.20	5.22	**11.63**	2.11	3.73	5.24	3.53	**3.41**
SC	13.95	18.78	31.58	20.77	30.98	18.96	**20.64**	10.12	14.75	23.37	9.52	**14.00**
OBC	46.15	48.38	40.99	40.63	40.70	42.99	**44.08**	45.77	35.40	47.40	39.69	**41.06**
FC	20.27	16.44	9.51	21.53	7.76	23.51	**16.73**	31.28	38.71	14.81	40.34	**32.82**
Muslim	4.10	10.11	4.72	4.76	6.72	7.39	**5.50**	8.20	4.82	7.89	3.69	**6.26**
Other	1.63	1.64	0.65	2.18	0.64	1.93	**1.41**	2.53	2.59	1.30	3.25	**2.44**
Tot Employment	100	100	100	100	100	100	**100**	100	100	100	100	**100**

Notes: The table presents the percentage share of the different caste groups by employment types - in rural and urban areas- in 2012.

Table 2.39. Share of Caste Group by Employment Type within Rural Top Decile

	2002						2012						
Caste Group	Self-Employed Agriculture	Self-Employed Non-Agri	Agricultural Labour	Casual Labour	Other	**Caste Share**	Self-Employed Agriculture	Self-Employed Non-Agri	Regular Wage Salary	Agricultural Labour	Casual Labour	Other	**Caste Share**
ST	3.03	2.17	4.59	2.46	3.50	**3.01**	3.75	1.63	7.09	5.03	6.33	0.85	**3.85**
SC	4.47	5.45	16.64	7.56	6.26	**4.97**	5.29	4.88	10.50	6.89	9.45	4.89	**5.93**
OBC	41.60	46.72	40.58	46.03	38.59	**41.80**	46.32	59.56	42.53	49.77	61.62	54.64	**48.10**
FC	39.43	30.51	22.43	29.08	39.34	**38.18**	34.20	20.48	26.56	15.70	11.23	29.43	**30.88**
Muslim	3.96	6.83	7.03	7.11	3.27	**4.25**	2.88	6.53	5.09	12.07	5.62	0.49	**3.61**
Other	7.51	8.32	8.73	7.76	9.05	**7.79**	7.56	6.92	8.23	10.54	5.74	9.70	**7.63**
Column Sum	*100.00*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100.00*	*100*	*100*	*100*
Employment % share	**7.55**	**0.91**	**0.11**	**0.23**	**1.20**	**9.99**	**7.07**	**1.07**	**1.08**	**0.13**	**0.27**	**0.39**	**10.00**

Notes: The table presents the percentage share of different caste groups by their employment types within the top 10% in rural areas in 2002 and 2012. The deciles are created based on the wealth of rural residents. Self-employed in agriculture forms the major occupation within the top 10% rural population. The share of SC and ST have increased in within the rural rich class in 2012, though they are still under-represented compared to their population share.

TABLE 2.40. Share of Caste Group by Employment Type within Urban Top Decile

	2002					**2012**				
Caste Group	Self-Employed	Regular Wage Salary	Casual Labour	Other	**Caste Share**	Self-Employed	Regular Wage Salary	Casual Labour	Other	**Caste Share**
ST	1.10	0.79	0.19	2.06	**1.10**	0.87	2.33	6.51	2.83	1.86
SC	1.84	5.48	7.28	1.37	**3.16**	1.84	4.80	7.27	2.93	3.34
OBC	26.93	17.55	44.30	19.99	**22.76**	34.07	23.94	47.93	31.67	29.77
FC	53.01	65.11	12.12	69.48	**59.19**	50.45	58.97	30.50	46.17	53.07
Muslim	5.07	2.35	16.92	2.14	**3.80**	5.30	2.21	5.36	6.44	4.14
Other	12.05	8.72	19.19	4.95	**10.01**	7.47	7.75	2.43	9.96	7.81
Column Sum	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100.00*	*100*
Employment % share	**5.04**	**3.67**	**0.07**	**1.22**	**10.00**	**4.27**	**4.22**	**0.21**	**1.30**	**10.00**

Notes: The table presents the percentage share of different caste groups by their employment types within the top 10% in urban areas in 2002 and 2012. The deciles are created based on the wealth of urban residents. From 2002-12, the regular wage/salary earners have increased their share within the urban rich class. In 2002, 3.67% out of the 10% belonged to the salaried class, which grew to 4.22% in 2012. Within this wealthy salaried class, the share of SC declined, but ST rose.

TABLE 2.41. Transforming Sampling Design of 2018 to 2012

(1) SSS (2018)	(2) SSS SSS	(3) # of HH	(4) Survey Weights (Original)	(5) New SSS (like 2012)	(6) # of HH	(7) Design Weights (New)
HH with MPCE>A and Indebted type A	1	2	d1	HH Indebted type A	4	(d1+d4)/2
HH with MPCE<=A and Indebted type A	4	2	d4			
HH with MPCE>A and Indebted type B	2	2	d2	HH Indebted type B	4	(d2+d5)/2
HH with MPCE<=A and Indebted type B	5	2	d5			
HH with MPCE>A and Indebted type C	3	2	d3	HH Indebted type C	4	(d3+d6)/2
HH with MPCE<=A and Indebted type C	6	2	d6			
Total HH Selected		**12**	**D**		**12**	**D**

Notes: The table presents the household sample selection in the second stage strata in NSS AIDIS 2018. Col(1) shows the six SSS using MPCE and three Indebtedness categories. Col(2) is simply the numbering of SSS. Col(3) shows the number of households sampled from a SSS. Col (4) is the survey weights provided in 2018. To make the sampling design similar to the 2012 (where only three Indebtedness categories were used), I combine SSS1 and 4; SSS2 and 5; SSS3 and 6 and select 4 households randomly from each newly created SSS. Col (7) provides the new design weights to keep the overall weights same.

Table 2.42. Educational Profile of the Indian Married Couples

EDUCATIONAL CHARACTERISTICS						
	All-India		Rural		Urban	
	Men	**Women**	**Men**	**Women**	**Men**	**Women**
N (Total Samples)	34713	34713	65.58%	65.58%	32.86%	32.86%
Total Couples	203,933,402	203,933,402	68.64%	68.64%	31.36%	31.36%
Average Age (in years)	40.1	35.1	39.57	34.71	41.25	35.96
Education (%of Population)						
No Education 0	21.97	39.61	27.06	48.28	10.84	20.65
Less than Primary	9.06	7.21	10.22	7.86	6.52	5.78
5th Pass	9.28	8.97	10.27	9.33	7.12	8.17
8th Pass	26.69	22.68	26.66	21.12	26.77	26.09
Secondary	14.33	9.9	12.6	7	18.09	16.22
H.Secondary and Diploma (<3yrs)	9.11	6.36	7.39	4.26	12.89	10.96
Bachelors(BA,Bsc,Diploma 3+)	6.88	3.88	4.1	1.6	12.97	8.87
BTech,MBBS,MD,CA,PhD	2.67	1.4	1.7	0.55	4.79	3.26
Total (%)	*100%*	*100%*	*100%*	*100%*	*100%*	*100%*

Notes: The table presents the educational characteristics of the couples' dataset created using IHDS 2011 datasets. Design weights are used to estimate these values. Educational categories are defined in Appendix F.7. The rural sample is 66% and the rest urban. In the full sample, 22% men and 35% women fall into 0 educational categories.

TABLE 2.43. Employment and Earnings of Married Couples

Employment Characteristics of Couples

	Total		Rural		Urban	
	Men	Women	Men	Women	Men	Women
Primary Activity Status						
%age of Total Couples Covered	**92.69**	**22.71**	**90.89**	**26.11**	**96.63**	**15.29**
Salaried/Professional	21.86	15.27	11.07	8.53	44.06	40.45
Small Business/Artisan	15.73	7.05	10.83	4.5	25.81	16.58
Cultivators	24.23	20.07	34.67	24.99	2.74	1.7
Non Agri Wage Labour	26.49	23.36	27.13	21.19	25.17	31.49
Agri Wage Labor	11.69	34.24	16.3	40.79	2.22	9.78
Total (%)	100%	100%	100%	100%	100%	100%
Occupation Types						
% of Total Couples Covered	**99.93**	**31.8**	**99.94**	**37.97**	**99.92**	**18.28**
Professional	5.08	5.13	3.61	2.76	8.28	15.9
Admin,Exec,Managers	1.26	0.13	0.44	0.05	3.06	0.52
Clerical	6	2.54	3.13	1.59	12.28	6.84
Sales	14.88	4.37	8.97	2.47	27.81	13.01
Service Providers	3.96	4.85	2.97	2.2	6.11	16.86
Farmers,Cultivators	29.36	51.89	40.45	60.21	5.07	14.08
Labourers (Non-Agri)	39.46	31.09	40.43	30.72	37.39	32.8
Total (%)	100%	100%	100%	100%	100%	100%
Earnings						
% with positive income	71.3	26.88	71.86	32.23	70.08	15.18
Average Annual Income	66179	22182	44061	15000	115810	55547
Imputed Earnings						
% with positive income	99.09	99.86	99	99.89	99.29	99.78
Average Annual Income	55165	19702	36748	11735	95352	37156

Notes: The table presents the educational characteristics of the couples' dataset created using IHDS 2011 datasets. Design weights are used to estimate these values. The categories are per the definition in Appendix F.7. Percentages are calculated out of the total population for which data is present. 21.86% of men have salaried/professional status out of the sample (92.69%) for which the primary activity status information is available.

TABLE 2.44. India France Comparison: Assortative Matching

	(1)	(2)	(3)	(4)	(5)	(6)
	Highest Degree					
PANEL A: Education	**France**		**India (all)**		**India excluding 0 education**	
	Polychoric	Spearman	Polychoric	Spearman	Polychoric	Spearman
Couple's Education	0.593	0.559	0.64	0.63	0.51	0.503
Father's Education	0.506	0.437	0.611	0.487	0.133	0.113
Mother's Education	0.476	0.401	0.695	0.512	0.001	0.005

	Gross Wage Earnings					
PANEL B: Earnings	**France**		**India (given earnings)**		**India (imputed earnings)**	
	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
Couple's Earnings (including 0)	0.175	0.175				
Couple's Earnings (excluding 0)	0.31	0.269	0.64	0.63	0.439	0.502

	Occupational Status					
PANEL C: Occupation	**France**		**India (Occupation)**		**India (Primary Activity Status)**	
	Polychoric	Spearman	Polychoric	Spearman	Polychoric	Spearman
Couple's Occupation	0.531	0.453	0.434	0.352	0.728	0.649

Notes: The table compares the education (Panel A) and economic (Panel B and C) assortativity between India and France. The education assortativity in India is slightly lower than in France but economic assortativity is higher. The France data comes from Frémeaux and Lefranc, 2020.

TABLE 2.45. Average Village-Level Percentage of Landless Population

	Landlessness (%)		
State	**Household level**	**Individual Level**	**Adult Level**
Rajasthan	35.54	32.62	30.99
Karnatka	40.87	37.37	36.36
UP	42.62	39.77	38.45
Maharashtra	50.14	47.08	45.13
Madhya Pradesh	53.18	50.86	48.78
West Bengal	58.47	55.76	55.21
Bihar	61.15	58.61	58.06
TN	69.53	67.13	65.91
Kerala	70.78	69.68	69.18
Punjab	76.53	75.58	74.08

Notes: The table presents the average of the village-level landlessness (%) from SECC-2011. The three columns are at three levels - household, individual and adult level. The states are ranked from low to high level of landlessness.

TABLE 2.46. Top Landholders share (%) in villages

State	Top Landholders share of land (%) Top 1	Top 2	Top 3
Maharashtra	10.03	15.57	19.76
Rajasthan	11.30	17.03	21.30
Karnatka	11.48	17.17	21.37
Tamil Nadu	11.82	18.01	22.45
Uttar Pradesh	12.48	18.53	23.03
Madhya Pradesh	13.80	20.73	25.87
Kerala	14.28	20.32	24.45
West Bengal	17.87	25.65	31.04
Bihar	21.36	30.08	35.81
Punjab	22.15	33.24	40.70

Notes: The table presents the average village-level land share (%) owned by the top 1 (Col (1)), top 2 (Col (2)) and top 3 (Col (3)) households in a village, computed from SECC-2011. The states are ranked in the ascending order. The richest household in villages of Punjab on average own 22.15% of land compared to 10% in Maharashtra.

TABLE 2.47. Land Inequality (Gini Coeff HH) and SC and ST population share by States

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	tn	karnatka	kerala	maharashtra	up	bihar	mp	rajasthan	punjab	wb
SC share	0.117***	0.112***	0.105**	0.128***	0.125***	0.147***	0.116***	0.138***	0.172***	0.0539**
	(0.009)	(0.012)	(0.036)	(0.020)	(0.010)	(0.011)	(0.013)	(0.017)	(0.020)	(0.019)
Observations	14,689	26,734	1,013	40,059	88,734	37,280	47,914	42,558	11,046	33,411
R-squared	0.482	0.355	0.463	0.373	0.219	0.318	0.205	0.209	0.404	0.437
Mean	0.826	0.644	0.924	0.710	0.673	0.827	0.717	0.618	0.844	0.789
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports OLS estimations based on a village-level sample for the ten states separately. Robust standard errors in parentheses are clustered at the district level. The dependent variable is the Gini coefficient at the village level computed based on the HH-owned agricultural land. All estimations include SC and ST population share in the village. All the columns include district fixed effects; Demographic controls (total population, population density, literacy rate, working population share, and agricultural population share); Geographic controls (distance from the nearest statutory town in km, total village land, and forest density); and Climatic controls (elevation, roughness, decadal average temperature 2001-11 and decadal average precipitation 2001-11). The coefficient on the SC share is positive and statistically significant in all the states.

TABLE 2.48. Land Inequality (Gini Coeff Individual) and SC and ST population share

VARIABLES	(1) gini indiv	(2) gini indiv	(3) gini indiv	(4) gini indiv	(5) gini indiv	(6) gini indiv
SC population share	0.133***	0.124***	0.121***	0.121***	0.121***	0.121***
	(0.0127)	(0.00778)	(0.00643)	(0.00649)	(0.00648)	(0.00542)
ST population share	-0.00765	0.00846	0.0175**	0.0183**	0.0203***	0.0203***
	(0.0151)	(0.00863)	(0.00756)	(0.00746)	(0.00734)	(0.00630)
total population			1.38e-05***	1.36e-05***	1.36e-05***	1.35e-05***
			(7.44e-07)	(7.43e-07)	(7.37e-07)	(7.11e-07)
Distance to Nearest Town (in km)				-0.000202***	-0.000205***	-0.000214***
				(4.48e-05)	(4.24e-05)	(2.82e-05)
Constant	0.687***	0.687***	0.874***	0.879***	0.943***	0.841***
	(0.00737)	(0.00198)	(0.0106)	(0.0107)	(0.0814)	(0.0709)
Observations	349,882	349,882	349,446	340,730	340,686	340,685
R-squared	0.027	0.321	0.390	0.395	0.397	0.445
Mean	0.71	0.71	0.71	0.71	0.71	0.71
District FE	No	Yes	Yes	Yes	Yes	No
SubDistrict FE	No	No	No	No	No	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes
Geographical Controls	No	No	No	Yes	Yes	Yes
Climatic Controls	No	No	No	No	Yes	Yes
Num_cluster	336	336	336	336	336	336

Notes: The table reports OLS estimations based on a village-level sample for the ten states. Robust standard errors in parentheses are clustered at the district level. The dependent variable is the Gini coefficient at the village level computed based on the HH owned agricultural land equally split within household members. All estimations include SC and ST population share in the village. Column 1 is without controls; Column 2 adds District fixed effect; Column 3 further adds Demographic controls (total population, population density, literacy rate, working population share, and agricultural population share); Column 4 further adds Geographic controls (distance from the nearest statutory town in km, total village land, and forest density); Column 5 further adds Climatic controls (elevation, roughness, decadal average temperature 2001-11 and decadal average precipitation 2001-11); Column 6 replaces District FE with Sub-district FE while keeping all the controls. Demographic and Geographic controls come from the Census 2011.

F. Appendix Notes

F.1. NSS AIDIS Surveys: Stratification Methodologies. The sampling design is two stage-stratification in each stratum in rural and urban sectors separately. Census villages and urban blocks form the First stage units (FSU) in rural and urban areas respectively. [58] Households form the Second-stage units (SSU). The list of villages are drawn from the recent most available census information.

1971-72: The rural area was divided into 66 agricultural regions having similar crop patterns and population density. Each region was further divided into strata by grouping geographically contiguous tehsils/districts having similar characteristics based on altitude, transport and communication facilities. The effort was to keep the population dependent on agriculture same across all strata. In total 379 rural strata were formed and 12,452 villages were sampled.[59] The second-stage stratification (SSS), within all the sampled villages, was done based on the area of operated land by the households. The four second-stage strata were- non-cultivators, small, medium and large class cultivators.[60] On average three households were randomly selected[61] within each second-stage stratum in every sampled village (or 8 HH's per village).

1981-82: The rural area was divided into 77 agricultural regions having similar crop patterns and population density. The rural and urban strata were 497 and 363 respectively.[62] In total, 7718 rural FSU's and 5169 urban FSU's were sampled.[63]. The second-stage stratification (SSS), within all the sampled villages, was done based on the area of land possessed by the households. The four second-stage strata were- landless (land

[58] The larger FSU's are split into sub-strata, to keep similar population level.

[59] Within each stratum, in total 36 villages were selected - 12 forming Central Sample, 12 State Matching sample and 12 Additional Matching Sample by Reserve Bank of India (RBI)

[60] A non-cultivator household is defined as an household having an operational holding of area less than 0.005 acre and/or having land which has wholly put to non-agricultural uses. The cut-off points for small, medium and large cultivators was decided on regional basis to equalise the total land area operated in each of the sub-strata at regional level.

[61] The selction was done based on linear systematically with specified intervals and random starts, which were pre-determined to make the design self-weighting at the regional level for each sub-stratum

[62] The state level allocations were determined based on the investigator strength and expected work load per investigator. The allocation to strata within a state were based on the proportion of total number of 1981 census house-listing blocs in rural sector and the total number of blocks in the frame in urban sector

[63] The initial plan was to sample 8408 villages and 5300 urban blocks. The decline is primarily due to two reasons- operational difficulties and shortage of staff leading to no survey in some regions and non-availability of some records

possessed<0.05acres), small land possessing HH's (0.05acres-A), medium land possessing HH's (A-B), large land possessing HH's (>B acres).[64] On average two households were randomly selected[65] within each second-stage stratum in every sampled village. In urban sector, three SSS were formed based on the monthly per capital expenditure (MPCE).[66] On average two households were selected in each sub-stratum. [67]

1992: Similar to previous rounds, first each state/union territories was divided into 78 similar agro-economic regions (population density and crop pattern) and then into several basic strata from which FSU's were randomly selected.[68] In total, 4231 rural FSU's and 2419 urban FSU's were sampled. A major change occurred in the second-stage strata formation in this round, where indebtedness status of the households was also considered (in combination with land possession) to capture the debt situation in the country better. In rural sector 7 second-stage strata were formed. First, based on land possession 4 groups were formed, namely non-cultivators (<.005 acres), small (.005-X), medium (X-Y) and large (>Y) cultivators with increasing land area. X and Y varied with state to ensure that in all the three cultivators group total land area remained same. The first two groups were sub-divided into "indebted" and "non-indebted" groups to form AIDIS sub-strata 1 to 4. Medium and Large cultivators were then merged and subdivided into three sub-strata 5 to 7- "indebted to institutional agencies with or without being indebted to non-institutional agencies", "indebted to non-institutional agencies alone" and "non-indebted". In urban sector also, 7 second-stage strata were formed based on MPCE and indebtedness of households.[69] In both sectors, from SSS 1 to 7: 1,1,1,2,1,1,

[64]The cut-off points A and B were decided for each village to form three other sub-strata, in such a way that total area of land possessed was nearly the same for each sub-stratum.

[65]The selection was done based on circular systematically with a random start. The households of sub-stratum 1 were arranged by their means of livelihood - agricultural labour, artisan and others before sample selection.

[66]The boundary points for them were fixed at state level.

[67]The selection was done based on circular systematically.

[68]The selection of sample villages was done by Probability proportional to size with replacement (PPSWR) with population as the size of variable and the sample blocks were selected by simple random sampling without replacement.

[69]Households were first grouped into 3 classes - less than x, x-y and greater than y. The cut-off x and y was decided at state level, based on NSS consumption survey to allocate 30%, 60% and 10% of the urban population of the state. Similar to in rural sector these three groups were divided into 7 sub-strata based on indebtedness.

and 2 households were sampled respectively. [70] In total 57,031 households were sampled.

2002-03: The stratification exercise was similar to previous round of survey with slight modifications.[71] A total of 10,608 FSU's was decided to be sampled, though ultimately 10,309 were surveyed with 1.5 weightage to urban sector. [72] Sub-strata formation was similar as in 1991-92, with some slight changes. The cutoff points X and Y were determined at state level to ensure that 40% of HH's possess land area less than X, 40% possess land are between X-Y and 20% possess land area greater than Y. In Urban sector, based on MPCE, 4 groups were formed (instead of 3 in 1991-92)- less than A, A-B, B-C, >C. The cut-off points A, B and C was decided at state level to ensure 30% households fall into first, second and third groups with 10% falling into fourth. The first two groups (in both rural and urban sectors) were then divided into indebted and non-indebted categories. The last two groups were merged and sub-divided into institutional indebted, non-institutional indebted and non-indebted. In this way 7 second-stage strata were formed and 2 HH's from each SSS were sampled (i.e. 14 HH's from each sample village/block.) In total 91,192 in rural and 52,903 in urban sector were sampled.

2012-13 The first difference with 2002-03 was that at the stage of allocation of FSU's between rural and urban sectors, relatively lesser rural FSU's (and eventually lesser rural HH's) were sampled. [73] The second difference was that the second-stage (or ultimate stage) strata formation was done only with HH's indebtedness. In each selected FSU's three SSS were formed - institutional indebtedness, non-institutional indebtedness and

[70] Selection of households in each sub-stratum was done by Simple Random sampling without replacement (SRSWOR).

[71] Though technically, it became multi-stage design with First stage, Intermediate stage and Ultimate stage units. The intermediate stage of sampling was in case of large FSU's. Ultimate stage unit was households, which was called Second stage unit in previous rounds.

[72] Out of 6,784 rural FSU and 3824 urban FSU's, 6,552 and 3,824 FSU's were eventually surveyd. The allocation of these FSU's to the different states and union territories was in proportion to provisional population as per census 2001. The state/ut level sample size was then allocated between the rural and urban areas in proportion to census 2001 population with an weightage to 1.5 to urban areas. The sample to be selected from each strata (within rural and urban sectors) was based on the proportion of population

[73] Even though the report mentions that double weightage was decided to be given to urban sector the micro-datasets available shows only 1.2 times more rural FSU's were surveyed, compared to 1.76 in 2002-03 and 1.74 in 1991-92. This resulted into higher share of urban households being sampled compared to previous rounds. The FSU's were selected by Simple Random Sampling without Replacement in both rural and urban sectors.

non-indebtedness. A total of 14 HH's were selected within each FSU- 6, 4 and 4 HH's from SSS 1, 2, and 3 respectively. A total of 4,529 rural and 3,507 urban FSU's comprising 62,135 and 48,665 households were surveyed.

2018-19: A total of 5940 rural and 3995 urban FSU's comprising 69,455 and 47,006 households were surveyed. [74] The weightage to urban sector was 1.5, i.e. for every 1 HH's in urban sector 1.5 HH's in rural sector was sampled. The second-stage strata formation again got changed. In each selected FSU's (in both rural and urban sectors) six SSS were formed based on indebtedness and MPCE. SSS 1 to 3 were formed from HH's in the top 20% MPCE category combined with indebtedness - institutional indebtedness, non-institutional indebtedness and non-indebtedness. The SSS 4-6 were formed from HH's in the bottom 80% MPCE category combined with indebtedness - non-institutional indebtedness(split into MPCE>A and MPCE<=A) and non-indebtedness. Two HH's from each SSS were sampled, resulting into a total of 12 HH's within each FSU.

F.2. NSS AIDIS Surveys: Valuation of Assets. In 1961-62, all the physical assets were evaluated using the average market value prevalent in the visit. The financial assets (except shares of companies and cooperatives) were evaluated at their face value. The values of the shares of companies and co-operatives were determined by the paid-up value. All dues receivable on loans in kind were evaluated using average wholesale prices. The valuation of assets changed from 1981. Due to the lack of book value for the valuation of assets in the household sector, the procedure followed is as below:

- Value of physical asset acquired prior to the 30th June 1981 (1991, 2001, 2011, 2018) was evaluated in its existing condition at the current market price prevailing in the locality on the date of the survey if the asset is owned on the date of survey, or on the date of disposal if the asset is disposed of during the reference period in a manner other than sale.
- In case the asset is sold/purchased during the reference period, sale/cost price is considered the asset's value. If the asset is acquired through construction, the expenditure incurred on construction is taken as its value.

[74] The selection of FSU's within each strata was based on the Simple Random Sampling without Replacement as in the previous round

- If an asset is acquired/disposed off other than purchase, then the asset's value in its existing condition as prevailing in the locality at the time of acquisition is reported.
- If an asset is acquired and disposed of during the reference period, the disposal value is reported.

F.3. Converting from Household Level to Individual Level. The first step, before estimating adult individual level wealth, is to estimate AVAH from the next round of micro-surveys files. I use a non-parametric approach to estimate the average adult size at HH Level (AVAH) for 1961, 1971, and 1981 survey years from the next survey years. To predict 1981-82 AVAH, I assume the decadal rate of change of AVAH during 1991-2002 to be the same during 1981-1991. The demographic factors usually are slow-moving variables, which makes the assumption valid. However, it might fail if there was a significant demographic shock (like some natural calamity/disease/war wiping out a big share of the population or strict government policy like the One-child Policy), making the decadal growth rates different. It is impossible to say by certainty, but there were not a large scale demographic shock to impact all-India averages. The wars between 1961-1981[75] were not on a large scale in terms of human capital. Forced sterilization during the Emergency (1975-77) is a potential threat[76] as it affected 1.5% population. A varying decadal migration rate (even temporary) could also impact AVAH. Census provides the share of migrating population, which was 30.8%, 28.7%, 29.4%, 26.6% and 29.3% in 1961, 1971, 1981, 1991 and 2001 respectively. The decadal variation is not too significant to yield concern. Keeping the above potential issues in mind, I follow the following steps:

(1) Assuming the same decadal rate at the decile level, predict AVAH and Average HH size (AVH) for 1981.

(2) Correct the predicted 1981 AVH and AVAH values using "correction factors". These correction factors are:

a) AVH: A factor to make the predicted population level equal to the representative population of the survey.

[75] 1962 India-China war- resulted in 10,000(0.002% of the total population) human loss; 1965 India-Pakistan war resulted in 3000 casualties; 1971-72 India Pakistan war - 4000 casualties.

[76] According to Shah Commission Third Report 19789:2007, it resulted in 8 million extra sterilization which was 1.5% of the 1971 census population. It could make big the adult population share after 20 years, i.e. 1991-2001 decade.

b) AVAH: A factor to make the predicted adult population share the same as in Census.[77] [78]

(3) Use the predicted AVH and AVAH for 1981 and repeat a) and b) to predict the year 1971.
(4) Repeat step c) to predict for 1961 level.

The second step involves generating a full distribution using Generalized Pareto Interpolation- first at the household level and then applying the decile-wise estimated AVAH to recalibrate the p values, bracket average, and threshold values at an individual level and re-running the process at the adult-individual level.

Using predicted AVH and AVAH 1981 levels, I repeat the process to estimate the 1971 level. And further, using the estimated 1971 level, I predict the 1961 level.

F.4. Sampling Design Correction. The NSS-AIDIS surveys have slightly different sampling methodologies in different years. One difference in the 2018 (compared to 2012) survey was that the households were randomly selected from the SSS formed based on the household's MPCE and indebtedness compared to only the household's indebtedness in 2012. The households were chosen so that half of the selected HHs come from the top 20% MPCE category. The design survey weights underweight these over-sampled high consumption households to generate all-India estimates. However, if these over-sampled (high-consumption) households have more non-zero wealth compared to the low-consumption HHs with zero wealth, then it implies more HHs with some wealth will be sampled compared to the scenario when MPCE is not used. And this could impact the wealth inequality estimate (even after using the survey design weights). It is hard to say in which direction the wealth inequality will be biased due to this change in sampling design.

To make both surveys comparable, I emulate the 2012 strategy in 2018.[79] First, I generate a full population dataset using the provided weight for each household and

[77]The data for adult population share comes from the website: https://www.populationpyramid.net/india., I use the same adult share for Rural and Urban areas. It will be better to use it separately. The lack of good sources has resulted in compromising here.

[78]Different sampling of adults in Census and survey might be a threat. For 1981, I have the comparison between survey sampling and Census at the household level, which I use here. For 1971 and 1961, I use census population, which means the survey population was perfect for the census population

[79]It is not possible to convert 2012 to 2018 because MPCE information is not present in the 2012 survey.

randomly select the households without using the MPCE condition. Imagine in a village of population D; the 2018 survey selected 2 HHs in six SSS (with survey weights d1-d6) using MPCE and indebtedness criteria. The 2012 survey used only the indebtedness type of the households. I combine SSS1 and SSS4, SSS2 and SSS5 and SSS3 and SSS6 to create three SSS - based on the households' indebtedness type (similar to 2012). Next, from each newly created SSS, I randomly select four households (still selecting the same number of HHs) and compute the new weights for each HH. Table 2.41) describes tabular form. I re-compute the wealth inequality based on the newly selected dataset and repeat the process 72 times. The distribution is provided in the Appendix Figure 2.18. The average value of the top 10% net wealth share comes out as 52.33% which is 0.4% higher than the original share (52.1%). It shows that the change in the sampling methodology is not the reason behind the declining survey-based estimates.

F.5. Population Share by Caste Groups using Other Surveys. Figure 2.10) presents the population share using NSS Consumption surveys (1983, 87, 93, 99, 2004 and 2009). The population share of different caste groups and their evolution is very similar to what we observe using NSS-AIDIS surveys. There is an increasing OBC's population share and decreasing FC's share. The decline in the population share of Non-Hindus (Muslims plus Others) in 1999 is an artefact of the classification method where the OBC category takes precedence. E.g. OBC Muslims are categorized under OBC. There is an increasing share of other religions among STs and OBCs. In 1983, Hindus comprised 90% of ST, which dropped to 86% in 2009. In the OBC, the share of the Hindus dropped from 88% to 84% from 1999 to 2009. Regarding the rural-urban divide, a high share of the FC group lives in Urban areas, and a high share of STs live in rural areas. (Refer Appendix Figure 2.11.)

Appendix Figures (2.12 and 2.13 show the proportion of SC/ST/OBC in different religions in surveys. Pre-1993 surveys have no OBC. In Hindus, the share of SC and ST is almost stable at around 9.5% and 19-21%, respectively. For later rounds, in 1999, OBCs formed 38%, 30%, 20%, 13.6% in Hindus, Muslims, Christians and Sikhs, respectively which increased to 43%,43%,25%,21% in 2009 in the same order. The increase in the OBC population is in all religions, chiefly due to the reclassification of more castes into

OBC.[80]

IHDS datasets allow splitting the FC category into Brahmins and the rest of the FC. The population share of Brahmins is $\sim$ 4.86%. Table 2.28 provides the percentage of population in different caste groups. The information on the Brahmin community, considered the highest caste in the caste hierarchy makes this table interesting. Another important observation is that child and young population among SC, ST, and OBC is higher than their respective population. In contrast, FC's adult and the old population is higher than their respective population. It suggests lower fertility level of FC compared to the other caste groups.

NFHS allows to further split the FC into different caste categories.[81] Table 2.29 provides the population shares. I split the FC into Brahmins (4.7%), Rajputs (a proxy for Kshatriya- including Marathas 4.9%), Bania (merchant class; 2.1%), Kayasth (0.6%) and Others (9.3%). The population share of Brahmins is close to what we get from the IHDS dataset- which is a robustness check for caste categorization.[82] The lower fertility is true for other FC groups too. They have a higher proportion of adults and old than their overall population.

F.6. Representational Inequality in Rural and Urban Areas. Rural

Fig 2.14 shows the percentage share of different castes by wealth deciles in Rural areas. There is an over-representation of FC in the top wealth decile, 19pp in 2002 and 14pp in 2012. There was a slight under-representation of OBC in the top wealth decile (-2pp) in 2002, which improved to +2pp in 2012. There is under-representation of SC, ST, and Muslims in the top wealth decile (or even in all top 5 deciles). The situation is the worst for SC. They were under-represented by 41 pp within the top 50% in 2002, which also means by construction, they are over-represented in the Bottom 50% by 41 pp. The

[80]There may be slightly higher fertility levels among OBCs, but it can not explain this big gap.

[81]Currently, the categorization is done through google searches and crowdsourcing, hence prone to errors. Some researchers are in the process of cleaning the caste names from NFHS.

[82]The NFHS dataset provides names of jatis, which I categorize into Brahmins, Rajputs, Bania and Kayastha, using different sources. I emphasize that there is scope for improvement here. Bania population is higher in NFHS (compared to the 1901 census), which could be due to errors in categorization. However, business as a profession is taken up by other jatis, which are now associated with the Bania community.

under-representation of ST and Muslims in the Top 50% is 15.5 pp and 9 pp respectively in 2002. Between 2002-2012, SC and Muslims reduced the under-representation (SC: 20.4 pp, Muslims: 5pp) in the top 50%. For the ST group, the under-representation in the top 50% increased to 29 pp.

Fig 2.15 shows the percentage share of different castes in different deciles in Urban areas. The results in urban areas are different from those in rural areas. In 2002, FC over-representation in the top wealth decile by 25pp, which remained the same in 2012. OBCs are under-represented in the top wealth decile by 12 pp in 2002 and 2012. Like rural areas, there is an under-representation of SC, ST, and Muslims in the top wealth decile (or even in all top 5 deciles). The situation is the worst for SC. They were under-represented in the top 50% wealthy population by 27.7 pp in 2002, which improved to 24.2pp in 2012. The under-representation of ST and Muslims in the Top 50% was 5pp and 13pp respectively in 2002. Between 2002-2012, it remained the same for ST but improved to 9pp for Muslims.

F.7. Ranking of Employment and Educational categories. Ordered categories of Employment: Two variables are used for the categories of occupation. First is Primary Activity Status, which identifies an activity as primary for the household member. I created five groups, namely- 1) Salaried/Professional- combining Organized Business, Salaried and Profession. 2) Small Business/Artisan 3) Cultivators- Cultivators and Allied Agriculture 4) Non-Agricultural Wage Labour and 5) Agricultural Wage Labor.

The second is using the occupation type using two variables - occupation codes (which uses National Classification of Occupation 1968; available only for the wage earners) and primary activity status (for non-wage earners).[83] The categories are equivalent to the professions and socio-professional (PCS) categories of INSEE France (to compare the assortativity level with France). The seven hierarchical groups are:

(1) "Professional" - comprising scientists, engineers, teachers, jurists etc.
(2) "Admin, Exec, Managers"- elected and legislative officials, executive and managerial workers etc.

[83]For example, if the occupation code is missing and the primary activity status is cultivator, I categorize the individual under the "Farmers, Cultivators" category. Organized business (from Primary Status activity) is clubbed under "Profession" as they are related to a well-established unit ($>$ 10 employees). The salaried class is under clerical. Artisans and independent workers are under "Sales".

(3) "Clerical"- clerical and supervisors, stenographers, conductors and guards etc.
(4) "Sales" - merchants and shopkeepers, manufacturers, sales workers etc.
(5) "Service Providers"- hotel keepers, maids, cooks, waiters, etc.
(6) "Farmers, Cultivators"- Agricultural labourers, cultivators, fishermen etc.
(7) "Labourers (Non-Agri)"- miners, tailors, carpenters, plumbers, construction workers etc.

Ordered categories of education: Similarly for estimating education level AM, the categories are defined based on the questions related to completed years of education, highest degree obtained and whether the individual has even attended school. There are two sets of categories for education. One with 8 categories, namely-

(1) "No Education" - zero years of education
(2) "Less than Primary"- < 5 years of education
(3) "5th Pass" - $\geq 5yrs$ and $< 8yrs$
(4) "8th Pass" - $\geq 8yrs$ and $< 10yrs$
(5) "Secondary" - $\geq 10yrs$ and $< 12yrs$
(6) "H. Secondary and Diploma (< 3yrs)"[84] - $\geq 12yrs$ and $< 15yrs$
(7) "Bachelors (BA, Bsc, Diploma 3+)"[85] - $\geq 15yrs$
(8) "BTech, MBBS, MD, CA, PhD"[86] - $\geq 15yrs$

The last two groups differentiate between elite and common higher educational degrees in education. The last group is synonymous to *Grand Ecole* of France, but since college information is not present, it is not perfectly comparable. The classification based on highest degree is imperfect.[87]

F.8. Caste and Employment Type. I check the employment type of different caste groups for the 2002-03 and 2012-13 surveys.[88] The employment type is based on the major source of income during 365 days preceding the survey of the household head. In rural areas, the five classifications in 2002 include Self-employed in agriculture and

[84] Higher Secondary
[85] BA- Bachelors in Arts, Bsc- Bachelors in Science
[86] BTech- Bachelors in Technology, MBBS- Bachelor of Medicine, MD-Medical Degree, CA- Chartered Accountant
[87] For example, A BA degree from St. Stephen's college will have more weight than a B. Tech from not well-known engineering college, if applying to some non-technical positions in labor market.
[88] Tabulated data before 1981 does not allow the analysis of caste-class analysis. The 1991 census has no OBC information.

non-agriculture; Labour (agriculture and non-agriculture); Others. In 2012 in rural areas, we have an additional classification of Regular/wage salary classification. In urban areas, there are four employment types, namely- Self-employed, Regular Wage/Salaried, Casual Labor and Other.

Rural Landscape: A look into the Table 2.36 and 2.35 show the percentage of employment type in different caste group i.e. where different castes are employed. There was an increase in agricultural self-employment in 2012 from 2002 across all castes (except Muslims) in rural areas. In 2012, 51% of the ST and FC population were engaged in the Self-Employed in Agriculture category, followed by 45%, 32%, 29% of the OBC, Muslim and SC population. The second biggest employment type is Agricultural Labor, which saw almost 10 pp decline across all caste groups.

Next, I check the share of employment types in different deciles. Self-employed households in agriculture are concentrated in higher deciles, and almost all the agricultural labourers' households are confined to lower deciles (Figure 2.21). There is a clear divide in the type of caste groups in the self-employment category. The wealthy self-employed population is predominantly FC, while the poor self-employed are ST/SC. Within the top wealth decile (where 60% of the total rural wealth is concentrated), there are only 5% SC (and a meagre 1.6% ST) and 34% FC with employment type Self-Employed in Agriculture in 2012 (Appendix Table 2.39).

Urban Landscape: In urban areas, 40% are employed in regular wage/salary. 45% of SC, ST and 50% of FC are employed in the regular wage sector. On the other hand, only 36% of OBC and 32% of Muslims are in wage employment. The next higher employment type is self-employment at 37% in 2002, which declined to 32% in 2012. The decline is observed across all the caste groups. A higher proportion of the SC(25%) and ST(23%) population are engaged in casual labour, increasing from 2002 to 2012. (Appendix Table 2.36 and 2.35)

There is over-representation of FC in regular wage employment for both years. There is an observed drop in the share of FC, and a corresponding increase in OBC share, which again hints towards the re-classification of FC castes with a higher share in regular wage employment. (See Appendix Table 2.38 and 2.37)

In urban areas, self-employed and regular wage earners are predominant in top deciles as we can see below.

F.9. Assortative Matching. There are several prominent characteristics of marriages in India. The role of parents in Indian marriages is very high.[89] Religion and castes are the first checkpoints in arranging marriages. I create a dataset of married couples using the IHDS 2011 dataset by restricting the age of couples between 15-60 years. I remove the couples where even one of the members is retired, students or non-workers. I do not remove housewives from the dataset though more than 70% of women identify themselves as housewives. The final dataset has 34,713 married couples. These couples represent 204 million population using the survey design weights.

F.9.1. *Educational Characteristics of Partners.* Rural couples form 68.6% of the total couples. The average age of men is 40 years and women 35 years. There is age hypergamy (husbands are older than wives) in the society with an age difference of 5 yrs. Age hypergamy is almost at a similar level in both Rural and Urban areas. There is a big chunk of the population with no education. 22% of men and 39.6% of women have zero years of education. Further, 9% of men and 7.2% of women have less than five years of education. For simplicity, I will use the term "Low-Level Education" for these two categories, "Medium Level Education" for the next two categories (5-8 years of education) and "High-Level Education" (more than eight years of schooling). 36% men and 32% women have a Medium level education. 32% men and 22% women have a High level of education. (See 2.42)

F.9.2. *Employment Characteristics of Partners.* Table 2.43 provides the detail on employment characteristics for both men and women. Occupation types and Primary Activity status are present for more men than women. Under *Primary Activity Status* we see 44% of men and 15.3% of women in the Urban area have salaried or professional status. Only 11% of men and 8.5% of women have salaried status in rural areas. These percentages are out of those couples for which information is available. Indeed very few women

[89]To illustrate, in the IHDS 2011 survey, one of the questions asked to eligible married women is- " How long knew your husband before marriage'. 82% of the women's response was less than a month. On the question, "Who chose your husband?",95% responded with Parents or other relatives.

(22.7%) and almost 93% men have this information.

Occupation Type: Almost all the men have some occupation type compared to only 32% women having some occupation. 40% of the men are non-agricultural labourers, and 30% have occupations related to agricultural activity. In comparison, 31% of women (for whom occupation type is provided) are non-agricultural labourers, and 51.9% of women are involved in agriculture-related occupations. The representation in the Professional class is almost the same at 5% for both men and women. As more men and women are involved in agriculture in rural areas, we observe a difference of 20 pp with more women working as farmers/cultivators. It shows the high contribution of women in the agriculture sector in India. In the urban area, a higher share of both men and women are involved in Professional (16% women and 8% men), Sales (13% women and 27.8% men) and Clerical occupations.

Wage Earnings for only 71.3% men and 26.9% women is present in the data, with the average annual wage income of men at Rs 66k and women at Rs. 22k. The annual mean wage for urban men is 2.63 times the annual mean wage of rural men. Similarly, urban women earn 3.7 times more than rural women. Based on imputed wages, the average annual wage goes down as I assign wages to people who are not in wage employment. The average wage drops to Rs 55k for men and 19.7k for women, a drop of 16.6% for men and 11.2% for women.

F.9.3. *India France Comparison.* I compare the educational and earnings level assortative matching between India and France. For France, I use the estimated values (Frémeaux and Lefranc, 2020). The level of assortative matching in India is comparable to France in education. Including the full sample (i.e. 0 years of education couples), the correlation is higher in India (.64) compared to .59 for France. Excluding the cases when both couples have 0 education (22% of men and 40% of women), the correlation in India (.51) is lower than in France. The AM level at social origin (parents' education) is higher in India if we compare all the samples in India. However, excluding zero education cases from Indian data, the coefficients are much lower in India (.113) than in France at 0.51. Further, the correlation in annual wage-earning is higher in India among the couples where both earn.

F.10. Correlation Coefficients. **Pearson correlation coefficient** The Pearson correlation coefficient is the most commonly used (also referred to as Pearson product-moment correlation coefficient). This provides the strength of linear association between two variables. It is given by the covariance between two variables divided by standard deviation of each variable, i.e., for variables X and Y, $\rho_{pearson} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$.

Polychoric correlation coefficient Polychoric correlation is used to estimate the Pearson correlation coefficient between two continuous, bivariate-normally distributed variables from categorized versions of those variables (Hershberger, 2005). This coefficient, thus, measures the association between two ordinal variables. The maximum likelihood method is used for calculation. For the case where two variables are binary, the coefficient is called Tetrachoric correlation.

Spearman's rank correlation coefficient The spirit of Spearman's rank correlation is the same as the polychoric correlation coefficient, i.e. it estimates the correlation between two ordinal (or rank-ordered) variables. However, the intuition is similar to the Pearson correlation coefficient, and the estimation is much simpler than the polychoric correlation coefficient. While the Pearson correlation coefficient measures a linear relationship, it measures a *monotonic* relationship between two variables. No assumptions (e.g. bivariate normality) are required compared to the polychoric coefficient. For ranked variables X_R and Y_R that have no tied ranks (i.e. no same rank is assigned to individuals more than once), Spearman's rank correlation coefficient is given by $\rho_{spearman} = 1 - \frac{6\Sigma_i d_i^2}{n(n^2-1)}$, where d_i is the difference in paired ranks and n is the number of observations. If they have tied ranks, then $\rho_{Spearman}$ is calculated in the same way as Pearson's correlation coefficient.

CHAPTER 3

The Early Origins of Judicial Stringency in Bail Decisions: Evidence from Early-Childhood Exposure to Hindu-Muslim Riots in India

With Sutanuka ROY.[1]

Abstract

We estimate the causal effects of judges' exposure to communal violence during early childhood on pretrial detention rates by exploiting novel administrative data on judgments and detailed resumes of judicial officers born during 1955–1991. Our baseline result is that judges exposed to communal violence between ages 0 and 6 years are 16% more prone to deny bail than the average judge, with the impact being stronger for the experience of riots between ages 3 and 6 years. The observed judicial stringency is driven by childhood exposure to riots with a higher duration of state-imposed lockdowns and low riot casualties.

[1]We would like to thank Guilhem Cassan, Yan Chen, Mathieu Couttenier, Oliver Vanden Eynde, James Fenske, Kareem Haggag, Namrata Kala, John List, Ameet Morjaria, Thomas Piketty, Herakles Polemarchakis, Dominic Rohner, Marc Sangnier, and Rabee Tourky for their comments and support. We would like to thank participants of the Young Economist Symposium (Princeton), Applied Econometrics Conference (Hitotsubashi University), KVS (Leiden University) and seminar participants at the Paris School of Economics and the University of Namur for suggestions and comments. We gratefully acknowledge Sudhir Gupta, Rahil Vora and Yucheng Lu for their research assistance. We thank Lakshmi Iyer for providing us the dataset from their paper Bhalotra et al., 2014.
Sutanuka Roy is currently Assistant Professor at Research School of Economics, Australian National University (sutanuka.roy@anu.edu.au)

A. Introduction

About three million people are held as pretrial detainees worldwide (Walmsley, 2018). The use of pretrial detention as a crime policy tool is motivated by its potential to reduce pretrial flight (Kling, 2006), recidivism (Ribeiro and Ferraz, 2019) via capacitation, and deterrence (W. Dobbie, Goldin, and C. S. Yang, 2018). However, pretrial detention has been associated with an increase in the likelihood of being convicted and in the length of incarceration sentences (Stevenson, 2018), loss of formal employment (W. Dobbie, Goldin, and C. S. Yang, 2018), an increase in the accumulation of debts (Stevenson, 2018), and nontrivial criminogenic effects (Kling, 2006; Leslie and Pope, 2017). Decisions on pretrial detention have consequences for both the defendant and society (Kleinberg et al., 2018) for such detentions typically last for several months with low-income and minority communities bearing a significant portion of the economic costs of pretrial detentions (W. Dobbie and C. Yang, 2021; Henrichson, Rinaldi, and Delaney, 2015). This is of particular concern, since several studies under various settings have documented evidence of judicial stringency and racial disparities in bail decisions (Kleinberg et al., 2018; Arnold, W. Dobbie, and C. S. Yang, 2018). But little is known about the origins of such judicial biases.

In the present paper, we examine the origins of judicial stringency in bail decisions. In particular, we test whether variations in judges' early-childhood exposure to social disorder explain their decisions on bail. The decision on whether a defendant should await trial in jail or at home potentially reflects the trade-offs that judges make between the perceived risks of new crimes that a defendant may commit while awaiting trial out of jail and the incarceration costs (Kleinberg et al., 2018). Therefore, variations in judicial decisions could be driven by differences in either fundamental preference parameters or beliefs that define these trade-offs for judges. In this regard, a growing body of causal early-childhood research in economics shows that early-life exposure to a sociopolitical environment engenders the development of fundamental parameters, such as later-life social preferences (Cappelen et al., 2020), preferences for honesty (Abeler, Falk, and Kosse, 2021) and political identity (Billings, Chyn, and Haggag, 2020), as well as intergroup behavior during adulthood (Couttenier et al., 2019; Fisman et al., 2020). Guided by this literature, we examine whether the early childhood sociopolitical experiences of judges can explain variations in pretrial detention decisions. Motivated by Cappelen et al., 2020, who show that early childhood interventions between ages 3 and 4 years affect later-life preferences for redistribution and views on fairness, we focus on examining the effects of exposure to inter-group religious clashes between the ages of 0 to 6 years, controlling for exposure in later years, on bail decisions.

Our study setting is the Indian judicial system, which has one of the highest shares of pretrial detainees in the world: 70% of the total prisoners in India are under trial, compared with 23% in the United States, 33% in France, and 62% in Pakistan. In particular, we analyze the decisions on pretrial detentions of 668 judges born between 1955 and 1991, who handled 323,380 bail cases from 2014 to 2018 in Uttar Pradesh (UP), the largest Indian state, with a population of 199.81 million (Census of India 2011). According to Prison Statistics of India[2], the percentage of pretrial detainees in UP increased from 70.6% in 2015 to 72.5% in 2019, in line with the national trend.[3] Owing to the high pendency rates of cases in courts, about 32% of pretrial detainees in UP remain incarcerated for more than a year, compared with the national average of 25%. This fact is striking, since "Bail is rule, jail is an exception" was established as a legal principle by the Supreme Court of India in a landmark judgment (State of Rajasthan v. Balchand alias Baliya) in 1978.

We focus on the Hindu-Muslim riots in UP because such inter-group clashes are not specific to the state of UP but occur throughout India.[4] They are recurrent events that continue to plague the country and potentially have unmeasured consequences in terms of social trust, social segregation, economic damages, and human capital depletion (Mitra and Ray, 2014). These riots have claimed 6,565 lives, injured 21,429 people, and resulted in 87,903 arrests between 1950 and 2000, with an average of five days of lockdown per riot.

Our conflict dataset includes data from two sources. One is that Mitra and Ray, 2014 for the period 1950–2000.[5] The other is a novel dataset with data we collected on lockdowns during riots, which we sourced from the original historical newspaper articles that Varshney and S. Wilkinson, 2006 used to prepare their dataset. We use a novel

[2]These data are for 2014–2019 and are published by the National Crime Records Bureau.

[3]In India, the percentage of pretrial prisoners has increased from 67.6% in 2014 to 69.1% in 2019

[4]It would also be interesting to study political emergencies. However, these are usually aggregate shocks with not much within–country or within–state variation.

[5]This dataset includes the dataset of Varshney and S. Wilkinson, 2006 for the period 1950–1995, which has been used in several studies, such as in those by Fisman et al., 2020; Sarsons, 2015.

dataset on judiciary officers[6], which we retrieved from state administrative records containing information on judges' date of birth, home district, date of recruitment, judicial posting details, and academic qualifications. We combine the riot data with these administrative records to ascertain judges' exposure to conflict in early childhood. We obtained all the original bail judgment files for 2014–2018 from the judiciary website and extracted case-level characteristics and bail decisions. Our extracted sample consists of 423,000 bail applications from the entire pool of cases that we downloaded (two million cases). We link these data with judiciary officers' data to arrive at judge-level panel data and the pretrial detention rate (which equals the total bail cases denied/the total bail cases assigned) as our primary outcome variable.

We adopt the empirical approach used in the recent seminal empirical studies on the impact of early exposure to violence that have research setting identical to ours: Couttenier et al., 2019, which estimate the causal impact of early childhood exposure to conflict on asylum seekers criminal behavior, and Fisman et al., 2020, which provides causal estimates of the exposure to Hindu-Muslim riots in India on bank managers' lending decisions. Following the above-mentioned literature, we identify the effect of riot exposure using two key variations. The first variation is based on the variations in *early childhood riot exposure* of judges born in same home-districts but belong to different birth-cohorts, as well cross home-district variations in *early childhood riot exposure* within the same birth-cohorts. The second variation relates to the exogenous rotation policy of judiciary officers -some judges with riot exposure and others not- during *adulthood*. The second variation experienced by judges during *adulthood* allows us to distinguish the effect of riot exposure from the location attributes of the district assigned to the judges as well as control for unobserved time-varying differences in the districts.

Our research setting and data have several unique features that allow us to identify causal effects on judicial decisions. First, our focus on Hindu–Muslim riots provides substantial within-region variations in early childhood exposure to social disorder. Further, the recurrent nature of these riots allows us to test for the robustness of their impact across generations. Second, our analysis of exposure to conflict between the ages of 0 to 6 years helps us to rule out self-selection into violence exposure. One concern could be about systematic relocation decisions by families owing to the communal riots; for example, judges in non-exposed home districts could be affected by other families migrating

[6]We use the terms judge and judiciary officer interchangeably.

to their districts in response to violence. This could lead to a violation of identifying the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980). We test for the possible violation of our identifying assumption using migration data. We find that the district-years affected by the communal riots do not have differential rates of migration. In addition, we collect data on districts where the judges completed their secondary school (age 15), high school (age 17), and undergraduate (age 21) studies. We show that there is no treatment effect of early-childhood exposure to riots on migration from home districts to the districts where they completed their secondary schooling degree, their higher secondary degree, or their undergraduate degree. Further, we use home-district fixed effects to account for family selection into location. We include home-district time trends to filter out current trends in home districts affecting judicial decisions from those affecting through early-riot exposure. Third, we exploit an exogenous rotation policy for the judicial officers in UP, which generates plausibly exogenous spatial variations in the judicial postings of riot-exposed and not-exposed judges away from their home districts during adulthood. This approach allows us to isolate the effects of riot exposure from the attributes of the district (where the judge is posted) and the home district. It also rules out the self-selection of judges into less crime-prone districts. Fourth, since cases are exogenously assigned, it mitigates the possibility of judges selecting into particular types of cases. We implement several empirical strategies to show evidence of no differential selection along a wide range of cases characteristics, judge peer characteristics, and judge characteristics. Fifth, bail is an ideal outcome to detect bias because it is purely discretionary and the judges are required to make decisions with limited information and almost no interaction with the defendants.

We find that exposure to communal violence when aged 0–6 years causes an increase of 6 percentage points ($p < 0.01$) in the share of pretrial detentions, which is an increase of 16% compared with the mean. The effect is robust to the use of various estimation techniques; the inclusion and exclusion of controls such as judge-level covariates (e.g., gender, religion, performance in the Bachelor of Law (i.e., LLB) examination, and on-the-job experience); the use of placebo checks; and the removal of outliers. More importantly, we control for time-varying share of Muslim population, urban population, share of Muslims in the urban population, and log of total population, which are shown to be predictors of occurrences of Hindu-Muslim ethnic violence (Corbridge, Kalra, and Tatsumi, 2012). Further, we sort the judges by their influence on the regression coefficient and remove them one by one to test whether a few judges are driving our results, but do not find evidence in this regard. A key concern could be that it is difficult to assess whether exposure to civil conflict in early childhood directly affects any given judges' preferences or is associated with changes in the selection of who ultimately becomes a judge. To address this concern, we present the results of two exercises. First, using a representative sample of the working population from an independent data source (the 66th round of the Employment and Unemployment Survey in 2011 by the National Sample Survey Organization (NSSO)), we find that the share of the early-childhood (0–6 years old) riot-exposed population in the entire working population in UP (in all sectors and all types of employment) is close to the share of the early-childhood riot-exposed judges in UP. We find no under- or over-representation of this riot-exposed population in the sample of judges. Second, we test whether the number of judges or, the number of judges of a particular gender or religion are disproportionately, drawn from a riot-affected home district-year. We do not find any statistically significant difference between the proportions of early-childhood riot-exposed versus non-exposed judges by home district-year along gender or religion. Hence, we can rule out the possibility of selection into the judiciary due to childhood exposure to communal violence.

A part of the total effect on bail decisions could be driven by differences in the ability of the judges. We find that ability measured as the division[7] obtained in the LLB examination does not explain the increase in the pretrial detention rate. Guided by the active economics literature on endogenous preference formation, which shows early childhood as a formative period for social and political preferences,[8] we explore the behavioral explanations of the early-childhood exposure effect.

We find that high-intensity state interventions, such as a high duration of lockdowns and a high number of arrests, which are associated with limiting the riot casualties, explain the increase in the observed pretrial detention rates. This finding suggests that early-childhood exposure to state-imposed lockdown measures that proved effective in containing violence possibly generated higher support for the institutions of the state in law-and-order matters. Further, we do not find any evidence of religious bias in the observed stringency in the bail decisions of early-childhood riot-exposed judges, which

[7]We classify the grades obtained in the LLB degree course into three divisions: I (Grades: $\geq$ 60%), II (Grades: $\geq$ 45% and $<$ 60%), and III (Grades: $\geq$ 33% and $<$ 45%).

[8](Kohlberg, 1984; Piaget, 1997; Harbaugh, Krause, and Vesterlund, 2002; Sutter and Kocher, 2007; Fehr, Bernhard, and Rockenbach, 2008; Almås et al., 2010; Bauer, Chytilová, and Pertold-Gebicka, 2014; Angerer et al., 2015; Ben-Ner et al., 2017; Cappelen et al., 2020)

rules out the possibility of the inter-group hostility mechanism underlying the effect. In line with Cappelen et al., 2020, who found that early interventions between ages 3 and 4 years have lasting effects on social preferences, we find that exposure to violence between ages 3 and 6 years, robust to multiple hypothesis testing, is the key driver of the observed judicial biases toward pretrial detentions.

This paper makes contributions to several strands of literature. Its' first contribution is to provide the first causal evidence, to the best of our knowledge, on the early origins of judicial bias. We expand the rich literature on judicial bias by providing evidence on the long-term determinants of judicial decisions. Our focus on linking interventions during the formative years of judiciary officers with stringency in their decisions relates particularly to the emerging evidence on the impact of early-childhood interventions on long-term social preferences, such as that found by Gould, Lavy, and Paserman, 2011; Giuliano and Spilimbergo, 2014; Cappelen et al., 2020; Billings, Chyn, and Haggag, 2020, more broadly, and on the impact of early-childhood exposure to violence on inter-group behavior, in particular (Couttenier et al., 2019; Fisman et al., 2020). We contribute to this literature by examining the effects of bureaucrats' early childhood exposure to violence on their public service decisions. Our study also adds to the robust empirical evidence on the influence of early-childhood interventions on various long-term outcomes, such as cognitive skills (J. J. Heckman, 2006; Bleakley, 2007; Almond, Edlund, and Palme, 2009; Maccini and D. Yang, 2009; Aizer and Cunha, 2012; Bharadwaj, Løken, and Neilson, 2013), health outcomes (Currie, 2009; Maccini and D. Yang, 2009; Almond and Currie, 2011; Currie and Vogl, 2013; Adhvaryu, Fenske, and Nyshadham, 2019), and labor market outcomes (Almond, 2006; Bleakley, 2010; Gould, Lavy, and Paserman, 2011).

Second, our analysis reveals that human capital achievements, as measured by the division achieved in the LLB examination, do not explain the observed pretrial detention rates of early-childhood riot-exposed officers. Heterogeneity analyses indicate that these observed biases are possibly driven by behavioral effects. Our results add to the literature that demonstrates the importance of early investments during the formative years in generating noncognitive outcomes (J. J. Heckman, 2007; Cunha, J. J. Heckman, and Schennach, 2010; J. Heckman, Pinto, and Savelyev, 2013), such as motivation, dependability (J. J. Heckman, 2006), and distributive preferences (Cappelen et al., 2020), which have economic consequences independent of cognitive achievements (J. J. Heckman and Rubinstein, 2001; J. J. Heckman, Stixrud, and Urzua, 2006).

Studies on judicial bias are now common in the literature on the economics of crime, which mostly focuses on the Organization for Economic Co-operation and Development (OECD) countries. A large body of literature focuses on the criminal justice system in the United States (W. Dobbie and C. Yang, 2021; W. Dobbie, Goldin, and C. S. Yang, 2018; Kleinberg et al., 2018; Kling, 2006; Stevenson, 2018; Agan and Starr, 2018; Arnold, W. S. Dobbie, and Hull, 2020; Doleac, 2021). Thus, our third key contribution is expanding this empirical examination using data from a country with a weak institutional context. Given the large share of pretrial detainees in Indian prisons, the examination of judicial bias in India is important in understanding the potential welfare consequences of institutional imperfections. Our study highlights the presence of bias in bail decisions in India, which adds to a recent study on India that found no in-group (by gender or religion) bias in judicial sentencing (Ash et al., 2021).

Lastly, this paper also contributes to a growing body of evidence on extraneous factors in judicial decision-making such as judicial distortions due to media (Lim, 2015), Presidential selection (Mehmood, 2021), religion (Shayo and Zussman, 2011, gift-giving (Bakhtawar and Mehmood, 2022), rituals (Mehmood, Seror, and Chen, 2021) and race (Bielen, Marneffe, and N. Mocan, 2021).

An understanding of the causal processes that shape social preferences is of interest to academics and policymakers alike. Our study reveals the importance of sociopolitical institutions early in life in shaping long-term outcomes. More crucially, we show that the impact of early-childhood exposure to institutions is robust across generations, that is, regardless of whether the judiciary officer was born in 1955 or 1980.

The remainder of the paper is organized as follows. Section 2 presents the context of the study and the data used. Section 3 explains the empirical strategy, and Section 4 presents the results of the balance test. Section 5 presents the results of the core analysis and a series of robustness tests. Section 6 explores potential mechanisms underlying riot effects. Section 7 concludes.

B. Context and Data

Our study setting includes the judiciary in UP, the largest Indian state, which has a population of about 199.81 million (Census of India, 2011)[9] and a demographic composition that is similar to that of India as a whole.[10] Next, we explain the unique features of the data and the study setting that allow us to estimate the causal effects of violence on bail decisions.

B.1. Indian Judiciary System. The Indian judiciary can be divided vertically into three levels. The apex court, the Supreme Court of India, is based in New Delhi. Its jurisdiction encompasses the entire country. Next in the hierarchy of courts are the High Courts. They are the highest court at the state level, and their jurisdiction is limited to the state boundaries. The third level is the district-level courts, at the district level, and their jurisdiction is restricted to the district.[11] Within district-level courts, the District and Session Court has appellate jurisdiction over all the other courts, such as civil, criminal and family courts.

The district-level courts in UP have an average of 27 judges per district (i.e., 2,048 unique judges in 75 districts in August 2018). Districts are the smallest administrative division in India to which the authority of law and order are delegated.[12] The total number of judges per million population is 9.1, and the average age of judges is 43.84 years. UP has 22.6% female judges, and 6.9% of its judges are Muslims.

B.2. Rotation Policy for Judicial Officers in UP. The UP judiciary follows an explicit geographical rotation policy with the stated objectives of reducing corruption and collusion in the judiciary, which induces exogenous spatial variations in the distribution of officers across districts. Generally, the tenure of judicial officers is 3 years of service in the district.[13] District judges are posted away from their hometown district

[9]See censusindia.gov.in

[10]Hindus and Muslims form 79.73% and 19.26% of UP's population, as against the national average of 79.8% and 14.2%, respectively.

[11]Some newly created districts do not have courts and rely on the services of courts in adjacent districts.

[12]District officials include an Indian Administrative Officer, tasked with administration and revenue collection, a Superintendent of Police, tasked with maintaining law and order, and a Deputy Conservator of Forests, tasked with maintaining environmental management. As per the Census of India, 2011, the country had 640 districts.

[13]The guidelines for transferring officers appear in circulars issued by the Registrar General of the High Court of Judicature at Allahabad. The tenure is of 2 years at an outlying court (courts far from district headquarters) or at Sonbhadra district.

by the rotation policy, whereas the following norms guide the transfers of other judicial officers:

(1) Officers will not be posted to their hometown.
(2) They will not be posted within 6 years to a district in which they were earlier posted.
(3) They will not be posted within 3 years to any district falling in the zone[14] in which they were earlier posted.
(4) They will not be posted to any adjoining district of another zone.
(5) The constraints on re-posting of officers in the zone will not apply if they had been posted for a short period of less than 6 months.

The Assistant Registrar also collects information on the list of stations in UP where judiciary officers have close relatives and a statement of places where they were educated, as required under C.L. No. 25/Admin (A)/DR(S)/78 dated March 16, 1978.

We observe that officers are located away from the hometown district. An average officer reallocation assigns a judge to a new district that is 325 km (std. dev. 165 km) away from the district of the previous assignment. The main advantage of the rotation policy is that it induces matching between judges and defendants that is plausibly uncorrelated with bail cases. In the next section, we explain the detailed data on judiciary officers, which we use to test the plausibility of the exogeneity of the judiciary rotation policy of the state.

B.3. Data on Judiciary Officers. We extract information on working and retired judges[15] from the Allahabad High Court website.[16] The collected data include details on judges' date of birth; their home-district; the dates on which they were promoted; their educational qualifications, dating back to the first school-leaving examinations; and the dates and locations of their postings and transfers. We use the data on home districts and the date of birth of the judges and match it with the data on riots to compute their exposure to riots at every age. Since some district boundaries in our sample have undergone changes over 50 years,[17] we first harmonize the districts in the two datasets by

[14]A zone is a collection of districts. The entire state is split into seven zones of contiguous districts.
[15]A few judges who judged cases during 2014–2018 retired during this period.
[16]`http://www.allahabadhighcourt.in/District/Officer/judge_id.html`, where the judges' ID is their unique identification. Since the information on retired judges has been removed from the Allahabad High Court website, we extract data from the archived web page.
[17]The total number of districts in UP is currently 75 and was 48 in 1950. Further, in 2000, a new state, Uttarakhand, was carved out of UP.

assigning every district to their parent (origin) district. We use the official census district (2011) records to trace the origin of every district in our data. Appendix Table 3.8 details the district formation and Appendix F.1 provides complete information on district harmonization.

B.4. Bail Jurisprudence. The fundamental right enumerated in Article 21 of the Constitution of India is that "No person shall be deprived of his life or personal liberty except according to procedure established by law." This right forms the basis of bail provision in India. Although bail is not defined legally in Indian codebooks,[18] it implies the release of a person detained by the police for a certain offence, by furnishing a guarantee of future attendance in the court for trial. The two categories of bail in India are as follows:

i) *Bail in bailable offences*: In Section 436 of Cr.P.C, bail is the right of a person who has been accused of committing an offence that is bailable in nature. This provision casts a mandatory duty on police officials as well the court to release the accused on bail if their alleged offence is bailable in nature.[19]
ii) *Bail in nonbailable offences*: When a person is charged with having committed a non-bailable offence(s), the court has to consider many factors:[20]
a) whether there is any prima facie or reasonable ground to believe that the accused had committed the offence; b) the nature and the gravity of the accusation; c) the severity of the punishment in the event of conviction; d) the danger of the accused absconding or fleeing; e) the character, behavior, means, position and standing of the accused; f) the likelihood of the offence being repeated; g) a reasonable apprehension of witnesses being influenced; h) the danger of justice being thwarted by granting bail.
The subjective nature of the factors a judge must consider during bail decisions is evident.[21] When a court gives bail, the accused must sign a personal bond and must usually provide two surety bonds (from relatives or others who can vouch for the defendant) for

[18]The Criminal Procedure Code (Cr.P.C.) details the bail process but does not define bail. All offences are categorized as bailable or nonbailable.
[19]In 2005, the Cr.P.C. was amended by adding section 436-A: A person who has undergone detention for a period that is half the maximum period of imprisonment imposed for a particular offence shall be released on his/her personal bond with or without sureties.
[20]State of U.P. through CBI v. Amarmani Tripathi, 2005 (8) SCC 21; Prahlad Singh Bhati v. NCT, Delhi & Anr. 2001 (4) SCC 280; Ram Govind Upadhyay v. Sudarshan Singh & Ors., 2002 (3) SCC 598.
[21]We do not examine Anticipatory Bail. Under Section 438 of the Cr.P.C., the High Court or Court of Sessions can issue bail before a person is arrested, which is known as Anticipatory Bail, if there is an apprehension or a reason to believe that a person may be arrested on an accusation of having committed a nonbailable offence. The court considers the same list of factors mentioned in point (ii).

a certain amount. If the accused breaks the bail condition, the court is liable to recover the amount from the defendant. A granted bond can also be cancelled later if it is found that bail conditions are not complied with.

B.5. Data on Defendants and Cases Registered. We web-scraped all the case-level pdfs from the district e-court website by court establishment[22] in August 2018. We segregated about 423K bail cases[23] from the entire pool of two million downloaded cases. We performed optical character recognition, translated the documents to English and then extracted all the relevant variables at the case level using text analysis. The primary details we extracted are the bail decision (whether granted/denied), the name of the defendants (which is used to identify their religion, following Bhalotra et al., 2014),[24] and the criminal section codes under which a case is registered. The criminal section codes pertain to either the Indian Penal Code (IPC: the comprehensive list of offences and associated punishments) or special laws (the Acts to augment the IPC). We created 11 crime categories from these criminal section codes, mostly following the chapters of the IPC codebook.[25] Appendix F.2 provides detailed information on the procedure we adopted.[26]

B.6. Hindu-Muslim Communal Riots: 1950-2000. The data on Hindu–Muslim riots are from two sources: the datasets of Varshney and S. Wilkinson, 2006 and Mitra and Ray, 2014. The combined dataset provides detailed information on the Hindu–Muslim riots in India as reported by a national English daily, *The Times of India*. We use the information on the district, month, and year to identify a unique riot.[27] For each recorded clash, the dataset also has information on the riot duration, the number of people killed

[22]Website: https://districts.ecourts.gov.in/up; the District Courts for Chandauli, Etawah, Hardoi, Kheri, Pratapgarh, and Sant Kabir Nagar districts have not uploaded judgments. In the district of Varanasi, very few bail cases have been uploaded.

[23]The bail cases are identified from the bail application marker provided with the case number.

[24]The accuracy of the algorithm is in the range of 5-6%. Details are provided in the Appendix F.3

[25]Arms and Explosives, Body Crime, Cow Slaughter, Electricity Theft, Gangster and Dacoity, Property Crime, Forgery, Criminal Intimidation, Public Tranquility, Public Health, and Other

[26]Since the lengthy process of text extraction could entail errors, we manually digitized all the variables for 60k cases—30k bail cases handled by Muslim judges and an equal number of randomly chosen cases handled by Hindu judges—and show the error rates for each variable extracted (see Appendix F.4 for the selection of the cases). The measurement error is 5% in the bail outcome which is the main outcome variable, and we show that it is not correlated to our main explanatory variable (Details in Appendix F.5)

[27]For some entries, district information is missing, but city/village names are provided. We use this to assign districts to a riot.

or injured, and the number of people arrested. Further, we add the duration of lockdown for each riot from the source articles of Varshney and S. Wilkinson, 2006.

In line with the literature (Couttenier et al., 2019), in our main specification, we use the extensive margin of exposure to conflict, which is a dummy variable of exposure to communal violence between the ages of 0 and 6 years that takes the value 1 if the home district of the judiciary officer experienced Hindu–Muslim communal clashes when the officer was 0–6 years old. Appendix Table 3.9 shows that 31.4% of the judges have been exposed to violence when aged 0–6 years.[28].

Since the judges in our sample have UP as their home state, our treatment variations in conflict exposure derive from the variations in communal clashes in this state. In all, 33 of the total of 48 home districts have experienced at least one riot during 1950–2000.[29] Within the districts experiencing at least one riot, the mean of the number of riots per year across districts is 0.12 with standard deviation 0.38. During 1950–2000, the average number of Hindu–Muslim riots per year per state was 7.6 in India as a whole and 8.2 in UP. The average number of deaths and injured per year per state are 50 and 140, respectively, for the entire country and 65 and 116 for UP. In terms of state response variables, the average duration of state-imposed lockdowns following a riot was 5 days both in UP and in India as a whole, the average number of arrests was 144 in UP. In terms of the intensity of violence in each riot, 6.7 people were killed, on average, in a riot in UP, which is similar to the average (6.3) for the whole of India.

We set the following restrictions to arrive at our sample of judges. Since bail outcomes are our main outcome variable, we retain judges who are assigned to bail cases ($N = 1,268$), of which the names of the home districts and the home states of 35 judges and 67 judges, respectively, were not available in the administrative data. Following Arnold, W. S. Dobbie, and Hull, 2020, we drop judges who were assigned too few cases (the bottom 5 percentile in terms of the number of cases (<97 cases) assigned to the judges, which amounts to 493 judges).[30] We also drop 10 judges because we did not have information on their LLB examination results. The LLB degree is the minimum

[28]The proportion of exposed bank managers in the study by Fisman et al., 2020, based on at least one death in the riot, is 14.4%

[29]Since the treatment in our case starts from 1950, we use the districts that were present in 1950 by merging the districts as detailed in Appendix Table-F.1. Currently, there are 70 districts in Uttar Pradesh.

[30]Our results are robust to changing the threshold from the bottom 1 to the bottom 10 percentile; we provide the results of the robustness checks in Section 7. The choice of the bottom 5 percentile as the threshold,, that is, judges handling a minimum of 97 cases in 4 years, is to maintain a balance between not dropping too many judges and not keeping too many judges who have handled very few cases.

requirement for a judiciary officer, and 29% of the judges had passed this course with first division and 71% with second division.The final sample consists of 668 bail judges handling 323,380 cases aggregated at the judge-district-quarter level, which yielded a sample size of 5,530. The descriptive statistics presented in Appendix Table 3.10 show that the analysis sample is similar to the total sample of judges, along observables in the data. Figure 3.5 reveals that the majority of the district judge transfers occurred mostly in the second quarter. Figure 3.6 reveals that the pattern is similar for the sample of bail judges. In Appendix Table 3.9, we show that judges exposed to communal violence between the ages of 0 and 6 years are similar to judges without such exposure along the district covariates and the judge peers assigned to them. In terms of judges' characteristics, the judges exposed to communal clashes in early childhood are more likely to be female, older, and more experienced.

C. Empirical Strategy

C.1. Specification. Following recent empirical studies estimating the causal impact of early childhood exposure of violence,[31] we identify the effect of riot exposure based on the two variations. First, we exploit the variations in *early childhood exposure* of judges born in same home-districts but belong to different birth-cohorts, as well cross home-district variations in *early childhood exposure* within the same birth-cohorts. Second, we exploit the variation generated by the exogenous rotation policy of judiciary officers *during adulthood*, which allows us to distinguish the effect of early childhood riot exposure from location attributes of the districts assigned to the judges as well as control for time-varying unobservables affecting bail outcomes, such as variation in crimes registered in districts.

We begin with our case-level data, where the unique identifier is a bail case. Each bail case is uniquely matched to a judge; that is, only one judge handles each bail case. We observe the bail decisions of a judge corresponding to each bail case. For each judge, we aggregate the pretrial detention rate at the district-quarter level. We conduct our analysis of pretrial decisions using the judge–district–quarter level data.[32] Our results

[31] Couttenier et al., 2019; Fisman et al., 2020

[32] We address concerns about the clustering of bail decisions at the judge level by first aggregating case-level outcomes at the judge–district–quarter level (Bertrand, Duflo, and Mullainathan, 2004). Noting that our treatment variation is at the judge level (Abadie et al., 2017) and that there may be correlations across outcomes for a judge across quarters and districts (Bertrand, Duflo, and Mullainathan, 2004), we cluster our standard errors at the judge level.

are robust to case-level regression. Section 7, the robustness checks section, explains why we have judge-level regressions rather than case-level regressions as our preferred specification. Guided by our data depicted in Figure 3.7 in the Appendix, which shows that the proportion of bail cases trends in quarterly periods in the data, we aggregate bail decisions at the quarterly level. Our results are robust to aggregation at the monthly level.

Our key econometric specification is as follows:

$$B_{j,d,t} = \alpha + \eta_{d,t} + \delta_{h,t} + \mathcal{F}.\mathcal{E}_b + \beta \times kid[0-6]_j + \sigma X_j + \sum_{k=7}^{9} \gamma(k) \times exposure(k)_j + \epsilon_{j,d,t} \quad (8)$$

where $B_{j,d,t}$ is the share of bail denied by judge "j" assigned to district "d" at quarter "t". The covariate in the regression, $kid[0-6]_j$, is the binary variable of exposure to communal riots when aged 0–6 years. The variable $expo(k)_j$ is the exposure to violence at the k th year of a judge j. α is an intercept.

We next explain how we control for unobserved differences between riot-exposed and non-exposed judges that could potentially confound our riot-impact. First, riots may be triggered in certain years when certain political parties are in power. Thus, judges born in riot-years may pick up not only riot exposure effects but also the effects of unmeasured socio-political preconditions correlated with religious riots. We include birth-cohort fixed effects, $\mathcal{F}.\mathcal{E}_b$, which control for unobserved differences by birth year. Second, we include $\delta_{h,t}$, which are the home-district-quarter fixed effects, to filter out the current trends in home districts affecting judges' preferences from those affecting through early-childhood exposure. Some districts could have more cases registered in certain quarters because the police were more active and successful in those districts at those times. Districts could also vary by types of crime committed owing to seasonal weather shocks (Blakeslee and Fishman, 2018). Although, we establish in our balance tests that judge rotation policy yields randomization of judges into districts and cases, spatial and time-varying differences in the detection and registration of crimes in districts assigned to judges are additionally accounted for through district-quarter fixed effects $\eta_{d,t}$. We also account for time-varying share of Muslim population, urban population, share of Muslims in urban population, log of total population, which are shown to be predictors of the occurences of Hindu-Muslim ethnic violence (Corbridge, Kalra, and

Tatsumi, 2012).[33] X_j is a vector of judge-level characteristics, such as religion, gender, division obtained in the LLB examinations, and on-the-job experience.

In our setting, the policy-induced exogenous rotation of judicial officers addresses endogeneity concerns related to judges selecting into districts and hence types of cases, as well as generates substantial heterogeneity across birth cohorts in their early-childhood exposure in all home districts such that the $kid[0-6]_j$ and $expo(k)_j$ dummies are not collinear with the home-district fixed effects, which allows us to separate unobserved confounders that vary at the home-district level. In addition, our focus on the exposure of judges to violence when aged 0–6 years alleviates the endogeneity concern of self-selection into conflict, whereas home-district fixed effects account for household selection into riot-exposed districts. Standard errors are clustered at the judge level to account for within-judge correlations in bail decisions over time and across the assigned districts.

In equation 8, our coefficient of interest β denotes the difference in bail decisions between judges exposed and not exposed to communal riots in early childhood (0–6 years old). We augment the equation by including a dummy variable that takes the value 1 if there was a riot 1–5 years before a judge was born, to control for the direct effects of pre-birth exposure to conflict.

The riot information is available until 2000, and in our sample, the youngest judge is born on October 7, 1991. Hence, we can calculate exposure to violence up to the first 9 years after birth for the full sample of judges in UP.[34] However, as a robustness check, we test for early-life exposure to violence by controlling for exposure to violence in later life for the subsample for which we can control for exposure to violence in later years.

D. Balance Test

In this section, we test whether the exogenous rotation policy of judiciary officers resulted in selection along observables. One likely concern is that riot-exposed judges select into cases involving specific types of crimes. We check for differences between riots-exposed and not exposed judges across a host of case attributes using two empirical strategies. Our first empirical specification is as follows:

[33]These variables are at home-district year level which comes from the decennial census data for 1961, 1971, 1981, 1991 and 2001. The yearly level values are interpolated from these values.

[34]Controlling for the later years of riot experience is possible only at the cost of sample size reduction. For example, if we add exposure to violence up to the first 10 years, we will have to drop judges born in 1990 because we do not have information on riots in 2001.

$$Y_{j,d,t} = \alpha + \eta_{d,t} + \delta_{h,t} + \mathcal{F}.\mathcal{E}_b + \beta \times kid[0-6]_j + \sigma X_j + \sum_{k=7}^{9} \gamma(k) \times exposure(k)_j + \epsilon_{j,d,t} \quad (9)$$

where $Y_{j,d,t}$ are the characteristics of cases at the level of judge j assigned to district d at quarter t. The set of covariates $kid[0-6]_j$, $exposure(k)_j$, α, and $\delta_{h,t}$ are the same as those in the core econometric specification in equation-8.

We show that there is no treatment effects on case characteristics. In other words, conditional on fixed effects, there is no selection of riot-exposed judges into cases. Column (1) of Table 3.1 shows that there is no statistical difference in the total number of cases assigned to the early-childhood riot-exposed judges and not-exposed judges. Columns 2-10 show no treatment effects along the following case characteristics - share of cases with Muslim defendants, share of non-bailable cases, share of cases booked under special Acts, share of cases booked under 1 IPC section, share of cases booked under 2 IPC sections,share of cases booked under 3 IPC sections,share of cases booked under 4 IPC sections,share of cases booked under 5 IPC sections and share of cases booked under 6 or more IPC sections. All the coefficients are small and insignificant showing that exposed and non-exposed judges handle similar type of cases.

We define crime categories using the IPC, which is the official criminal code of India, and the special Acts passed by the central and state governments (see Appendix-F.2.iii for details). We show in Appendix Table 3.11 that case assignment (based on the 11 types of crime categories explained in the data section) is not correlated to the exposure variable. Considering the potential concern about measurement error due to errors in the data extraction of crime categories, we test for selection in a manually digitized random sample of judges (Appendix Table-3.12) and find no evidence of the selection of judges into crime types.

Possibly a more convincing way is to show that there are no pre-trends in the case composition corresponding to the timing of judge rotation. Our second empirical specification is as follows:

$$Y_{j,d,t} = \alpha + \eta_d + \delta_h + \mathcal{F}.\mathcal{E}_b + \sum_{k=-12}^{12} \beta_k \times kid[0-6]_j \times months_k + \sigma X_j + \sum_{k=7}^{9} \gamma(k) \times exposure(k)_j + \epsilon_{j,d,t} \quad (10)$$

where $Y_{j,d,t}$ are the characteristics of cases at the level of judge j assigned to district d in month t, where the month is relative to the transfer date of the judge in the district. The set of covariates: $kid[0-6]_j$, $exposure(k)_j$, α, and $\delta_{h,t}$ are the same as those in the core econometric specification in equation-8. Though we are adding 12 months prior and after the event of judge rotation, the table reports the coefficients only for 6 months prior and after the judge rotation events. Appendix Table 3.13 and 3.14 present the results. We find that no pre-trends around the transfer date of a judge in terms of case characteristics.

Next, we apply the method used by Couttenier et al., 2019 to demonstrate the exogeneity in the allocation of judges determined by the rotation policy. The notion is to test whether the judicial postings across different district-quarters is non-random. More specifically, we test whether the average characteristics of the judges from the same home district are similar to those of the judges (from the same home district) posted in different district-quarters.[35] We test for the difference in means along the judge-level treatment and non-treatment covariates across district-quarters. Formally, we estimate the following equation separately for judges from each home district for every quarter:

$$J_{h,b,q,d} = \sum_{d=1}^{75} \beta_{d,q} \times \mathcal{I}_{h,b,q,d} + \epsilon_b \tag{11}$$

where $J_{h,b,q,d}$ are the judge-level characteristics of judges from home district "h", birth cohort "b", at quarter "q" in district "d". $\beta_{d,q}$ are the district-quarter specific coefficients corresponding to the indicator dummy for judges, denoted as $\mathcal{I}_{h,b,q,d}$, that takes the value 1 if the judge from the home district "h", birth cohort "b", is allocated to district "d" at quarter "q". The dependent variables are judge-level characteristics, such as exposure to communal conflict when aged 0–6 years, gender, religion, age, time to promotion, and age when joining the judiciary. For each home district, we examine the number of district-quarters for which the F-test of the null hypothesis $\hat{\beta_{d,q}} = \hat{\mu_h}$ is rejected where $\hat{\mu_h}$ are the average characteristics of the judges at the home-district level. If the allocation is exogenous, then the district-quarter specific coefficient $\hat{\beta_{d,q}}$ should not differ from the home-district average, and the F-test should not be rejected for this district-quarter. If there is no selection in the spatial allocation of the judges, then the observable judge

[35]Suppose X judges are from home district A and the average age of these judges is 46 years. Out of these X judges, say x1 judges are posted in district B and (X-x1) judges are posted in district C at a given time. If the judges are posted randomly, then the average age of x1 and (X-x1) judges would also be 46 years.

characteristics in some districts with respect to the home-district average should not be over- or under-represented. Each row in Table 3.15 represents the share of home districts for which the F-test is rejected at the 10% cutoff in at most 0, 1, 5, and 10 districts. For instance, it shows that for 95% of the home districts, we do not have any district-quarter specific coefficients that differ from the home-district average of $Kid[0-6]$ and for 100% of the home districts, less than five district-quarter coefficients differ from the home-district average. We observe similar results for the home-district averages related to the average of female judges, Muslim judges, and judges with first division in their LLB examination. Regarding the home-district average age of judges, time to promotion, and joining age, for almost all home districts, less than 10 districts have district-quarter specific coefficients that differ from the home-district averages. Therefore, we do not find any evidence of selection along observables in the spatial allocation of the judges.

Eren and N. H. Mocan, 2020 show that peers of judges influence judicial decisions. Columns 1–6 in Appendix-Table 3.16 present the results of a leave-one-out regression we run to test whether the peers assigned to judges differ by early-childhood exposure to riots. We estimate the following equation, that is, equation-8, without district-quarter fixed effects with covariates at the level of districts assigned to judges as the outcome variables.

$$Y_{j,d,t} = \alpha + \delta_{h,t} + \mathcal{F}.\mathcal{E}_b + \beta \times kid[0-6]_j + \sigma X_j + \sum_{k=7}^{9} \gamma(k) \times exposure(k)_j + \epsilon_{j,d,t} \tag{12}$$

where $Y_{j,d,t}$ are the peer characteristics assigned to judge j at quarter t. The set of covariates $kid[0-6]_j$, $exposure(k)_j$, α, and $\delta_{h,t}$ are the same as those in the core econometric specification in equation-8. We find that there is no statistically significant difference between peer groups assigned to judges by early-childhood exposure along dimensions such as religion, age, and on-the-job experience. However, the group of early-childhood exposure judges do have 10% fewer female peer judges than the group of judges with no such exposure to violence. Female judges are correlated with high pretrial detention rates. We account for the differences in peer attributes with district-quarter fixed effects in our empirical specification measuring the causal impact of early riot exposure.

E. Main Impact of Exposure to Communal Violence

We present the results of our key econometric specification as represented by equation-8 in Panel A of Table 3.2. We report our coefficient of interest β controlling for exposure to violence in later years. The coefficient of interest demonstrates the causal effect of exposure to communal violence when 0–6 years old on the shares of bail denied (that is, pretrial detention rates), where the control group consists of judiciary officers with either no experience of violence or who have not been exposed to violence when 0–6 years old. Column (1) of Panel A provides the treatment effect estimates using a variant of regression equation-8, which is a specification without controls for judge-level characteristics. The treatment effects of exposure to riot are positive and statistically significant at the 1% level of significance. The shares of bail denied by judiciary officers exposed to violence in early years are 6.3 percentage points higher, which is an increase of 17% (= 0.063/.37) compared with the baseline mean, than the shares of bail denied by judiciary officers without such exposure. In Column (2), we add controls for judge-level characteristics, such as experience, gender, religion, and LLB examination grades. The treatment effect estimates show a 6.1 percentage points increase in pretrial detention rates, which is an increase of 16.4% (= 0.061/.37) compared with the baseline mean, which is statistically significant at the 1% level of significance. In Column (3), we add controls for the occurrence of communal riots 5 years before birth. Our coefficient of interest remains almost unaffected, with 17% (= 0.062/.37) increase in detention rates. The effects are statistically significant at the 1% level of significance.

In Column (4), we add birth-year-quarter fixed effects (and exclude birth-year fixed effects) to flexibly account for the unobserved current time trends by judges' birth cohort. The coefficient remains stable at 0.06 and is statistically significant at the conventional level. Even though we have shown that riot-exposed judges do not select into types of crime, we perform one more check to alleviate the concern. We change the specification in Column (5) where we aggregate the data at the judge-crime type-district-quarter level (which increases the number of observations) and include crime-type fixed effects explicitly. The size of the coefficient estimate relative to the mean is 16.87% (=.054/0.32) which is very similar to the estimates from Column(2). Finally, in Column(6) we add potential predictors of riots (Corbridge, Kalra, and Tatsumi, 2012) - share of urban population, share of Muslims population, share of Muslims in urban areas and log of total population at home-district-year level - our result remain unchanged.

Recent literature on the causal effects of exposure to violence during early-childhood in the context of asylum seekers in Switzerland (Couttenier et al., 2019) and bank managers in India (Fisman et al., 2020) have used ages 0–12 and 0–10 years old (for early-childhood), respectively.[36] In light of this evidence, we estimate our main regression equation 8 and add controls for exposure to violence in the years after age 6 in Panel B. In Column (1) of Panel B, we control for exposure until age 14 years, in Column (2) of Panel B we control for exposure until age 18 years, and last, in Column (3), we control for exposure until age 22 years. Adding controls for exposure reduces our sample from Columns (1) to (3) in Panel B, but our results remain positive with a similar magnitude and are statistically significant at the 5% level of significance.

In our main table, we also report various estimates of the standard errors of the treatment effect of early-childhood exposure to communal violence. The Moulton-corrected standard errors and the wild bootstrap standard errors are both stable and demonstrate that the coefficients of interest across specifications are significant at the 5% level of significance. Further, in Appendix Table-3.18, we show that our results are robust to clustering at the level of home-district -year of birth and home-district level. The coefficients are stable across specifications.

F. Interpretation: Selection or Exposure Effect

A key concern about the interpretation of the impact coefficient is that the coefficient could also include sorting into the judiciary. We adopt two empirical strategies to show that impact of early-childhood exposure to violence is not driven by the selection of exposed individuals into the judiciary.

In the presence of the selection effect, we would observe either under- or over-representation in the judiciary of judges exposed to riots in early-childhood. In the first method, we compare the share of the early-childhood riot-exposed population in the entire working population (in all sectors and all types of employment) in UP with the share of early riot-exposed judges in this state. To this end, we exploit the data from the Employment and Unemployment Survey of the National Sample Survey Organization (NSSO) (66th round, 2011). We restrict our analysis to the riot-affected UP districts. This

[36] We show in Appendix Table 3.17 that exposure to riots between the age of 0-9, or 0-10 or 0-12 do not cause statistical significant effects on the share of bail denied. The coefficient is positive, consistent with the effect of riot exposure between the age of 0-6. But the riot exposure coefficients are imprecise.

survey captures information about individuals' age (but not their birth date) and the district in which they were residing at the time of the survey (i.e., their current district, but not their birth district), which we use to ascertain their riot exposure. One constraint is that no all-India survey captures information on survey respondents' birthplace (or even birth district). However, the migration literature has shown a low migration rate (5–6%) for India. Further, 99% of the migration is within a district. Hence, for this exercise, we assume that the current district is the birth district. Next, we select the sample born after 1950 (since our riot data start from 1950) who are employed (all types of employment). In Appendix Table-3.19, we find that the percentage of the total working population exposed to riots when aged 0–6 years is 39.2%, whereas the percentage of riot-exposed judges in the total population of judges in UP is 38.64%, in the sample of bail judges is 38.88%, and in our analysis sample is 38.8%. It is reassuring to note that there is no over- or under-representation of riot-exposed judges in the judiciary compared with the representation of the riot-exposed population in the total working population.

In the second approach, we ask whether different numbers and types of judges, where type is defined by gender and religion, are drawn from different riot-affected districts. In particular, we test whether the home-districts that experience a riot in a given year are more likely to have different total number and types of judges using the following specification at the home-district-riot-year level.

$$Judges_{h,y} = \alpha + \eta_h + \delta_y + \beta Exposed_{h,y} + \epsilon_{h,y} \tag{13}$$

where h and y denote home-district and riot-year. η_h is the home-district fixed effect, and δ_y is the riot-year fixed effect. The outcome variables $Judges_{hy}$ are the total number of judges, the proportion of females, and the proportion of Muslim judges. Since our focus is on the first 6 years of exposure, the outcome variable includes judges born 6 years before any given home-district riot-year. For instance, if the district Agra had a riot in 1970, the total number of judges affected by this riot (in their early-childhood, 0-6 years) would be the judges born in Agra in 1965-1970.

Appendix Table 3.20 shows that there is no selection of the type or total number of judges by riot-affected home-districts in any given year.

G. Robustness

G.1. Few Judges per District Concern. Our analytical sample has 668 judges over several districts, possibly resulting in a small number of judges per district, thereby raising the concern that the impact is attributable to riot exposure among a few judges. To address this concern, we sort the judges by their influence on the regression coefficient, where Cook's distance measures the influence. Figure 3.2 plots the coefficients from the estimates of our main specification (Column (2) of Panel A of Table 3.2), by excluding one judge at a time—starting with the judge having the highest influence on the regression coefficient—and ultimately excluding 300 judges. The coefficient is stable and statistically significant at the 5% level of significance until the exclusion of the first 280 judges (out of 668 judges), alleviating the concern that a few judges may be driving our result.

G.2. Case-level regressions. Our outcome variable is aggregated at the judge–district–quarter level. It may be argued that using case-level outcome data could account for the differences in workload by judges within a court-quarter (i.e., across courtrooms in the same District Court in a given quarter), which are not addressed by the district-quarter fixed effect, especially for larger districts with multiple police stations and multiple courtrooms adjudicating criminal trials. In Appendix Table 3.21, we test the regression at the case level instead of aggregating at the judge–district–quarter level. Column (1) has the same controls and fixed effects as in our baseline results (i.e., Column (2) of Table 3.2). In Column (2), we add the crime-type fixed effect, and in Column (3), we further refine our specification by adding two more controls—a dummy for whether the defendant is a Muslim and a dummy for the nonbailable nature of the case. The coefficients range from 0.038 to 0.043, which is 11 to 12% over the mean, and are close to our main result.

Although case-level data account for the different workloads per judge, there are concerns about correct inference owing to the clustering of outcomes. Following the design-based uncertainty approach of Abadie et al., 2017, since the random variation of treatment is at the judge level, the data should be clustered at the judge level. However, if each judge has a different number of cases, case-level data lead to misleading inferences because of the varying cluster sizes (MacKinnon and Webb, 2017). Assuming a sampling-based approach to clustering, in line with (Cameron and Miller, 2015), then the level at which the data should be clustered because of correlation is ambiguous. It can be suggested that with case-level outcomes, since the same defendant(s) can be represented across cases assigned to judges, the correct inference would require accounting for serial correlation across cases with the same defendant in addition to clustering at the judge level. In a similar quasi-random judge assignment study, W. Dobbie, Goldin, and C. S. Yang, 2018 account for two-way clustering by including the defendant- and the judge-level clusters. This approach is not feasible for our data because we do not have a unique defendant ID.

The benefits of adding case level controls are very limited owing to data limitations. However, there are inference issues as mentioned above arising from clustered data in case-level regressions. Hence, our preferred specification aggregates the outcome data at the judge level using judge-level clustering for inference.

G.3. Sample Selection. We follow Arnold, W. S. Dobbie, and Hull, 2020 and exclude the bottom 5 percentile judges (i.e., judges handling less than 97 cases) from our primary analysis sample, to allay concerns related to judges dealing with very few cases driving our outcomes. In Appendix Table 3.22, we present the results using alternative thresholds for the exclusion of judges from analysis samples. From Column (1) to Column (10), we change the threshold of exclusion from the bottom 1 to 10 percentile. The coefficients are very stable and close to our main result in all the specifications.

G.4. High-rank Judges. Another concern is that high-ranked judges may influence the cases assigned to them. In Appendix Table 3.23, we exclude District and Session Judges and Chief Judicial Magistrates—the two most influential judges in the district-level judiciary—and find that our coefficient magnitudes range from 5.4 to 8.4 percentage points and are statistically significant at the 95% confidence interval.

G.5. Outlier Tests. The next set of robustness tests is to check for potential outliers in our baseline results. In Appendix Table 3.24, we show that judges from home districts exposed to a high number of riots do not drive our results. Column (1) in Table 3.24 presents the results after the exclusion of judges from the home districts that have experienced the highest number of Hindu–Muslim riots, Column (2) presents the results after the exclusion of judges from the home districts with the second-highest number of riots, and so on. We observe that the effect of early-childhood exposure to riots is positive and statistically significant at the conventional levels, with its magnitude ranging from an increase of 5.7 to 7.5 percentage point in pretrial detention rates. In Appendix Table 3.25, we remove the home districts with the highest number of riots cumulatively. Here, again we find that the treatment effect of exposure to riots when in the age group of 0–6 years is positive, and its magnitude ranges from 6.8 percentage points to 8.8 percentage points, significant at the 1% level of significance. Lastly, we test our baseline results by removing observations that are 3, 2, and 1 standard deviation away from the residual mean in Column (1), Column (2), and Column (3) in Appendix Table 3.26, respectively. In addition, we remove observations with high leverage, which shift estimates to at least one standard error and to at least $4/N$. The results are positive, with magnitudes ranging from 5.3 percentage point to 6.4 percentage points, and are significant at the conventional level of significance across all specifications.

G.6. Placebo Test. In our placebo check, we follow a Monte Carlo approach and randomly reassign our treatment variable $kid[0-6]$ following a binomial distribution, based on the observed distributions of $kid[0-6]$, keeping all other characteristics unchanged. We estimate our main specification (Column (2) of Panel A of Table 3.2) on the simulation data. We implement 1,000 simulations. The sampling distribution of the treatment effects of $kid[0-6]$ Monte Carlo draws is centered around zero. Figure 3.8 demonstrates that the probability of the treatment effect found in our main specification being spurious is negligible.

H. Threats to Identification:

H.1. Migration. The migration of households from riot-hit districts to districts less likely to experience Hindu-Muslim riots would violate SUTVA (Rubin, 1980), which is our identifying assumption. Therefore, we test whether the migration rates are affected by the communal riots. We use the NSSOs' microdata from the Employment and Unemployment Survey 1983, which captures migration information.[37] The important migration-related information we exploit are the age at which migration occurs, the district from where migration takes place (i.e., the origin district), and whether migration occurs within the district or to another district. The data allow us to perform analysis only for the migrating population. Since the violation of SUTVA in our setting occurs in case of migration from a riot-hit district to another district, and not within a riot-hit

[37] We could not find any nationally representative survey capturing both the origin and destination districts. The later rounds of the NSSO's Employment and Unemployment Surveys do not provide data on origin districts.

district, we show that the share of out-migration from the district in the total migration is not correlated with the riots.

We build the data at the district-year level and run the following regression.

$$MigrationRate_{h,y} = \alpha + \eta_h + \delta_y + \beta Exposed_{h,y} + \epsilon_{h,y} \tag{14}$$

where h and y denote the origin of migration district and the year of migration. η_h is the district fixed effect, and δ_y is the year fixed effect. The outcome variable $MigrationRate_{hy}$ is the ratio of migration across districts to the total migration. The coefficient to focus upon is β associated with the explanatory variable $Exposed_{h,y}$ that captures whether the district-year cell had a riot. The β coefficient as shown in Table 3.3 is close to zero and statistically insignificant.

Next, we collect administrative data on the districts where the judges completed their secondary schooling, higher secondary schooling and undergraduate studies for a subsample of judges. In Table-3.4 we establish there is no early childhood riot exposure effect on migration away from home districts to the districts where they completed their secondary schooling (at age 15), higher secondary schooling (at age 17) and undergraduate studies (at age 21).

H.2. Underlying Political Determinants of Riots. Our riot-impact estimates could potentially be confounded by political variables that trigger riots (S. I. Wilkinson, 2006). We follow Besley and Burgess, 2002 and measure political competition in state legislative elections in two ways. First, it is measured by the difference in the number of seats between the incumbent party (Congress) and its main political competitor party (Janta Party before the 1990s and Samajwadi Party afterward). Second, it is measured by the absolute difference in the proportion of seats between the main party and the main political competitor. In Appendix-Table 3.27, we show that riot-exposed and non-exposed judges coming from the same home district do not experience differential political competition or differential voter turnout in elections in the first six years of their lives. Therefore, our riot-impact estimate is not confounded by time-varying attributes that underpin ethnic violence.

I. Mechanisms

A growing body of economics literature on endogenous preference formation emphasizes early childhood as a period in which fundamental preference parameters and character skills develop. [38] More importantly, recent studies have highlighted that the social environment during early childhood can have persistent causal effects on preferences, such as the preference for honesty (Abeler, Falk, and Kosse, 2021), risk (Giuliano and Spilimbergo, 2014), and redistribution (Cappelen et al., 2020). In particular, seminal recent studies on exposure to violence in early childhood, such as those by Couttenier et al., 2019; Fisman et al., 2020), have found that high casualties resulting from intergroup conflict produce lasting intergroup hostility. One possibility of social environment affecting children is through parental influence. Parental traits can shape preferences of their children, for example children have been shown to become long-term oriented when observing a long-term oriented adult (Bandura and Mischel, 1965). It is possible that parents experiencing effective state-intervention in civil clashes develop positive attitude towards state institutions. The children who observe their parents' confidence in the institutions and the functioning of the state may develop greater support for such institutions. We follow the emerging literature on early childhood and explore whether there is intergroup hostility in the observed judicial stringency or whether judicial stringency potentially represents support for the state. We, additionally, use our data to rule out other non-behavioral channels, such as differences in cognitive abilities.

I.1. No Intergroup Bias Behavior. Early-life exposure to an intergroup conflict could generate animosity between groups, as evidenced in the high-intensity Hindu–Muslim violence in the Indian context in the case of bank managers (Fisman et al., 2020). To estimate the intergroup hostility effect, we would need to identify the religion of the judges and the defendants but do not have such administrative data. Following Bhalotra et al., 2014, who use names to infer the religion of electoral candidates in India, we use names to infer the religion of the judges and the defendants.

We manually assign each judge to a religious group using the judges' name and their fathers' name. For defendants, first, we use the "Stanford Named Entity Algorithm" to extract their names from the judgments (see Appendix F.3 for details). Then, we use the Nilabhra name2community algorithm to identify Urdu-sounding names, which we classify as Muslim names. To address the concern about the likely scope for error in identifying Muslim names, we test it on the dataset of Bhalotra et al., 2014. We find that

[38] Kautz et al., 2014; Alan, Boneva, and Ertac, 2019; Falk et al., 2021; Kohlberg, 1984; Piaget, 1997; Harbaugh, Krause, and Vesterlund, 2002; Sutter and Kocher, 2007; Fehr, Bernhard, and Rockenbach, 2008; Ben-Ner et al., 2017; Almås et al., 2010; Bauer, Chytilová, and Pertold-Gebicka, 2014

this algorithm predicts the religion from names with a 6% error rate. However, the error rate in the classification of the defendants' religion is higher (20%) owing to additional errors in the process of extracting names from the judgment pdfs.

In our sample of 668 judges, only 51 are Muslims, of which only 13 Muslim judges were exposed to communal violence when 0–6 years old. In comparison, we have 617 Hindu judges in our sample, out of which 197 Hindu judges were exposed to religious riots between ages 0 and 6 years.[39] However, about 20% of cases involve only Muslim defendants, as measured by the algorithm.

We perform a subsample analysis to test whether the bail decisions of early-childhood riot-exposed Hindu judges differ in cases where all the defendants are Hindus from their decisions in cases where all the defendants are Muslims. Columns (2) and (3) of Table 3.7 reveal that the coefficient measuring the causal effect on early-childhood exposure to communal riots remains positive for both Hindu and Muslim defendants, with the coefficient for Hindu defendants being 5.1 percentage points (14% increase in pretrial detention rates) and 7.3 percentage points (20% increase in pretrial detention rates), both statistically significant at the 5% level of significance. We do a Chow test from a pooled regression and find that the early-riot exposure coefficient for Hindu defendants is not statistically different from the early-riot exposure coefficient for Muslim defendants (F-stat= 0.15).

I.2. Riot Intensity and State Lockdowns. Hindu–Muslim religious clashes in India affect socioeconomic outcomes not only through riot casualties or social segregation but also through state-imposed lockdowns. S. I. Wilkinson, 2006 argues that the state response to Hindu–Muslim riots in the form of arrests, lockdowns, and increased police presence plays a huge role in determining riot damages. In other words, an effective state response can prevent the escalation of a riot. The early-childhood exposure of individuals to a sociopolitical environment in which strong state action resulted in fewer riot-related deaths can potentially generate in them support for, or confidence in, the state relative to the individual. Therefore, we hypothesize that judicial stringency could be driven by judges with a positive childhood experience of state intervention to curb civilian misconduct.

[39]The sparse presence of Muslim judges is not surprising, and several studies have shown the under-representation of Muslims, including Fisman et al., 2020 who considered the exposure to violence of bank managers.

To examine this hypothesis, we analyze the heterogeneity impact of riot casualties interacted with state lockdowns. First, we compute casualties per land area (instead of population size to avoid reverse causation bias) experienced by each judge between ages 0 and 6 years. We select the median value of casualty experienced by the riot-exposed judges as the threshold below which we term a riot as a low-severity riot. Similarly, a state response measure, such as a high lockdown, is defined as a lockdown that lasts for more than 5 days, which is the median days of lockdown experienced by our sample of riot-exposed judges. High arrests are defined as arrests that exceed 170, the median value of arrests that occurred in riots experienced by judges between ages 0 and 6 years.

We compare the bail decisions of non-exposed judges with that of judges experiencing high state action in terms of high lockdowns or arrests, but varying levels of riot casualties, using the following equation:

$$B_{j,d,t} = \alpha + \eta_{d,t} + \delta_{h,t} + \mathcal{F}.\mathcal{E}_b + \beta_1 \times high - casualty - high - state - action[0-6]_j + \beta_2 \times low - casualty - high - state - action[0-6]_j + \sum_{k=7}^{9} \gamma(k) \times exposure(k)_j + X_j + \epsilon_{j,d,t} \tag{15}$$

where $high - casualty - high - state - action[0-6]_j$ denote high casualty and a higher period of lockdowns or police arrests, and $low - casualty - high - state - action[0-6]_j$ denote low levels of casualty and higher period of lockdowns or police arrests. The remaining variables are same as in our main specification in equation- 8.

Table-3.5 presents the heterogeneity impact by intensities of lockdown and riot-related casualties. We observe that among the judges who experienced early-childhood riots with intense state response measured by the total number of police arrests or the total number of days of state-imposed lockdowns, it is the judges with an early-childhood experience of riots resulting in low casualties who drive judicial stringency in bail decisions.[40]

The above pattern is consistent with the hypothesis that the early-life experiences of judges regarding effective lockdowns or police arrests that have effectively controlled

[40]We report riot severity results using an alternative specification in Table 3.28. We find that by all measures of riot intensity, low-intensity riots explain a high share of bail denied by early-childhood riot-exposed judges, which is statistically significant at the 1% level of significance (compared with non-exposed judges). Further the results are robust to using alternative methods of calculating the intensity thresholds, as presented in Appendix Table 3.31

civilian violence generate in them persistent confidence in the state relative to the individual.

I.3. Heterogeneity by Age of Exposure. Motivated by Cappelen et al., 2020, who show that interventions during ages 3 to 4 years have a long-term impact on social preferences, we examine the heterogeneity by riot exposure to test whether our interpretation of the interaction term of conflict severity is driven by changes in the support for the state.

We group judicial officers based on their exposure to riots into three mutually exclusive age groups of 0–3, 3–6, and 6–9 years to test whether the effects are determined by exposure between ages 3 and 6 years.[41] We estimate the following regression specification, which is a variant of our main specification in equation 8:

$$
\begin{aligned}
B_{j,d,t} = \alpha + \eta_{d,t} + \delta_{h,t} + \mathcal{F}.\mathcal{E}_b + \beta \times pre-birth[0-3]_j + \beta_1 \times kid[0-3]_j + \\
\beta_2 \times kid[3-6]_j + \beta_3 \times kid[6-9]_j + X_j + \epsilon_{j,d,t}
\end{aligned}
\tag{16}
$$

In our data, a similar number of judges in each age bin were exposed to communal riots. The families of 125 judges (approximately 19% of the total judges in the estimation sample) were exposed to violence between 0 and 3 years before the judges' birth. Further, 131 judges were exposed to violence when 0–3 years old, and 132 judges when 3–6 years old (which is approximately 20% of the total judges in the estimation sample). Last, 163 judges were exposed to communal clashes when 6–9 years old (which is approximately 24% of the total judges in the estimation sample). In Appendix Table 3.29, we show the results on estimating the above equation for older cohorts of judiciary officers for whom we can control for potential exposure to violence up to age 22 years. We find a statistically significant ($p < 0.05$) positive treatment effect of exposure for the age group of 3–6 years. Figure 3.1 plots the coefficient estimate of each age group. We note that the effects are primarily driven by exposure between ages 3 and 6 years. Our results are robust to multiple hypothesis testing, with Bonferroni p-value for the treatment effect of exposure for the age group of 3–6 years being 0.0058.

Next, we test the impact on bail decisions by the age of first exposure to communal conflicts. We construct conditional extensive margins by estimating the effects on

[41]The term mutually exclusive means that if a judge is exposed when aged 0–3 years as well as when aged 3–6 years, then (s)he will be categorized into the 0–3 age category.

bail decisions by the age at first exposure, denoted by $firstexposure(k)$. Our regression equation is as follows:

$$B_{j,d,t} = \alpha + \eta_{d,t} + \delta_{h,t} + \mathcal{F}.\mathcal{E}_b + \beta \sum_{k=1}^{n} firstexposure[k]_j + \sigma X_j + \epsilon_{j,d,t} \tag{17}$$

where $firstexposure(k)$ is defined as the age at first exposure at k. All other variables are the same as specified for equation 8.

Figure 3.3 and Appendix Figure3.9 plot the coefficient estimates with 95% confidence intervals of equation 17 for the full sample, for which the effects of the age at first exposure can be estimated only up to year 9, and for the subsample, for which the effects of the age at first exposure can be estimated up to year 22. Both plots demonstrate the causal effect on bail decisions of the age at first exposure being 4 and 5 years. Appendix Table-3.30 shows the effect of age at first exposure to communal violence for the full analysis sample. We observe that the effect of the age at first exposure being 4 years is 8.1 percentage points, and the effect of the age at first exposure being 5 years is 11.8 percentage points, which is statistically significant at the 5% and 1% levels of significance, respectively. Given the findings in the early childhood literature, this result provides further support for our interpretation in the above section that there is a causal link between early-childhood riot exposure to the support for the state in controlling civilian misconduct.

I.4. Do Recent Riots Exposure Matter? In Figure-3.4, we plot the post-period effects of current riots (2014-2017) at the weekly level for up to 12 weeks separately for early-exposed and non-exposed judges. We establish that the effects of current riots are not persistent. There is no effect of riots on shares of bail denied for early exposed judges. For the non-exposed judges there are marginal positive effects up to 5 weeks post the riots, with most post-period effects being statistically insignificant.

I.5. Judicial Education and Judicial Stringency. Next, we test whether differences in cognitive skills as measured by performance in the mandatory LLB examination explain differences in bail decisions across judges. Table-3.6 reports that the inclusion of LLB examination results do not affect our coefficient estimates of the early-childhood riot exposure effect. Therefore, we conclude that heterogeneity in skills in law training does not explain our results.

J. Conclusion

In this study, we examine the population of judges and show that their exposure to communal violence at ages 0–6 years has persistent economic and statistically significant effects on pretrial detention rates. Unlike studies that have focused on estimating bias and discrimination in judicial decisions, we investigate the origins of judicial bias. We show that early-childhood exposure to the sociopolitical environment has robust effects on adult decisions across generations. We show that judges exposed to communal violence between the ages of 0 and 6 years are 16% more prone to deny bail than the average judge. The effect is driven by exposure to a low number of riot-related deaths and injuries and a low riot duration, as well as by exposure when between 3 and 6 years of age.

We provide some evidence in support of our interpretation that the experience of riot de-escalation efforts by the state that result in low riot-related damages during judges' formative years has long-term effects on judicial outcomes. Further research on how preferences and beliefs are formed owing to sociopolitical events during the formative years of childhood would provide decision-makers with insights for designing effective policy tools.

Bibliography

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge (2017). *When Should You Adjust Standard Errors for Clustering?* Working Paper 24003. National Bureau Economic Research.

Abeler, Johannes, Armin Falk, and Fabian Kosse (2021). *Malleability of Preferences for Honesty*. Working Paper 14304. IZA Discussion Paper.

Adhvaryu, Achyuta, James Fenske, and Anant Nyshadham (2019). "Early Life Circumstance and Adult Mental Health". In: *Journal of Political Economy* 127.4, pp. 1516–1549.

Agan, Amanda and Sonja Starr (2018). "Ban the box, criminal records, and racial discrimination: A field experiment". In: *The Quarterly Journal of Economics* 133.1, pp. 191–235.

Aizer, Anna and Flavio Cunha (2012). *The Production of Human Capital: Endowments, Investments and Fertility*. Working Paper 18429. National Bureau of Economic Research.

Alan, Sule, Teodora Boneva, and Seda Ertac (2019). "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit". In: *The Quarterly Journal of Economics* 134.3, pp. 1121–1162.

Almås, Ingvild, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden (2010). "Fairness and the Development of Inequality Acceptance". In: *Science* 328.5982, pp. 1176–1178.

Almond, Douglas (2006). "Is the 1918 Influenza Pandemic over? Long-Term Effects of in Utero Influenza Exposure in the Post-1940 US Population". In: *Journal of Political Economy* 114.4, pp. 672–712.

Almond, Douglas and Janet Currie (2011). "Killing me Softly: The Fetal Origins Hypothesis". In: *Journal of Economic Perspectives* 25.3, pp. 153–72.

Almond, Douglas, Lena Edlund, and Mårten Palme (2009). "Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden". In: *The Quarterly Journal of Economics* 124.4, pp. 1729–1772.

Angerer, Silvia, Daniela Glätzle-Rützler, Philipp Lergetporer, and Matthias Sutter (2015). "Donations, Risk Attitudes and Time Preferences: A Study on Altruism in Primary School Children". In: *Journal of Economic Behavior & Organization* 115, pp. 67–74.

Arnold, David, Will Dobbie, and Crystal S Yang (2018). "Racial Bias in Bail Decisions". In: *The Quarterly Journal of Economics* 133.4, pp. 1885–1932.

Arnold, David, Will S Dobbie, and Peter Hull (2020). *Measuring racial discrimination in bail decisions*. Tech. rep. National Bureau of Economic Research.

Ash, Elliott, Sam Asher, Aditi Bhowmick, Daniel Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, and Bilal Siddiqi (2021). *Measuring Gender and Religious Bias in the Indian Judiciary*. Working Paper.

Bakhtawar, Ali and Sultan Mehmood (2022). *Judicial Capture by Gift Exchange*. Working Paper.

Bandura, Albert and Walter Mischel (1965). "Modifications of self-imposed delay of reward through exposure to live and symbolic models." In: *Journal of personality and social psychology* 2.5, p. 698.

Bauer, Michal, Julie Chytilová, and Barbara Pertold-Gebicka (2014). "Parental Background and Other-Regarding Preferences in Children". In: *Experimental Economics* 17.1, pp. 24–46.

Ben-Ner, Avner, John A List, Louis Putterman, and Anya Samek (2017). "Learned Generosity? An Artefactual Field Experiment with Parents and their Children". In: *Journal of Economic Behavior & Organization* 143, pp. 28–44.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?" In: *The Quarterly Journal of Economics* 119.1, pp. 249–275.

Besley, Timothy and Robin Burgess (2002). "The Political Economy of Government Responsiveness: Theory and Evidence from India". In: *The Quarterly Journal of Economics* 117.4, pp. 1415–1451.

Bhalotra, Sonia, Irma Clots-Figueras, Guilhem Cassan, and Lakshmi Iyer (2014). "Religion, Politician Identity and Development Outcomes: Evidence from India". In: *Journal of Economic Behavior & Organization* 104, pp. 4–17.

Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson (2013). "Early Life Health Interventions and Academic Achievement". In: *American Economic Review* 103.5, pp. 1862–91.

Bielen, Samantha, Wim Marneffe, and Naci Mocan (2021). "Racial Bias and In-Group Bias in Virtual Reality Courtrooms". In: *The Journal of Law and Economics* 64.2, pp. 269–300.

Billings, Stephen B, Eric Chyn, and Kareem Haggag (2020). *The Long-Run Effects of School Racial Diversity on Political Identity*. Working Paper 27302. National Bureau of Economic Research.

Blakeslee, David S and Ram Fishman (2018). "Weather shocks, agriculture, and crime evidence from India". In: *Journal of Human Resources* 53.3, pp. 750–782.

Bleakley, Hoyt (2007). "Disease and Development: Evidence from Hookworm Eradication in the American South". In: *The Quarterly Journal of Economics* 122.1, pp. 73–117.

— (2010). "Malaria eradication in the Americas: A Retrospective Analysis of Childhood Exposure". In: *American Economic Journal: Applied Economics* 2.2, pp. 1–45.

Cameron, A Colin and Douglas L Miller (2015). "A Practitioner's Guide to Cluster-Robust Inference". In: *Journal of human resources* 50.2, pp. 317–372.

Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden (2020). "The Effect of Early-Childhood Education on Social Preferences". In: *Journal of Political Economy* 128.7, pp. 2739–2758.

Corbridge, Stuart, Nikhila Kalra, and Kayoko Tatsumi (2012). "The search for order: Understanding hindu-muslim violence in post-partition india". In: *Pacific Affairs* 85.2, pp. 287–311.

Couttenier, Mathieu, Veronica Petrencu, Dominic Rohner, and Mathias Thoenig (2019). "The Violent Legacy of Conflict: Evidence on Asylum Seekers, Crime, and Public Policy in Switzerland". In: *American Economic Review* 109.12, pp. 4378–4425.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Econometrica* 78.3, pp. 883–931.

Currie, Janet (2009). "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor health in Childhood, and Human Capital Development". In: *JoUrnal of economIc lIteratUre* 47.1, pp. 87–122.

Currie, Janet and Tom Vogl (2013). "Early-life Health and Adult Circumstance in Developing Countries". In: *Annu. Rev. Econ.* 5.1, pp. 1–36.

Dobbie, Will, Jacob Goldin, and Crystal S Yang (2018). "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges". In: *American Economic Review* 108.2, pp. 201–40.

Dobbie, Will and Crystal Yang (2021). *The Economic Costs of Pretrial Detention*. Working Paper. Brookings Paper on Economic Activity.

Doleac, Jennifer L (2021). "Racial Bias in the Criminal Justice System". In: *A Modern Guide to Economics of Crime*.

Eren, Ozkan and Naci H Mocan (2020). "Judge Peer Effects in the Courthouse". In: *NBER Working Paper* w27713.

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers (2021). "Socioeconomic status and inequalities in children's IQ and economic preferences". In: *Journal of Political Economy* 129.9, pp. 000–000.

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach (2008). "Egalitarianism in Young Children". In: *Nature* 454.7208, pp. 1079–1083.

Fisman, Raymond, Arkodipta Sarkar, Janis Skrastins, and Vikrant Vig (2020). "Experience of Communal Conflicts and Intergroup Lending". In: *Journal of Political Economy* 128.9, pp. 3346–3375.

Giuliano, Paola and Antonio Spilimbergo (2014). "Growing up in a Recession". In: *Review of Economic Studies* 81.2, pp. 787–817.

Gould, Eric D, Victor Lavy, and M Daniele Paserman (2011). "Sixty Years after the Magic Carpet Ride: The Long-Run Effect of the Early Childhood Environment on Social and Economic Outcomes". In: *The Review of Economic Studies* 78.3, pp. 938–973.

Harbaugh, William T, Kate Krause, and Lise Vesterlund (2002). "Risk Attitudes of Children and Adults: Choices over Small and Large Probability Gains and Losses". In: *Experimental Economics* 5.1, pp. 53–84.

Heckman, James, Rodrigo Pinto, and Peter Savelyev (2013). "Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes". In: *American Economic Review* 103.6, pp. 2052–86.

Heckman, James J (2006). "Skill Formation and the Economics of Investing in Disadvantaged Children". In: *Science* 312.5782, pp. 1900–1902.

— (2007). "The Economics, Technology, and Neuroscience of Human Capability Formation". In: *Proceedings of the National Academy of Sciences* 104.33, pp. 13250–13255.

Heckman, James J and Yona Rubinstein (2001). "The Importance of Noncognitive Skills: Lessons from the GED Testing Program". In: *American Economic Review* 91.2, pp. 145–149.

Heckman, James J, Jora Stixrud, and Sergio Urzua (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior". In: *Journal of Labor Economics* 24.3, pp. 411–482.

Henrichson, Christian, Joshua Rinaldi, and Ruth Delaney (2015). *The Price of Jails: Measuring the Taxpayer Cost of Local Incarceration*. Working Paper. Vera Institute of Justice.

Kautz, Tim, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. Working Paper w20749. National Bureau of Economic Research.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018). "Human decisions and Machine Predictions". In: *The Quarterly Journal of Economics* 133.1, pp. 237–293.

Kling, Jeffrey R (2006). "Incarceration Length, Employment, and Earnings". In: *American Economic Review* 96.3, pp. 863–876.

Kohlberg, Lawrence (1984). *Essays on Moral Development/2 The Psychology of Moral Development*. Harper & Row.

Leslie, Emily and Nolan G Pope (2017). "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments". In: *The Journal of Law and Economics* 60.3, pp. 529–557.

Lim, Claire SH (2015). "Media influence on courts: evidence from civil case adjudication". In: *American Law and Economics Review* 17.1, pp. 87–126.

Maccini, Sharon and Dean Yang (2009). "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall". In: *American Economic Review* 99.3, pp. 1006–26.

MacKinnon, James G and Matthew D Webb (2017). "Wild Bootstrap Inference for Wildly Different Cluster Sizes". In: *Journal of Applied Econometrics* 32.2, pp. 233–254.

Mehmood, Sultan (2021). "The impact of Presidential appointment of judges: Montesquieu or the Federalists?" In: *Program on Governance and Local Development Working Paper* 40.

Mehmood, Sultan, Avner Seror, and Daniel Chen (2021). *Rituals*. Working Paper.

Mitra, Anirban and Debraj Ray (2014). "Implications Of An Economic Theory Of Conflict: Hindu-Muslim Violence In India ". In: *Journal of Political Economy* 122.4, pp. 719–765.

Piaget, Jean (1997). *The Moral Judgement of the Child*. Simon and Schuster.

Ribeiro, Beatriz and Claudio Ferraz (2019). *Pretrial Detention and Rearrest Rates: Evidence from Brazil*. Working Paper.

Rubin, Donald B (1980). "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment". In: *Journal of the American Statistical Association* 75.371, pp. 591–593.

Sarsons, Heather (2015). "Rainfall and Conflict: A Cautionary Tale". In: *Journal of Development Economics* 115, pp. 62–72.

Shayo, Moses and Asaf Zussman (2011). "Judicial Ingroup Bias in the Shadow of Terrorism". In: *The Quarterly Journal of Economics* 126.3, pp. 1447–1484.

Stevenson, Megan T (2018). "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes". In: *The Journal of Law, Economics, and Organization* 34.4, pp. 511–542.

Sutter, Matthias and Martin G Kocher (2007). "Trust and Trustworthiness Across Different Age Groups". In: *Games and Economic Behavior* 59.2, pp. 364–382.

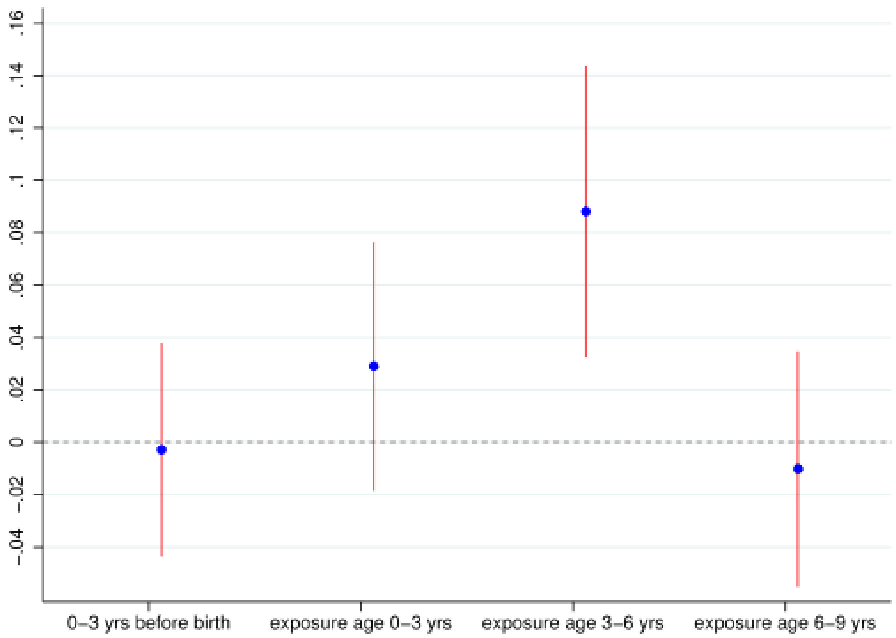
Varshney, Ashutosh and Steven Wilkinson (2006). *Varshney-Wilkinson Dataset on Hindu-Muslim Violence in India, 1950-1995, version 2*. Inter-university Consortium for Political and Social Research.

Walmsley, Roy (2018). *World Pre-trial/Remand Imprisonment List*. Working Paper. Institute for Criminal Policy Research.

Wilkinson, Steven I (2006). *Votes and violence: Electoral competition and ethnic riots in India*. Cambridge University Press.

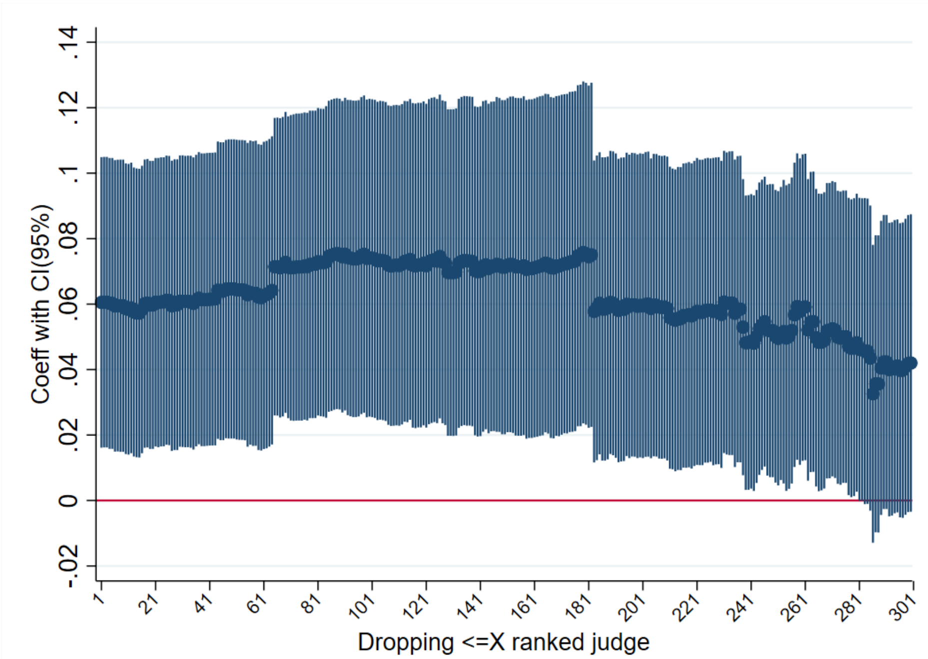
K. Figures

FIGURE 3.1. Treatment Effects Estimates of Riot-Exposure by Age



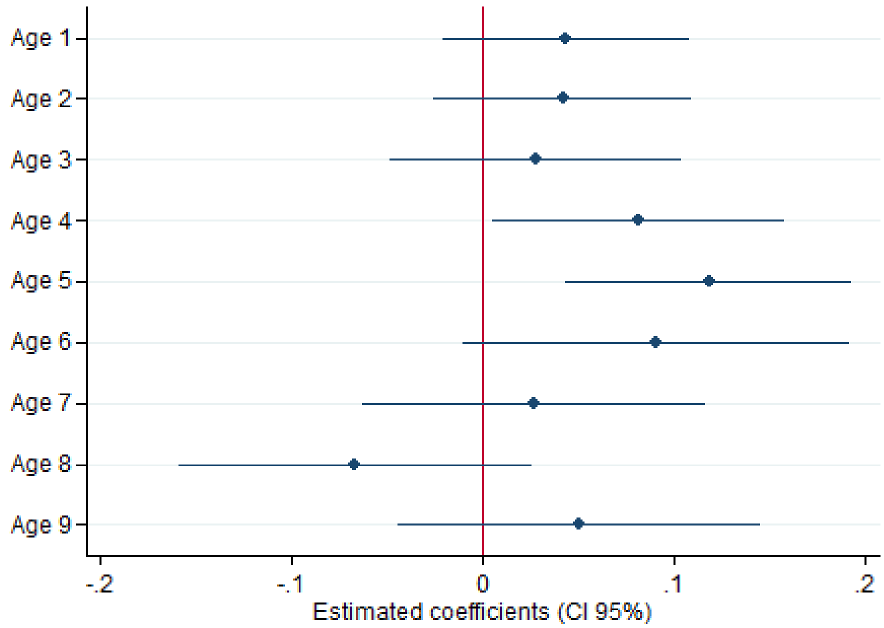
Notes: The figure reports the coefficients from equation-16, where we create four mutually exclusive age bins of 3 years each starting from 3 years before birth up to 9 years of age for exposure to riots. The figure illustrates that exposure to riots when aged 3-6 years of age is statistically significant.

FIGURE 3.2. Removing Judges serially with high Influence



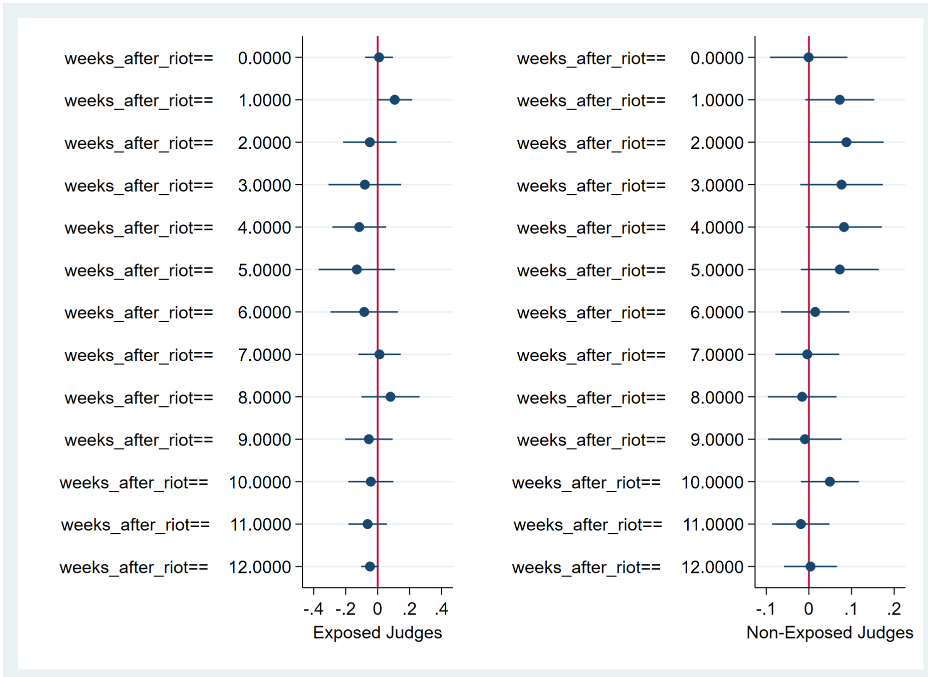
Notes: The figure reports the coefficients from running the main equation-16, 300 times by removing one by one the judges with the highest contribution to the early-childhood exposure effect. The contribution is measured by Cook's distance. The result goes away only after removing half of the judges from the sample.

Figure 3.3. Treatment Effects Estimates of Age at First Exposure



Notes: Estimated coefficients and the confidence interval at 95% are reported through ordinary least squares estimation using equation-17 on the total sample of 668 judges. The main dependent variable is the pretrial detention rate at the judge-district-quarter level. The estimation includes home-district-quarter, birth-year, and district-quarter fixed effects.

FIGURE 3.4. Impact of Current Riots



Notes: This figure presents event study plots of share of bail denied up to 12 weeks post-riot exposure for riots that occurred in 2014-2017. The figure presents estimates of the trajectory of the share of bail denied over time separately for early-exposed and non-exposed judges following the specification: $Y_{jdt} = \beta * T + \eta_j + \delta_{dt} + \epsilon_{jdt}$; where, Y_{jdt} is the share of bail denied by judge j, posted at district d in week t; T is the time dummy for the post and pre-exposure periods, η_j is judge fixed effect, δ_{dt} is district-weekly trends, and ϵ_{jdt} is the error term clustered at the judge level.

L. Tables

Table 3.1. Balance Test: Riot-Exposed Judges and Type of Cases

CASES CHARACTERISTICS	(1) Tot Cases	(2) Muslim Def	(3) Non Bailable	(4) With SLL	(5) IPC1	(6) IPC2	(7) IPC3	(8) IPC4	(9) IPC5	(10) IPC6
***Kid*[0-6]**	-0.921	0.0037	-0.0025	0.011	-0.0032	0.0015	0.011	-0.001	-0.001	0.0023
	(6.754)	(0.008)	(0.016)	(0.030)	(0.007)	(0.009)	(0.009)	(0.009)	(0.007)	(0.008)
Observations	5,530	5,394	5,503	5,503	5,530	5,530	5,530	5,530	5,530	5,530
R-squared	0.359	0.473	0.319	0.309	0.316	0.343	0.299	0.319	0.310	0.349
Mean Dep Var	55.28	0.10	0.83	0.44	0.06	0.14	0.18	0.16	0.10	0.13
Home Dist X Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DOB Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dist X Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Judge	Judge	Judge	Judge	Judge	Judge	Judge	Judge	Judge	Judge
No. of Clusters	660	660	660	660	660	660	660	660	660	660

This table reports ordinary least squares estimations based on the judge–district–quarter level sample of 668 judges with home districts in UP. Robust standard errors in parentheses are clustered at the judge level. The dependent variable are: Total Bail Cases (Column 1), share of cases with Muslim defendants (Column 2), share of Non-Bailable cases (Column 3), share of cases booked under special Acts (Column 4), Column 5-10 are the share of cases booked under one to six IPC sections, whereas the last column is 6 or more sections. The sample mean and the standard deviation of the dependent variables are reported. The explanatory variable is exposure to communal conflict between the ages of 0 and 6 years. All estimations include home district X quarter, yesar of birth fixed effects. Exposure to riots until 9 years of age is included as control with a set of binary measures. The reduction from 668 is due to the dropping of singleton observations.

Table 3.2. Bail Decisions and Early Riot Exposure

PANEL A: FULL SAMPLE DEPENDENT VARIABLE: SHARE DENIED	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
1-5 yrs ***P*re Birth**			0.008 *(0.020)*			0.003 *(0.020)*
***K*id[0-6]**	0.063*** *(0.022)*	0.061*** *(0.023)*	0.062*** *(0.023)*	0.060** *(0.024)*	0.054*** *(0.017)*	0.062** *(0.025)*
Observations	5,530	5,530	5,530	5,443	29,925	5,530
R-squared	0.327	0.332	0.332	0.391	0.235	0.335
Mean Dep Var	0.37	0.37	0.37	0.37	0.32	0.37
Controls	no	yes	yes	yes	yes	yes
home-district X Quarter F.E	yes	yes	yes	yes	yes	yes
Date of Birth F.E	yes	yes	yes	no	yes	yes
District X Quarter F.E	yes	yes	yes	yes	yes	yes
Date of Birth X Quarter F.E	no	no	no	yes	no	no
Crime Type F.E	no	no	no	no	yes	no
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge
Total Number of Clusters	660	660	660	657	668	660
Standard errors of *K*id[0-6]						
Moulton-corrected	0.0219	0.0220	0.0225	0.0236	0.0160	0.0235
Wild Bootstrap errors	0.0231	0.0232	0.0237	0.0253	0.0173	0.0170

PANEL B: SUBSAMPLES DEPENDENT VARIABLE: SHARE DENIED	**(1)**	**(2)**	**(3)**
***K*id[0-6]**	0.059** (0.023)	0.058** (0.024)	0.056** (0.024)
$\sum_{k=7}^{N} expo(k)_{h,b,t}$: **N**	14	18	22
Cluster Level	Judge	Judge	Judge
No of Judges	651	637	604
Observations	5,488	5,425	5,248
R-squared	0.335	0.339	0.343
Mean Dep Var	0.37	0.37	0.37
home-district X Quarter F.E	yes	yes	yes
Date of Birth F.E	yes	yes	yes
District X Quarter F.E	yes	yes	yes

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level sample of 668 judges (Panel A) and the subsample of judges (Panel B) from home districts in UP. Robust standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at the judge–district–quarter level. Kid[0–6] is a dummy of childhood exposure to communal conflict. Column 1 is without controls; Column 2 includes controls such as a dummy for Muslim, for female, first division in the Bachelor of Law examination, and the total tenure as a judge at the time of judgment, Column 3 further adds a binary measure of pre-birth exposure of judges' families to communal conflict. Column 6 further adds extra controls- share of urban population, share of muslim population, share of muslim in urban areas and log of total population. Column 4 adds a cohort-quarter fixed effects and Column 5 aggregates the data at judge-district-crime-quarter level and add crime fixed effect too in the regression. All estimations include a set of binary variables coding for exposure up to 9 years of age (since our conflict data are up to the year 2000 and the youngest judge in our sample is born in 1991). In Panel B, we extend the exposure control for later years: up to 14 years of age (Column 1: keeping judges born before 1986), 18 years of age (Column 2: keeping judges born before 1982), and 22 years of age (Column 3: keeping judges born before 1978) on a subsample of judges. All estimations include home district X quarter, year of birth, and district X quarter fixed effects.

TABLE 3.3. Does Migration rate affected by Riots?

	Inter-district Migration Rate
Riot (1/0)	0.00246
	(0.0289)
Observations	1,530
R-squared	0.361
Mean Dep Var	0.31
Origin District FE	yes
Year of Migration FE	yes

Notes: This table reports ordinary least squares estimation based on district-year level data prepared from the National Sample Survey Organization's Employment and Unemployment Survey 1983, following the specification in equation-14. Robust standard errors are provided in parentheses. The dependent variable is inter-district migration rate at the district-year level. The main explanatory variable is a dummy capturing whether the district-year cell has experienced a communal riot. The estimation includes district and year fixed effects. The main coefficient of interest captures whether the district-year cells affected by communal riots have more migration at the across-district level.

TABLE 3.4. Treatment Effect of Exposure on Migration of the Judges

	(1)	(2)	(3)	(4)	(5)	(6)
Class	X	XII	LLB	X	XII	LLB
***Kid*[0-6]**	0.00351	0.0281	-0.000214	0.0365	0.0473	-0.0571
	(0.051)	(0.049)	(0.041)	(0.073)	(0.068)	(0.060)
Observations	651	651	651	351	351	351
R-squared	0.167	0.165	0.436	0.190	0.234	0.443
Mean Dep Var	0.40	0.33	0.39	0.39	0.33	0.40
Home-district F.E	Yes	Yes	Yes	Yes	Yes	Yes
Year of Birth F.E	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All Judges	All Judges	All Judges	Bail Judges	Bail Judges	Bail Judges

Notes: This table tests for the effect of judges' exposure to riots during early childhood (0-6 years) on migration away from home district using the information on the districts where the judges completed their Class X (secondary schooling at age 15), XII (higher-secondary schooling at age 17) and LLB (undergraduate studies at age 21). The outcome variable is a dummy variable that takes the value 1 if the district where the judge completed a given degree is different from her home district. The estimation includes district and year fixed effects. The main coefficient of interest captures whether the exposed judges' migration for studies differ from that of non-exposed judges.

Table 3.5. Conflict Intensity and High State-Reponsiveness

CONFLICT INTENSITY SHARE DENIED	(1)	(2)
***K*id[0-6]**High Curfew*High Riot-Casualty	0.009 (0.031)	
***K*id[0-6]**High Curfew*Low Riot-Casualty	0.0774* (0.0418)	
***K*id[0-6]**High Arrests*High Riot-Casualty		-0.0173 (0.0331)
***K*id[0-6]**High Arrests*Low Riot-Casualty		0.134*** (0.0444)
Observations	4,573	4,609
R-squared	0.362	0.371
Mean Dep Var	0.36	0.36
Controls	yes	yes
home-district X Quarter F.E	yes	yes
Date of Birth F.E	yes	yes
District X Quarter F.E	yes	yes
Cluster Level	Judge	Judge
Total Number of Clusters	549	551

Notes: We report ordinary least squares estimations based on the judge–district–quarter level sample of bail judges. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at the judge–district–quarter level. All estimations include home district X quarter, year of birth, and district X quarter fixed effects, and a set of binary variables coding for past exposure up to 9 years of age. The main explanatory variable (Kid[0–6] is a binary measure of childhood exposure to communal conflict at 0–6 years of age) where high state action is interacted with low and high conflict severity. The median of the variable under consideration defines the threshold for severity to split treated judges equally into two groups. The median district areas' threshold value is 23 casualties (killed and injured). The median curfew is of 5 days duration and median number of arrests is 170. In Columns (1) and (2), we compare the judges experiencing high state action in terms of curfew and arrests, respectively, with judges not experiencing riots in their first 6 years of life.

Table 3.6. Judicial Stringency and Cognitive Skills

DEPENDENT VARIABLE: SHARE DENIED	(1)	(2)
***Kid*[0-6]**	0.0605***	0.0595***
	(0.0226)	(0.0223)
LLB Division	0.00669	
	(0.0163)	
Observations	5,530	5,530
R-squared	0.332	0.332
Mean Dep Var	0.37	0.37
home-district X Quarter F.E	yes	yes
Date of Birth F.E	yes	yes
District X Quarter F.E	yes	yes
Number of Judges	660	660

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level sample of 660 judges who are from home districts within UP. Robust standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at the judge–district–quarter level. Kid[0–6] is a dummy of childhood exposure to communal conflict. The controls include a dummy for Muslim, for female, and the experience of the judge at the time of judgment. Column 1 includes the dummy variable indicating whether the judge attained first division in the Bachelor of Law examination. Column 2 reports the estimates without including performance in this examination as control. All estimations include home district X quarter, year of birth, and district X quarter fixed effects.

Table 3.7. Heterogeneity by Religion of the Defendants

	(1) Hindu Judge	(2) Hindu Judge-Hindu Defendant	(3) Hindu Judge-Muslim Defendant
***Kid*[0-6]**	0.057**	0.051**	0.073**
	(0.025)	*(0.025)*	*(0.037)*
Observations	5,079	4,925	3,078
R-squared	0.348	0.321	0.399
Mean Dep Var	0.37	0.37	0.35
home-district X Quarter FE	yes	yes	yes
D.O.B Year FE	yes	yes	yes
District X Quarter FE	yes	yes	yes
Cluster	Judge	Judge	Judge
No of Judges	609	608	591

Notes: This table reports OLS estimations based on the judge-district-quarter level. Robust standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at the judge-district-quarter level. Kid[0-6] is a dummy for exposure to communal conflict between 0-6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home-district X quarter, year of birth, and district X quarter fixed effects. Column 1 includes cases handled by only the Hindu judges. Column 2 includes the cases handled by Hindu judges when all the defendants are Hindu. Column 3 includes the cases handled by Hindu Judges when all the defendants are Muslims.

D. Appendix Figures

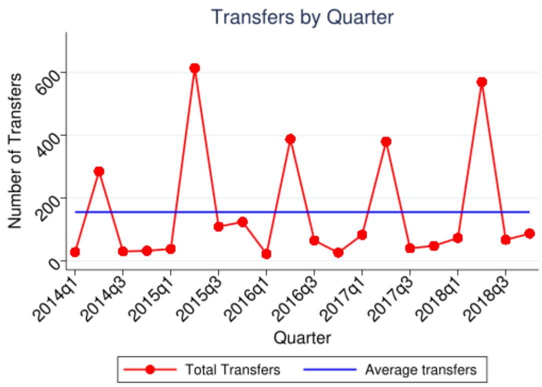


FIGURE 3.5. All Judges

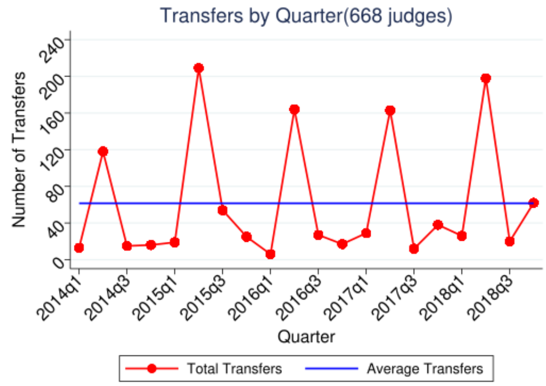
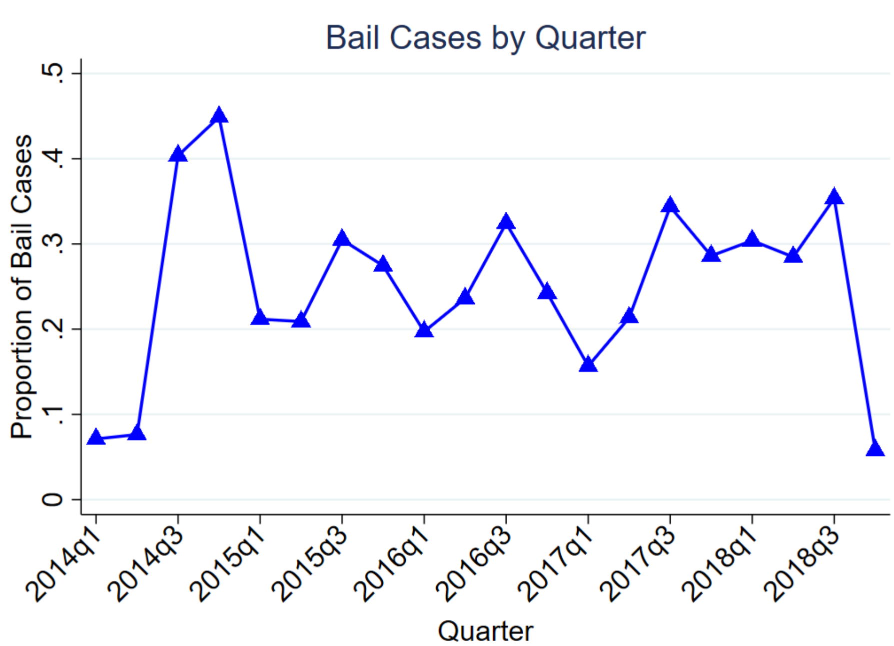


FIGURE 3.6. Sample Judges

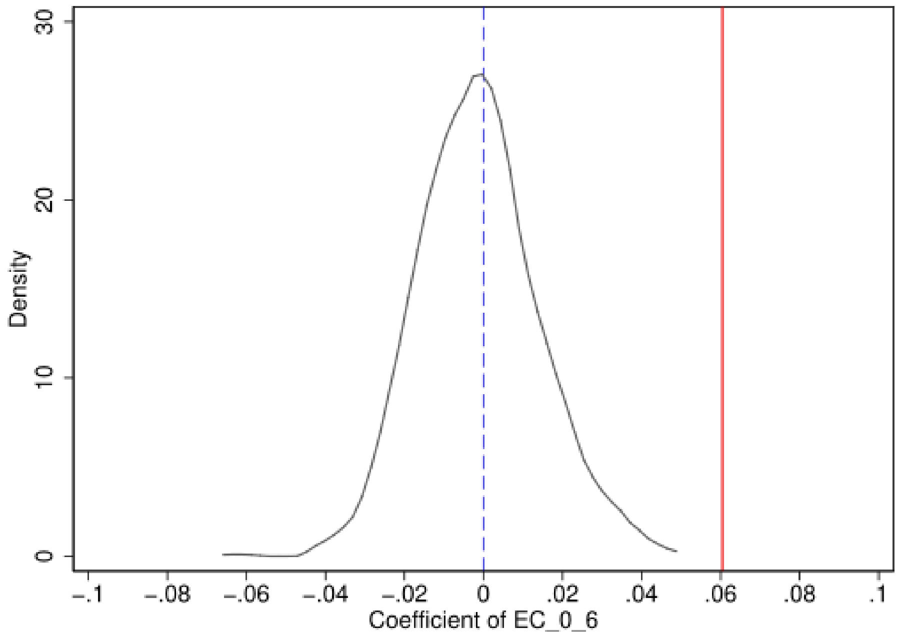
Notes: The figure reports the total and the average number of transfers of judges in the state of Uttar Pradesh for the full sample of judges and the 668 bail judges in our analysis sample. On average, 195 judges (7% of all judges) are transferred every quarter, with most transfers concentrated in the second quarter. Our analysis sample has a similar trend, with an average of 8.5% judges transferred every quarter.

FIGURE 3.7. Distribution of Bail Cases by Quarter



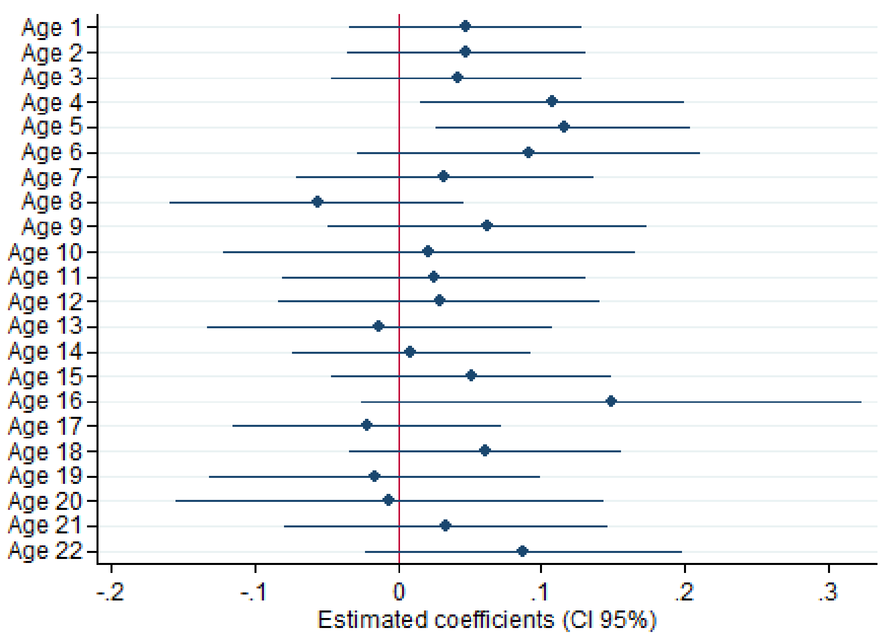
Notes: The figure highlights the cyclical nature of the reported bail case in UP district courts by quarter.

FIGURE 3.8. Distribution of Monte Carlo Treatment Effect Estimates



Notes: The figure displays the distribution of coefficients from our main regression (Column 2 of Table 3.2) obtained from 1,000 random draws of our variable Kid[0–6], keeping the proportion of judges (210/668) experiencing a communal conflict when aged between 0 and 6 years the same as in our original dataset.

FIGURE 3.9. Treatment Effects Estimates of Age at First Exposure



Notes: Estimated coefficients and the confidence interval at 95% are reported from ordinary least squares estimation using equation 17 on a sample of 612 judges born before 1979. The main dependent variable is the pretrial detention rate at the judge X district X quarter level. The estimation includes home district X quarter, birth year, district X quarter fixed effects.

E. Appendix Tables

Table 3.8. Creation of New Districts in Uttar Pradesh

State	New Districts	Parent Districts	Date of Formation	Comments	Source(link)
UP	Ambedkar Nagar	Faizabad	29-09-1995	Tehsils: Akbarpur, Tanda	https://ambedkarnagar.nic.in/
UP	Amethi	Sultanpur/Rae Bareilly	01-07-2010	Sultanpur Tehsils: Amethi, Gauriganj and Musafirkhana; Rae Bareilly Tehsils: Salon and Tiloi	https://amethi.nic.in/
UP	Amroha (Jyotiba Phule Nagar)	Moradabad	24-04-1997	Tehsils: Separating Amroha, Dhanora and Hasanpur	https://amroha.nic.in/
UP	Auraiya	Etawah	17-09-1997	Tehsils: Auraiya and Bidhuna	https://auraiya.nic.in/
UP	Balrampur	Gonda	25-05-1997		https://balrampur.nic.in/
UP	Baghpat	Meerut	01-05-1997	Tehsil: Baghpat	https://baghpat.nic.in/
UP	Bhadohi	Varanasi	30-06-1994		https://bhadohi.nic.in/
UP	Chandauli	Varanasi	20-05-1997		https://chandauli.nic.in/
UP	Chitrakoot	Banda	06-05-1997	Tehsils: Karwi and Mau	https://chitrakoot.nic.in/
UP	Gautam Buddha Nagar	Ghaziabad/ Bulandshahar	06-09-1997	Ghaziabad Blocks: Dadri and Bisrakh; Bulandshahar Blocks: Danakur and Jewar	https://gbnagar.nic.in/
UP	Ghaziabad	Meerut/ Bulandshahar	14-11-1976		https://censusindia.gov.in/2011census/dchb/ 0909_PART_B_DCHB_GHAZIABAD.pdf
UP	Hapur	Ghaziabad	28-09-2011	Tehsils: Hapur, Garhmukteshwar and Dhaulana	https://hapur.nic.in/
UP	Haridwar	Saharanpur	1988		https://haridwar.nic.in/
UP	Hathras (Mahamaya Nagar)	Aligarh/Mathura/ Agra	03-05-1997	Hathras tehsil came from Aligarh. Further it shares the Jail of Aligarh.	https://hathras.nic.in/
UP	Kannauj	Farrukhabad	18-09-1997		https://kannauj.nic.in/
UP	Kasganj (Kanshi Ram Nagar)	Etah	15-04-2008		https://kasganj.nic.in/
UP	Kaushambi	Allahabad	04-04-1997		https://kaushambi.nic.in/
UP	Kushinagar (Padrauna)	Deoria	13-05-1994		https://kushinagar.nic.in/
UP	Kanpur Dehat	Kanpur	23-04-1981		https://kushinagar.nic.in/
UP	Kanpur Nagar	Kanpur	23-04-1981		https://kanpurdehat.nic.in/
UP	Lalitpur	Jhansi	1974	Exact date unknown	https://lalitpur.nic.in/
UP	Maharajganj	Gorakhpur	02-10-1989		https://maharajganj.nic.in/
UP	Mahoba	Hamirpur	11-02-1995	Tehsil: Mahoba	https://mahoba.nic.in/
UP	Mau	Azamgarh	14-11-1988		https://mau.nic.in/
UK	Rudraprayag	Chamoli	18-09-1997		https://rudraprayag.gov.in/
UP	Sambhal	Moradabad	28-09-2011		https://sambhal.nic.in/
UP	Sant Kabir Nagar	Basti/ Siddarth Nagar	05-09-1997	Formed by carving : complete Khalilabad tehsil (from Basti), 131 villages of Basti tehsil and 161 villages of Santha block (Bansi tehsil, Siddarth Nagar)	https://sknagar.nic.in/
UP	Shamli	Muzaffarnagar	28-09-2011	Tehsils: Shamli and Kairana	https://shamli.nic.in/
UP	Shravasti	Bahraich	01-05-1997		https://shravasti.nic.in/
UP	Siddarth Nagar	Basti	29-12-1988		https://siddharthnagar.nic.in/
UP	Sonbhadra	Mirzapur	04-03-1989		https://sonbhadra.nic.in/
UK	Uddham Singh Nagar	Nainitaal	29-09-1997		https://usnagar.nic.in/

Notes: The table outlines the formation of new districts from parent (origin) districts in alphabetical order, their date/year of formation, and the source of this information. We assign all districts to their origin districts. With this harmonization, we arrive at 47 unique home districts for our sample of 668 judges. The details of district merging are explained in Appendix notes C.1.

Table 3.9. Descriptive Statistics

SUBSAMPLE:	**NON EXPOSURE 0-6 YRS.**	**EXPOSURE 0-6 YRS.**
TOTAL:	458	210
	A. JUDGES' CHARACTERISTICS	
Proportion of Females	0.096	0.157
(std. dev.)	*(0.014)*	*(0.025)*
Proportion of Muslims	0.083	0.062
(std. dev.)	*(0.013)*	*(0.017)*
Age	52.41	50.01
(std. dev.)	*(0.328)*	*(0.605)*
Experience	19.5	17.5
(std. dev.)	*(0.416)*	*(0.671)*
LLB division	1.74	1.65
(std. dev.)	*(0.021)*	*(0.033)*
	B. PEERS CHARACTERISTICS IN DISTRICT	
Number of Judges	34.96	33.78
(std. dev.)	*(0.278)*	*(0.421)*
Fraction of Female Judges	0.172	0.165
(std. dev.)	*(0.001)*	*(0.002)*
Fraction of Muslim Judges	0.063	0.058
(std. dev.)	*(0.001)*	*(0.001)*
Age of Peer judges	46.25	46.44
(std. dev.)	*(.034)*	*(0.051)*
Experience of Peer judges	14.16	14.39
(std. dev.)	*(0.038)*	*(0.054)*
	C. SOCIO-ECONOMICS OF DISTRICTS ASSIGNED TO JUDGES	
Proportion of Male	0.526	0.526
(std. dev.)	*(0.001)*	*(0.001)*
Proportion of Muslims	0.191	0.186
(std. dev.)	*(0.004)*	*(0.006)*
Proportion of SC/ST	0.209	0.215
(std. dev.)	*(0.002)*	*(0.003)*
Illiteracy Rate	0.325	0.325
(std. dev.)	*(0.003)*	*(0.003)*
Working Population	0.334	0.337
(std. dev.)	*(0.001)*	*(0.002)*

Notes: Notes: Subsection A presents the average observed characteristics of judges experiencing no riots (458 judges) and those experiencing riots (210 judges) in the first 6 years of childhood. Subsection B shows a subset of judge characteristics, as explained in Table 3.10. Subsection B presents average peers' characteristics (using the leave-me-out approach; that is, the judge is excluded when peers' characteristics are computed) at the district-quarter level. Subsection C presents the average district characteristics (computed from Census 2011 district-level data) of judges exposed to, and judges not exposed to, Hindu–Muslim riots when aged 0–6 years.

TABLE 3.10. Descriptive Statistics of UP District Court Judges

	All Judges			Bail Judges			Our Sample		
VARIABLES	N	mean	sd	N	mean	sd	N	mean	sd
Muslim Judge	2,434	0.061	0.240	1,236	0.067	0.250	668	0.0763	0.266
Female Judge	2,434	0.198	0.398	1,236	0.137	0.344	668	0.115	0.320
Age	2,434	46.55	11.29	1,236	49.58	8.537	668	51.66	7.687
Joining Age	2,434	31.81	4.687	1,236	32.76	4.920	668	32.81	5.114
Experience	2,434	14.74	11.26	1,236	16.82	9.283	668	18.85	9.202
Promotion Time Taken	1,651	7.471	2.260	974	7.169	2.356	542	7.560	2.352
Grade 10 Division	1,405	1.460	0.567	549	1.526	0.571	235	1.579	0.575
Grade 10 Age	1,446	14.99	0.984	579	14.93	0.970	246	14.91	0.963
Grade 12 Division	1,413	1.537	0.580	553	1.627	0.598	238	1.689	0.599
Grade 12 Age	1,446	17.11	1.080	579	17.06	1.056	247	17.05	1.085
LLB Division	2,385	1.621	0.488	1,206	1.697	0.460	668	1.711	0.454
LLB Age	2,405	23.72	2.134	1,219	23.64	2.119	668	23.42	2.031
Masters	2,434	0.386	0.487	1,236	0.360	0.480	668	0.356	0.479
PhD	2,434	0.015	0.121	1,236	0.016	0.126	668	0.0120	0.109
Number of Bachelors	2,431	1.863	0.363	1,234	1.938	0.272	668	1.958	0.253

Notes: The table presents the characteristics of all judges (including the judges who retired during 2014–2018 but had handled bail cases), bail judges, and our sample of judges (arrived at after removing judges handling less than 97 cases and judges with no information on the division obtained in the Bachelor of Law (LLB) examination). We notice that our sample of judges is very similar to the full sample of judges along different observable characteristics, such as the proportion of Muslim judges, the proportion of female judges, age (on December 31, 2018), joining age (age in years at which a judge enters into judiciary), experience, time to promotion (in years to become a Class I rank of judge), Grade 10 division (3 divisions: I ($\geq$ 60% marks), II ($\geq$ 45% and $<$ 60%), and III($\geq$ 33% and $<$ 45%), Grade 12 Division (3 divisions: I ($\geq$ 60% marks), II ($\geq$ 45% and $<$ 60%), and III($\geq$ 3% and $<$ 45%), LLB Division(I ($\geq$ 60% marks), II ($\geq$ 45% and $<$ 60%), and III($\geq$ 33% and $<$ 45%), Grade 10 Age (age at which a judge completes Class 10), Grade 12 Age (age at which a judge completes Class 12/Intermediate), LLB Age (age at which a judge obtains the LLB degree), Masters (dummy = 1 if a judge has a master's degree), PhD (dummy = 1 if a judge has a doctorate), and Number of Bachelors (number of bachelor's degrees of a judge). Only graduates were eligible to take the LLB examination earlier; currently, some institutions offer integrated courses such as BA + LLB, have started).

TABLE 3.11. Balance Table : Crime Categories (Full Sample)

	Body Crime	Prop Crime	Crim Intim	Cow Slaught	Elec. Theft	Others
***Kid*[0-6]**	-0.107	-0.770	-0.434	0.0624	-0.102	1.012
	(2.867)	(2.418)	(0.439)	(0.280)	(0.639)	(1.783)
Observations	5,530	5,530	5,530	5,530	5,530	5,530
R-squared	0.325	0.336	0.374	0.327	0.190	0.327
Mean Dep Var	15.47	14.16	4.52	1.21	1.25	10.82
Home Dist X Quarter F.E	yes	yes	yes	yes	yes	yes
Birth Year F.E	yes	yes	yes	yes	yes	yes
Dist X Quarter F.E	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge
No. of Clusters	660	660	660	660	660	660

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level on the full sample of judges. Standard errors in parentheses are clustered at the judge level. The dependent variable is the different categories of crime: Body Crime (Column 1), Property Crime (Column 2), Criminal Intimidation (Column 3), Cow Slaughter (Column 4), Electricity Theft (Column 5), and Others (Column 6). Kid[0–6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home district X quarter, year of birth, and district X quarter fixed effects. The classification of crime categories is based on Appendix Tables 3.32 and 3.33

Table 3.12. Balance Table : Crime Categories (Manual Sample)

	Body Crime	Prop Crime	Crim Intim	Cow Slaught	Elec. Theft	Others
***Kid*[0-6]**	-1.930	-1.002	-0.159	0.00155	0.587	-0.106
	(1.518)	(0.852)	(0.439)	(0.115)	(0.399)	(0.206)
Observations	3,811	3,811	3,811	3,811	3,811	3,811
R-squared	0.421	0.473	0.440	0.443	0.311	0.433
Mean Dep Var	5.14	4.34	1.76	0.40	0.32	0.82
Home Dist X Quarter F.E	yes	yes	yes	yes	yes	yes
Birth Year F.E	yes	yes	yes	yes	yes	yes
Dist X Quarter F.E	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge
No. of Clusters	525	525	525	525	525	525

This table reports ordinary least squares estimations based on the judge–district–quarter level of manually entered sample of judges. Standard errors in parentheses are clustered at the judge level. The dependent variable is the different categories of crime: Body Crime (Column 1), Property Crime (Column 2), Criminal Intimidation (Column 3), Cow Slaughter (Column 4), Electricity Theft (Column 5), and Others (Column 6). Kid[0–6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home district X quarter, year of birth, and district X quarter fixed effects. The classification of crime categories is based on Appendix Tables 3.32 and 3.33

Table 3.13. Pretrends in Case Characteristics

VARIABLES	(1) Tot Bail Case	(2) Def Religion Avail	(3) Muslim Def	(4) Bailable Info	(5) Non Bailable
1.ec_0_6	-5.240	-2.825	-0.409	-4.795	-4.059
	(10.05)	(6.448)	(1.165)	(8.590)	(7.103)
1.ec_0_6#6-months pre-transfer	11.98	9.226	1.698	10.64	9.641
	(14.08)	(8.942)	(1.640)	(12.27)	(9.954)
1.ec_0_6#5-months pre-transfer	9.697	7.551	0.501	8.855	7.845
	(12.39)	(7.880)	(1.448)	(10.79)	(8.796)
1.ec_0_6#4-months pre-transfer	4.210	3.269	0.496	3.841	3.719
	(11.26)	(7.042)	(1.377)	(9.659)	(7.897)
1.ec_0_6#3-months pre-transfer	10.50	7.219	0.0354	9.610	8.789
	(12.61)	(7.938)	(1.414)	(10.94)	(9.023)
1.ec_0_6#2-months pre-transfer	12.03	8.588	1.395	9.541	8.094
	(12.10)	(7.746)	(1.388)	(10.26)	(8.521)
1.ec_0_6#1-months pre-transfer	5.982	3.780	0.872	4.708	3.263
	(12.90)	(7.922)	(1.408)	(11.15)	(9.055)
1.ec_0_6#0-months pre-transfer	12.78	8.803	1.132	10.57	9.334
	(10.98)	(6.884)	(1.367)	(9.501)	(7.798)
1.ec_0_6#1-months post-transfer	5.545	4.283	0.0709	3.036	3.885
	(10.87)	(6.524)	(1.232)	(9.327)	(7.650)
1.ec_0_6#2-months post-transfer	6.580	4.934	1.079	5.268	4.097
	(11.25)	(6.868)	(1.339)	(9.758)	(7.983)
1.ec_0_6#3-months post-transfer	6.066	3.466	0.525	3.979	3.407
	(11.01)	(6.569)	(1.228)	(9.510)	(7.730)
1.ec_0_6#4-months post-transfer	6.921	2.341	0.726	4.366	3.496
	(11.40)	(6.931)	(1.300)	(9.947)	(8.282)
1.ec_0_6#5-months post-transfer	8.083	3.336	0.181	6.333	4.968
	(10.49)	(6.559)	(1.230)	(9.088)	(7.574)
1.ec_0_6#6-months post-transfer	-2.433	-2.309	-0.936	-2.417	-1.315
	(8.893)	(5.555)	(1.170)	(7.670)	(6.371)
Observations	19,656	19,656	19,656	19,656	19,656
R-squared	0.409	0.410	0.439	0.409	0.425
MeanDepVar	142.42	93.93	9.90	128.38	109.16
Home_dist FE	Yes	Yes	Yes	Yes	Yes
DOB YEAR FE	Yes	Yes	Yes	Yes	Yes
Dist FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
CLUSTER	id	id	id	id	id
NUM_clusters	649	649	649	649	649

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table reports ordinary least squares estimations based on the judge–district–month level sample of 668 judges with home districts in UP. Robust standard errors in parentheses are clustered at the judge level. The dependent variable are: Total Bail Cases (Column 1), Availability of defendant's religion (Column 2), Number of cases with Muslim defendants (Column 3), Availability of non-bailability information (Column 4) and Number of cases which are non-bailable (Column 5). The sample mean and the standard deviation of the dependent variables are reported. The explanatory variable is exposure to communal conflict between the ages of 0 and 6 years. All estimations include home district, year of birth and home district fixed effects. Exposure to riots until 9 years of age is included as control with a set of binary measures. Exposure variable is interacted with 12 months before and after. The results show 6 months pre and post transfer date with the reference month being 12th month post transfer.

Table 3.14. Pretrends in Crime Categories

VARIABLES	(1) Body Crime	(2) Arms Exp	(3) Prop Crime	(4) Forgery	(5) Cow Slaught	(6) Crim Intim
1.ec_0_6	-2.021	-0.311	-1.605	-0.0254	-0.0691	-0.307
	(2.765)	(0.539)	(2.736)	(0.620)	(0.415)	(1.080)
1.ec_0_6#6-months pre-transfer	4.543	0.616	4.518	0.319	-0.0840	0.0530
	(3.881)	(0.775)	(3.864)	(0.829)	(0.575)	(1.473)
1.ec_0_6#5-months pre-transfer	3.013	0.224	3.703	0.459	-0.136	0.598
	(3.583)	(0.644)	(3.446)	(0.799)	(0.524)	(1.399)
1.ec_0_6#4-months pre-transfer	2.468	0.539	1.746	0.264	0.0970	0.333
	(3.272)	(0.589)	(3.081)	(0.729)	(0.559)	(1.312)
1.ec_0_6#3-months pre-transfer	3.157	0.616	4.387	-0.0564	0.165	0.673
	(3.577)	(0.640)	(3.649)	(0.788)	(0.557)	(1.416)
1.ec_0_6#2-months pre-transfer	3.166	0.653	4.024	0.443	0.158	-0.519
	(3.313)	(0.595)	(3.482)	(0.735)	(0.589)	(1.265)
1.ec_0_6#1-months pre-transfer	2.694	0.488	1.543	-0.0764	-0.167	-0.763
	(3.685)	(0.675)	(3.434)	(0.748)	(0.606)	(1.349)
1.ec_0_6#0 months pre-transfer	3.406	0.648	3.992	0.995	0.310	-0.268
	(3.218)	(0.512)	(3.257)	(0.846)	(0.580)	(1.207)
1.ec_0_6#1-months post-transfer	1.592	0.373	2.849	0.355	-0.765	-1.005
	(3.033)	(0.550)	(3.065)	(0.745)	(0.555)	(1.159)
1.ec_0_6#2-months post-transfer	2.312	1.022	3.510	-0.233	-0.330	0.453
	(3.131)	(0.624)	(3.111)	(0.900)	(0.589)	(1.290)
1.ec_0_6#3-months post-transfer	2.432	0.764	1.941	0.0357	-0.257	-0.0398
	(2.968)	(0.572)	(3.000)	(0.885)	(0.545)	(1.311)
1.ec_0_6#4-months post-transfer	2.379	0.792	0.576	0.649	-0.259	0.485
	(3.080)	(0.627)	(3.097)	(0.895)	(0.564)	(1.383)
1.ec_0_6#5-months post-transfer	2.468	0.967	1.226	0.473	-0.375	1.375
	(2.926)	(0.638)	(2.701)	(0.757)	(0.523)	(1.298)
1.ec_0_6#6-months post-transfer	0.816	0.176	-0.813	-0.520	-0.298	-0.855
	(2.780)	(0.477)	(2.288)	(0.692)	(0.445)	(1.055)
Observations	19,656	19,656	19,656	19,656	19,656	19,656
R-squared	0.429	0.351	0.480	0.393	0.432	0.297
MeanDepVar	40.01	4.68	36.72	8.14	3.28	10.39
Home_dist FE	Yes	Yes	Yes	Yes	Yes	Yes
DOB YEAR FE	Yes	Yes	Yes	Yes	Yes	Yes
Dist FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Judge	Judge	Judge	Judge	Judge	Judge
Number of Clusters	649	649	649	649	649	649

This table reports ordinary least squares estimations based on the judge–district–month level sample of 668 judges with home districts in UP. Robust standard errors in parentheses are clustered at the judge level. The dependent variable is the different categories of crime: Body Crime (Column 1), Arms and Explosives (Column 2), Property Crime (Column 3), Forgery (Column 4), Cow Slaughter (Column 5) and Criminal Intimidation (Column 6). The sample mean and the standard deviation of the dependent variables are reported. The explanatory variable is exposure to communal conflict between the ages of 0 and 6 years. All estimations include home district, year of birth and home district fixed effects. Exposure to riots until 9 years of age is included as control with a set of binary measures. Exposure variable is interacted with 12 months before and after. The results show 6 months pre and post transfer date with the reference month being 12th month post transfer.

TABLE 3.15. F-test of Rotation Policy

Number of Districts X Quarter	**0**	**1**	**5**	**10**
***K*id[0-6]**	95	98.57	100	100
Female	92.45	98.67	100	100
Muslim	97.24	99.69	100	100
LLB First Division	93.37	98.57	100	100
Age	36.63	49.59	79.39	94.9
Joining Age	37.45	51.43	78.88	93.78
Time to Promotion	39.32	60.37	87.58	98.67

Notes: This table is based on the estimation results of Equation 11 on the main sample of judges at the district-quarter level. The null hypothesis $\hat{\beta_{d,q}} = \mu_h$ where μ_h refers to home district-level average of the judges' characteristics under consideration, such as childhood exposure to conflict when aged 0–6 years (Kid[0-6]), gender, religion, the division obtained in the Bachelor of Law examination, age on December 31, 2018, age at joining the judiciary and time to promotion as Class I officer. Each row represents the share of home districts for which the F-test of this null hypothesis is rejected at the 10% cutoff in at most 0, 1, 5, and 10 district-quarters.

TABLE 3.16. Balance Tests

PEERS At Dist-Qtly level	(1) Total	(2) Muslim	(3) Prop. Muslim	(4) Avg. Age	(5) Avg. Exp	(6) Female
***K*id[0-6]**	-0.354	-0.136	-0.004	0.118	0.140	-0.018**
	(2.016)	(0.173)	(0.005)	(0.198)	(0.218)	(0.009)
Observations	5,369	5,369	5,369	5,369	5,369	5,369
R-squared	0.199	0.197	0.175	0.207	0.211	0.226
Mean of Dep Var	34.72	2.05	0.06	46.32	14.25	0.17
Home-district X Quarter F.E	yes	yes	yes	yes	yes	yes
Year of Birth F.E	yes	yes	yes	yes	yes	yes
District X Quarter F.E	no	no	no	no	no	no
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge
Judge	642	642	642	642	642	660

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level sample of 668 judges with home districts in UP. Robust standard errors in parentheses are clustered at the judge level. The dependent variable are: Total Judges (Column 1), Number of Muslim Judges (Column 2), Proportion of Muslim Judges (Column 3), Average Age (Column 4), Average Experience (Column 5) and Proportion of Female Judges (Column 6). The sample mean and the standard deviation of the dependent variables are reported. The explanatory variable is exposure to communal conflict between the ages of 0 and 6 years. All estimations include home district X quarter, year of birth fixed effects. Exposure to riots until 9 years of age is included as control with a set of binary measures. The reduction from 668 is due to the dropping of singleton observations.

Table 3.17. Alternative Early Childhood Definition

SHARE DENIED	**(N=6)**	**(N=9)**	**(N=10)**	**(N=12)**
0-5 years **Pre Birth**	0.00791	0.00853	0.00816	0.00733
	(0.0200)	(0.0200)	(0.0201)	(0.0201)
Kid[0-N]	0.0623***	0.0421	0.0428	0.0397
	(0.0231)	(0.0277)	(0.0280)	(0.0260)
Observations	5,530	5,530	5,526	5,507
R-squared	0.332	0.329	0.329	0.329
Mean Dep Var	0.37	0.37	0.37	0.37
Home-district X Quarter FE	yes	yes	yes	yes
District X Quarter FE	yes	yes	yes	yes
Year of Birth FE	yes	yes	yes	yes
Controls	yes	yes	yes	yes
Cluster Level	Judge	Judge	Judge	Judge
Total Number of Clusters	660	660	659	655

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level. The dependent variable is pretrial detention rates. Kid[0–N] is a dummy of childhood exposure to communal conflict when aged 0–N years, where N is 6,9,10 and 12 from Column 1-4 respectively. All estimations include controls as in our main table. All estimations include home district X quarter fixed effects, district X quarter fixed effects and year of birth effects.

Table 3.18. Alternative Clustering Levels

SHARE DENIED	**(1)**	**(2)**	**(3)**	**(4)**
Kid[0-6]	0.0623***	0.0623**	0.0603**	0.0603**
	(0.0227)	(0.0250)	(0.0243)	(0.0250)
0-5 years **Pre Birth**	0.00791	0.00791	0.00123	0.00123
	(0.0194)	(0.0169)	(0.0205)	(0.0174)
Observations	5,530	5,530	5,443	5,443
R-squared	0.332	0.332	0.391	0.391
Mean Dep Var	0.37	0.37	0.37	0.37
Home-district X Quarter FE	yes	yes	yes	yes
District X Quarter FE	yes	yes	yes	yes
Year of Birth FE	yes	yes	no	no
Cohort X Quarter FE	no	no	yes	yes
Controls	yes	yes	yes	yes
Cluster Level	Home-District-Year of Birth	Home-District	Home-District-Year of Birth	Home-District
Total Number of Clusters	451	42	448	42

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level. The dependent variable is pretrial detention rates. Kid[0–6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home district X quarter fixed effects and district X quarter fixed effects. Year of birth effects are added in Column 1 and 2. Birth-Cohort X quarter fixed effects are added in Column 3 and 4. Standard errors in parentheses are clustered at the home district X year of birth level in Column 1 and 3; and at the home district level in Column 2 and 4.

TABLE 3.19. Selection into Occupation: Share of Riots Exposed Population

	Share of population exposed to Riots in 0-6 years of age			
	Total Working Population	**Total Judges**	**Total Bail Judges**	**Our Sample Judges**
% Exposed	39.20%	38.64%	38.88%	38.8%
# Population	26 993 272	2 197	890	500

Notes: The table reports the representation of early-childhood riot-exposed judges in the representative Indian sample (the National Sample Survey Organization's Employment-Unemployment Survey, 2011) and the sample of judges in Uttar Pradesh.

TABLE 3.20. Does riot affected district-year sends different types of judges ?

	Total Judges	Total Judges	Prop Female Judges	Prop Muslim Judges
Exposed	-0.129	0.00603	0.0196	-0.0178
	(0.498)	(0.346)	(0.0215)	(0.0160)
Observations	694	609	694	694
R-squared	0.811	0.581	0.526	0.508
Mean Dep Var	9.57	5.42	0.24	0.10
Home Dist F.E	yes	yes	yes	yes
Year F.E	yes	yes	yes	yes
Riot Exposure	0-6 yrs	3-6 yrs	0-6 yrs	0-6 yrs

Notes: This table reports ordinary least squares estimations based on the home district-year level sample using the following equation

$$Judges_{h,y} = \alpha + \eta_h + \delta_y + \beta Exposed_{h,y} + \epsilon_{h,y}$$

η_h and δ_y are home-district and year fixed effects. $Judges_{h,y}$ denotes dependent variable. Robust standard errors are in parentheses. $Exposed_h, y$ is a dummy that indicates whether the home district-year had a riot. The dependent variables are the number of judges (Columns 1 and 2), Proportion of Female Judges (Column 3), and Proportion of Muslim judges (Column 4) from the exposed home district-year in the first 6 years (after riot) in Columns 1, 3, and 4 and in 3–6 years in Column 2.

TABLE 3.21. Bail Decisions and Early Exposure to Riots (Case Level Regressions)

	(1)	(2)	(3)
***K*id[0-6]**	0.0381**	0.0435**	0.0409**
	(0.019)	(0.017)	(0.017)
Muslim Defendant			-0.018***
			(0.005)
Non Bailable			0.155***
			(0.014)
Observations	323,194	323,194	196,300
R-squared	0.081	0.160	0.186
Mean Dep Var	0.34	0.34	0.34
Home Dist X Quarter F.E	yes	yes	yes
Birth Year F.E	yes	yes	yes
Dist X Quarter F.E	yes	yes	yes
Crime Type FE	no	yes	yes
Controls	yes	yes	yes
Cluster Level	Judge	Judge	Judge
No. of Clusters	668	668	668

Notes: This table reports ordinary least squares estimations based on the case level. Standard errors in parentheses are clustered at the judge level. The dependent variable is a dummy for whether the bail is rejected. Kid[0–6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age and our usual controls. All estimations include home district × quarter, year of birth, and district × quarter fixed effects. Columns 2 and 3 further include crime-type fixed effects. Column (3) also adds a dummy for whether the defendant is Muslim and whether the case is nonbailable in nature (i.e., booked crimes are nonbailable according to the Indian Penal Code).

TABLE 3.22. Changing threshold of Judges Handling few cases

SHARE DENIED	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
***K*id[0-6]**	0.0378*	0.0466**	0.0508**	0.0525**	0.0605***	0.0601**	0.0647**	0.0667**	0.0681**	0.0760***
	(0.0206)	(0.0211)	(0.0216)	(0.0224)	(0.0226)	(0.0236)	(0.0250)	(0.0263)	(0.0269)	(0.0280)
Observations	6,427	6,119	5,889	5,695	5,530	5,370	5,179	5,020	4,877	4,706
R-squared	0.306	0.316	0.315	0.326	0.332	0.331	0.341	0.345	0.350	0.364
Mean Dep Var	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.36
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Home-district X Quarter F.E	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year of Birth F.E	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
District X Quarter F.E	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Bottom Threshold percentile	1	2	3	4	5	6	7	8	9	10
Num of Bail cases in 4yrs	32	53	69	87	97	109	118	130	143	167
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge	Judge	Judge	Judge	Judge
No. of Clusters	875	794	737	694	660	627	595	570	545	522

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level sample of judges from home districts within UP. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate. Kid[0-6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables, coding for past exposure up to 9 years of age. All estimations include home district X quarter, year of birth, and district X quarter fixed effects. From Columns 1–10, we exclude the judges falling in the bottom 1 to 10 percentiles. The threshold in terms of the number of bail cases handled during 2014–2018 is also provided.

Table 3.23. Excluding High Ranked Judges

	(1)	(2)	(3)
***K*id[0-6]**	0.0835***	0.0538**	0.0768***
	(0.0255)	(0.0231)	(0.0259)
Observations	4,772	5,290	4,538
R-squared	0.358	0.330	0.358
Mean Dep Var	0.37	0.36	0.36
Home Dist X Quarter F.E	yes	yes	yes
Birth Year F.E	yes	yes	yes
Dist X Quarter F.E	yes	yes	yes
Controls	yes	yes	yes
Excluding	**DSJ**	**CJM**	**DSJ+CJM**
Cluster Level	Judge	Judge	Judge
No. of Clusters	597	646	583

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level. Standard errors in parentheses are clustered at the judge level. The dependent variable is pretrial detention rates. Kid[0–6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home district X quarter, year of birth, and district X quarter fixed effects. We exclude District and Session Judges (DSJs) in Column 1; Chief Judicial Magistrates (CJMs) in Column 2; and both DSJs and CJMs in Column 3.

Table 3.24. Outlier Tests

SHARE DENIED	(1)	(2)	(3)	(4)	(5)	(6)	(7)
***K*id[0-6]**	0.0745***	0.0567**	0.0590**	0.0675***	0.0707***	0.0579**	0.0622***
	(0.0236)	(0.0229)	(0.0231)	(0.0247)	(0.0252)	(0.0230)	(0.0235)
Observations	5,283	5,231	5,416	5,201	5,243	5,430	5,276
R-squared	0.341	0.341	0.328	0.341	0.334	0.337	0.339
Mean Dep Var	0.37	0.37	0.37	0.37	0.37	0.37	0.37
Controls	yes	yes	yes	yes	yes	yes	yes
home-district X Quarter F.E	yes	yes	yes	yes	yes	yes	yes
Date of Birth F.E	yes	yes	yes	yes	yes	yes	yes
District X Quarter F.E	yes	yes	yes	yes	yes	yes	yes
Home Dist Removed	Aligarh	Meerut	Moradabad	Bulandshahr	Varanasi	Lucknow	Allahabad
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge	Judge
No. of Clusters	632	629	645	618	628	646	628

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level sample of judges from home districts within UP. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate. Kid[0-6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home district X quarter, year of birth, and district X quarter fixed effects. From Column 1 to Column 6, we exclude from the full sample judges coming from one of the top six home districts *one at a time* (Aligarh(26), Meerut(24), Moradabad(18), Bulandshahar(15), Varanasi(13), Lucknow(13), and Allahabad (13)) that have the maximum number of riots during 1950–2000.

Table 3.25. Outlier Tests

SHARE DENIED	(1)	(2)	(3)	(4)	(5)	(6)	(7)
***Kid*[0-6]**	0.0745***	0.0693***	0.0679***	0.0727***	0.0880***	0.0855***	0.0875***
	(0.0236)	(0.0240)	(0.0246)	(0.0266)	(0.0298)	(0.0304)	(0.0329)
R-squared	0.341	0.350	0.347	0.358	0.360	0.366	0.375
Observations	5,283	4,984	4,871	4,542	4,248	4,147	3,892
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge	Judge
No. of Clusters	632	601	586	544	512	498	466
Mean Dep Var	0.37	0.37	0.37	0.37	0.37	0.37	0.37
Controls	yes	yes	yes	yes	yes	yes	
Home Dist X Quarter F.E	yes	yes	yes	yes	yes	yes	yes
Birth Year F.E	yes	yes	yes	yes	yes	yes	yes
Dist X Quarter F.E	yes	yes	yes	yes	yes	yes	yes

This table reports OLS estimations based on the judge-district-quarter level sample of judges coming from home-districts within UP. Standard errors in parentheses are clustered at the judge level. Kid[0-6] is a dummy of childhood exposure to communal conflict between the age of 0-6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home-district X quarter, year of birth and district X quarter fixed effects. From Column 1 to Column 6 we exclude from the full sample judges *cumulatively* coming from one of the top seven home-districts (Aligarh(26), Meerut(24), Moradabad(18), Bulandshahar(15), Varanasi(13), Lucknow(13) and Allahabad (13)) that have the maximum number of riots in between 1950-2000.

Table 3.26. Outlier Tests

	(1)	(2)	(3)	(4)	(5)	(6)
Excl.	1σ	2σ	3σ	Leverage	Influence measure (dfbeta)	Influence measure (Cook's distance)
***Kid*[0-6]**	0.0532***	0.0577**	0.0605***	0.0639***	0.0605***	0.0578***
	(0.0171)	(0.0225)	(0.0226)	(0.0230)	(0.0226)	(0.0210)
Observations	5,252	5,519	5,530	4,992	5,524	4,991
R-squared	0.457	0.341	0.332	0.288	0.330	0.413
Mean Dep Var	0.35	0.37	0.37	0.36	0.37	0.35
Home Dist X Quarter F.E	yes	yes	yes	yes	yes	yes
Birth Year F.E	yes	yes	yes	yes	yes	yes
Dist X Quarter F.E	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes
Cluster Level	Judge	Judge	Judge	Judge	Judge	Judge
No. of Clusters	660	660	660	646	660	657

Notes: This table reports ordinary least squares estimations based on our main analysis sample at the judge–district–quarter level. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate. Kid[0–6] is a dummy for childhood exposure to communal conflict when aged 0–6 years. All estimations include a set of binary variables coding for past exposure up to 9 years of age. All estimations include home district × quarter, year of birth, and district × quarter fixed effects. In Column 1, Column 2, and Column 3, we remove observations that are 3, 2, and 1 standard deviation away from the residual mean. In Column 4, we remove observations with high leverage. In Column 5, we remove observations that shift estimates to at least one standard error. In Column 6, we remove observations that shift estimates at least to 4/N.

Table 3.27. Political Competition

	Pol Comp 1	Pol Comp 2	Turnout
***K*id[0-6]**	16.83	0.00455	-0.182
	(11.25)	(0.00988)	(0.475)
Observations	664	664	664
R-squared	0.100	0.097	0.201
Mean Dep Var	189.74	-0.57	53.67
Home Dist F.E	yes	yes	yes

Notes: This table reports ordinary least squares estimations based on the judge level. The dependent variables are at state legislative elections: difference in the number of seats between main party (Congress) and main competitor(Janta Party before 1990's and Samajwadi Party afterwards) in Column 1; absolute difference in the proportion of seats between main party and main competitor in Column 2 and proportion of eligible voters voting in Column3. Kid[0–6] is a dummy of childhood exposure to communal conflict when aged 0–6 years. All estimations include home district fixed effects. The dependent variables are created based on the Besley and Burgess, 2002.

Table 3.28. Conflict Intensity

PANEL A: CONFLICT INTENSITY SHARE DENIED	**(1)**	**(2)**	**(3)**	**(4)**
***K*id[0-6]**: Killed in Riots (High)	0.021			
	(0.031)			
***K*id[0-6]**: Killed in Riots (Low)	0.079***			
	(0.026)			
***K*id[0-6]**: Casualties in Riots (High)		-0.026		
		(0.028)		
***K*id[0-6]**: Casualties in Riots (Low)		0.096***		
		(0.026)		
***K*id[0-6]**: No of Riots (High)			0.016	
			(0.028)	
***K*id[0-6]**: No of Riots (Low)			0.076***	
			(0.024)	
***K*id[0-6]**: Duration of Riots (High)				0.018
				(0.033)
***K*id[0-6]**: Duration of Riots (Low)				0.070***
				(0.025)
Observations	5,530	5,530	5,530	5,530
R-squared	0.335	0.338	0.336	0.335
Mean Dep Var	0.37	0.37	0.37	0.37
Controls	yes	yes	yes	yes
home-district X Quarter F.E	yes	yes	yes	yes
Date of Birth F.E	yes	yes	yes	yes
District X Quarter F.E	yes	yes	yes	yes
Cluster Level	Judge	Judge	Judge	Judge
Total Number of Clusters	660	660	660	660

Notes: We report ordinary least squares estimations based on our main analysis sample at the judge–district–quarter level. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at the judge–district–quarter level. All estimations include home district X quarter, year of birth, and district X quarter fixed effects, and a set of binary variables coding for past exposure up to 9 years of age. The main explanatory variable (Kid[0-6]—a binary measure of childhood exposure to communal conflict when aged 0–6 years) is interacted with high and low conflict severity. The median of the variable under consideration defines the threshold for severity to split treated judges equally into two groups. The median district areas' threshold values are four persons killed for Column 1, 23 casualties (killed + injured) for Column 2, one riot exposed person in Column 3, and 3 days of riots in Column 4.

TABLE 3.29. Impact of Early Violence Exposure: Split into Three Age Groups

SHARE DENIED	(1)	(2)	(3)	(4)
0-3 years **Pre Birth**	-0.00283	0.00440	0.00189	-0.00961
	(0.0208)	(0.0204)	(0.0210)	(0.0237)
Kid[0-3]	0.0289	0.0270	0.0281	0.0181
	(0.0242)	(0.0247)	(0.0258)	(0.0280)
Kid[3-6]	0.0881***	0.0905***	0.0858***	0.0905***
	(0.0283)	(0.0294)	(0.0293)	(0.0312)
Kid[6-9]	-0.0102	-0.00384	-0.00801	-0.0156
	(0.0229)	(0.0236)	(0.0250)	(0.0275)
Observations	5,530	5,488	5,425	5,248
R-squared	0.334	0.337	0.340	0.345
Mean Dep Var	0.37	0.37	0.37	0.37
home-district X Quarter FE	yes	yes	yes	yes
Year of Birth FE	yes	yes	yes	yes
District X Quarter FE	yes	yes	yes	yes
Controls	yes	yes	yes	yes
exp[N]	9	14	18	22
Cluster Level	Judge	Judge	Judge	Judge
Total Number of Clusters	660	651	637	604

Notes: This table reports ordinary least squares estimations based on the judge–district–quarter level. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at judge–district–quarter level. 0–3 years Pre-Birth, Kid[0–3], Kid[3–6], and Kid[6–9] are mutually exclusive binary measures of childhood exposure to communal conflict. All estimations include home district X quarter, year of birth, and district X quarter fixed effects. Column (2) to Column (4) extends the past exposure control for later years: up to age 14 years (Column (2): keeping judges born before 1986), age 18 years (Column (3): keeping judges born before 1982), and age 22 years (Column (4): keeping judges born before 1978) on a subsample of judges.

TABLE 3.30. Age at First Exposure

SHARE DENIED	
Age at First Exposure at 1yr	0.043
	(0.033)
Age at First Exposure at 2 yrs	0.042
	(0.034)
Age at First Exposure at 3 yrs	0.028
	(0.039)
Age at First Exposure at 4 yrs	0.081**
	(0.039)
Age at First Exposure at 5 yrs	0.118***
	(0.038)
Age at First Exposure at 6 yrs	0.090*
	(0.052)
Observations	5,530
R-squared	0.336
Mean Dep Var	0.37
home-district X Quarter FE	yes
Year of Birth FE	yes
District X Quarter FE	yes
Controls	yes
Cluster	Judge
Number of Judges	660

Notes: This table reports the ordinary least squares estimation specification of equation-17. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at the judge–district–quarter level. The controls in the specification are gender, religion, performance in the Bachelor of Law examination, experience, and age at first exposure at 7, 8, and 9 years.

Table 3.31. Heterogeneity by Conflict Intensity: Alternative Thresholds

CONFLICT INTENSITY SHARE DENIED	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
***K*id[0-6]**: Killed in Riots (High)	0.025			-0.005	-0.004
	(0.036)			*(0.038)*	*(0.037)*
***K*id[0-6]**: Killed in Riots (Low)	0.068***			0.075***	0.073***
	(0.024)			*(0.024)*	*(0.024)*
***K*id[0-6]**: Casualties in Riots (High)		-0.004			
		(0.033)			
***K*id[0-6]**: Casualties in Riots (Low)		0.077***			
		(0.026)			
***K*id[0-6]**: Duration of Riots (High)			0.003		
			(0.039)		
***K*id[0-6]**: Duration of Riots (Low)			0.066***		
			(0.024)		
Observations	5,530	5,530	5,530	5,530	5,530
R-squared	0.333	0.335	0.334	0.335	0.335
Mean Dep Var	0.37	0.37	0.37	0.37	0.37
Controls	yes	yes	yes	yes	yes
home-district X Quarter F.E	yes	yes	yes	yes	yes
Date of Birth F.E	yes	yes	yes	yes	yes
District X Quarter F.E	yes	yes	yes	yes	yes
Cluster Level	Judge	Judge	Judge	Judge	Judge
Total Number of Clusters	660	660	660	660	660
Threshold Type	p75	p75	p75	***K*id[0-6]**: Killed per sq km >.002	***K*id[0-6]**: Killed > 10

Notes: This table reports ordinary least squares estimations based on our main analysis sample at the judge–district–quarter level. Standard errors in parentheses are clustered at the judge level. The dependent variable is the pretrial detention rate at the judge–district–quarter level. All estimations include home district X quarter, year of birth, and district X quarter fixed effects and a set of binary variables coding for past exposure up to 9 years of age. The main explanatory variable (Kid[0–6]—a binary measure of childhood exposure to communal conflict when aged 0–6 years) is interacted with high and low conflict severity. The threshold in Columns 1 to 3 is defined by the 75 percentile values from the distribution of variables under consideration at the district month level. Column 4 uses the threshold of killed per sq. km greater than 0.002, which implies a median area of the district having 10 persons killed. Column 5 uses an absolute threshold of 10 persons killed in the riot.

TABLE 3.32. IPC Categorization

IPC Chapters	Topic	Sections	Categories
	Of Offences Affecting Human Body		
16	Offences affecting Life	299-318	Body_crime
	Hurt	319-338	
	Wrongful restraint/Confinement	339-348	
	Criminal Force and Assault	349-358	
	Kidnapping, Abduction, Slavery, Forced Labour	359-374	
	Sexual Offences	375-377	
	Offences Against Property		
17	Theft	378-382	Property Crime
	Extortion	383-389	
	Robbery and Dacoity	390-	
	Criminal Misappropriation of property	403-404	
	Criminal Breach of Trust	405-409	
	Receiving of Stolen Property	410-414	
	Cheating	415-420	
	Fraudelent deeds and disposition of property	421-424	
	Mischief	425-440	
	Criminal Trespassing	441-462	
18	Forgery	463-489	Forgery
11	Of False Evidence and Offences against public justice	191-229	
12	Of Offences relating to Coin and Government Stamps	230-263	
22	Criminal Intimidation	503-510	Criminal Intimidation
8	Offences against Public Tranquility	141-160	Public Tranquility
15	Of offences relating to religion	295-298	
14	Of Offences Affecting Public health, safety, convenience, decency and morals	268-294	Public Health
20	Offences Relating to Marriage	493-498	Other
20A	Cruelty by Husband	498A	
5	Of Abetment	107-120	
5A	Criminal Conspiracy	120A, 120B	
6	Offences against State	121-130	
7	Offences relating to the Army, Navy and Air Force	131-140	
9	Of Offences by or relating to public servants	161-171	
10	Of Contempts of the Lawful authority of public servants	172-190	
13	Of Offences relating to Weights and measures	264-267	
19	Criminal Breach of Contracts of Service	490-492	
21	Defamation	499-502	
23	Attempts to Commit Offences	511	

Notes: This table presents the list of offences included in our crime categories based on the Indian Penal Code (IPC). The crime categories are created using the IPC chapters—in a way, following the legal classification. The cases are lodged under one or more IPC sections, which are retrieved from the judgment documents.

TABLE 3.33. Acts Categorization

Act	Section(s)	Categories
Arms Act	25	Arms and Explosives
Explosive Substances Act	5	
Cow Slaughter Act	8	Cow Slaughter
Prevention of Cruelty to Animals Act	11	
Electricity Act	135	Electricity
Gangsters Act	3	Gangs and Dacoits
UP Goonda Act	10	
UP Dacoity Affected Areas	12	
Copyright Act	63	Other
Dowry Prohibition Act	3	
Essential Commodities Act	7	
Examination Act	9/10	
Forest Conservation Act	26	
Gambling Act	3	
Immoral Traffic Prevention Act	3/5/6	
Indian Forest Act	26	
Indian Medical Council Act	15	
IT Act	66	
Mines and Minerals, Development and Regulation Act	21	
Motor Vehicle Act	207	
Narcotic Drugs and Psychotropic Substances Act	20	
Negotiable Instruments Act	138	
Petroleum Act	15	
Prevention of Atrocities Act (SC ST)	3	
Prevention of Children from Sexual Offences Act	4/8	
Prevention of Corruption Act	13	
Prevention of Damage to Public Property Act	3	
Prevention of Food Adulteration Act	16	
Railway Act	143	
Railway Property, Unlawful possession Act	3	
Representation of the People Act	136	
Trademark Act	103	
UP Excise Act	60/63	
UP Excise Act	63	

Notes: This table presents the list of offences included in the crime categories based on the special Central and State Acts. States have formed special laws from time to time to include different types of offences that are not dealt with in the Indian Penal Codebook (IPC) in detail or require special attention. If a case is filed under both IPC and some Acts, we first use IPC to create crime categories; otherwise, we use this table to create additional crime categories. The Arms and Explosives category includes offences related to illegal carrying/selling/use of arms or explosives. Cow Slaughter includes offences related to the slaughtering of cows and the infliction of unnecessary suffering on animals. Electricity thefts are included under the Electricity category. Gangs- and Dacoity-related Acts are passed to handle very severe crimes/criminals. Others include all the Acts that are not categorized above.

Table 3.34. Alternative Origin Districts

SHARE DENIED	(1)	(2)	(3)
Kid[0-6]	0.069***	0.070***	0.072***
	(0.0237)	(0.0239)	(0.0236)
home-district X Quarter FE	yes	yes	yes
Year of Birth FE	yes	yes	yes
District X Quarter FE	yes	yes	yes
Controls	yes	yes	yes
Observations	5,531	5,537	5,539
R-squared	0.333	0.334	0.334
Mean Dep Var	0.37	0.37	0.37
Judge	660	660	660
Allocations to Parent Districts	Ghaziabad to Meerut	Firozabad to Mainpuri	Hathras to Agra

Notes: This table reports the impact of exposure to communal violence between ages of 0 and 6 years under the alternative classification of the new districts that were carved out of the origin districts. We examine alternative assignments of the newly formed district to origin districts that were not tested in our main analysis in the main text. We find the effects of exposure to communal violence between the ages of 0 and 6 are similar across different district allocations and statistically significant at the 1% level of significance. Here we sequentially change the home districts. Column 1 assigns Ghaziabad to Meerut, Column (2) further assigns Firozabad to Mainpuri (keeping Ghaziabad with Meerut), and Column (3) assigns Hathras to Agra (maintaining the changes in Columns (1) and (2)).

Table 3.35. Treatment Effect of Riot Exposure on Measurement Error

ERROR IN TEXT EXTRACTION	(1) BAIL DECISIONS	(2) DEFENDANT RELIGION
Kid[0-6]	0.008	-0.003
	(0.147)	(0.005)
Observations	57,330	36,526
R-squared	0.125	0.228
Mean Dep Var	0.05	0.06
home-district X Quarter FE	yes	yes
Year of Birth FE	yes	yes
District X Quarter FE	yes	yes

Notes: This table reports ordinary least squares estimation based on case level data for the randomly selected data for manual entry. Robust standard errors are provided in parentheses. The dependent variable is the measurement error in the text extraction process which is equal to 1 if the text extraction values differ from the manual entry and 0 otherwise. In Column(1), the dependent variable is the error rate in the bail outcome and in Column (2) is a dummy variable if the religion of the dependent is identified correctly. All estimations include home district X quarter, year of birth, and district X quarter fixed effects, and a set of usual controls. Column(2) also controls for the religion of defendant. The main coefficient of interest on Kid[0–6] capturing whether the error in our text extraction is correlated to our dependent variables.

F. Appendix Notes

F.1. Merging of Districts in Uttar Pradesh. We use the information on the judges' home-district and birth year from their resumes and merge it with the information on years and districts of riot incidents from the conflict data to identify the judges' exposure to communal violence.

After independence, several changes have been made to the district boundaries in the state of Uttar Pradesh. In Appendix Table 3.8, we provide a list that tracks the formation of new districts from their origin districts. We adopt the following approach to provide a consistent measure of being a treated/control judge. The first case in our study context relates to districts that are broken down into smaller districts. The smaller districts that were a part of this origin district are assigned to the origin district. For example, if district D is split into D1 and D2, we treat both D1 and D2 as D. The logic is that if D1 and D2 at time "t" were treated as district D, then D1 and D2 at time "t + k" will also be treated as district D. The second case is when a new district is carved out by merging several other districts. For example, the district Hathras was formed from Aligarh, Mathura, and Agra in 1997. In this case, we have three possible ways to rename Hathras. We randomly choose one of the three origin districts. For our analysis, we assign Hathras to Aligarh. The third case is when districts are disintegrated sequentially over time. For example, a new district D1 is carved out of a district D at time "t." Then, at time "t + k" district D2 is carved out of D1. Here, we assign D1 and D2 to D. For example, the district Ghaziabad was formed from Bulandshahar and Meerut in 1976 and Hapur was created from Ghaziabad in 2011. We categorize Hapur as Ghaziabad, and then can reassign Ghaziabad in two ways: as Bulandshahar or as Meerut. In our main analysis, we assign Ghaziabad to Bulandshahar.

Further, we implement sensitivity checks by using alternative origin-district allocations and find that the results on the impact of early-childhood exposure to conflict between ages of 0 and 6 years are stable, robust, and statistically significant at the 1% level of significance across assignments to these districts, as demonstrated in Appendix Table-3.34.

F.2. Creation of Variables. The first step involves identifying the bail cases from the universe of all the downloaded cases. For every downloaded case, we use the information on whether the case is a "Bail Application" to identify whether the case is a bail case.

We also extract the case number and year. Our extracted sample consists of 423,000 bail applications from the entire pool of two million downloaded cases.

In Uttar Pradesh, Hindi and English are the official languages for the district-level judiciary. However, since Hindi is predominantly spoken in the state, most of the judgments are written in the Hindi (Devanagari) script. We translate all the Hindi judgments using Google Translate (via the inbuilt libraries of Python) into English to extract relevant information. Further, some of the judgments are uploaded in a format with different encoding to avoid duplication. For those judgments, we use optical character recognition to convert Hindi judgments (pdfs) into correctly encoded texts before translating them into English. We use these translated judgments to extract our essential variables.

The written judgments can be divided into three parts. The first part contains information about the case, such as the defendant(s)'s name (often accompanied by the father's name, age, and address), the criminal section under which they have been charged, the date of the judgment, and the name of the judge who delivered the judgment. The second part details the event that led to the filing of the case. Sometimes, it also contains the legal precedents followed in arriving at the decision. The last part contains the judges' decision on the bail application, either granting or denying bail and the bail amount if granting bail. We use the first and the last part of the written judgments in extracting our variables. The main content we obtain is through text extraction and is as follows.

i) Outcome of the bail decisions: The outcome of the bail decisions is present in the last part of the judgment, and we use negative words to identify whether bail was denied[42] and positive words to identify whether bail was granted.[43]

ii) Name of the defendant(s): The name of defendants is used to identify their religion. The information is present in the first part of the judgment. We use the "Stanford Named Entity Algorithm."[44] One concern about using this algorithm is that it is not perfectly suitable for Hindi names. It may have missed some names (or some part of the

[42]That is, "not granted," "not released," "not accepted," "not acceptable," "not approved," "not freed," "denied," "unacceptable," "cancelled," "canceled," "aborted," "dismissed," "rejected," "abrogated," "abortable," "to be canceled," "terminated," "cancellation," "suspended," and "revoked."

[43]That is, "released on bail," "granted," "released," "accepted," "acceptable," "approved," "freed," "acquitted," "surrender," "personal bond," "bond," "bondage," "security bond," "bond," "suretages bond," "interim," "amount," "money," "acceptance," "sureties," and "collateral."

[44]It associates each word to four tags: person (PER), location (LOC), organization (ORG), and miscellaneous (MISC). We consider the words tagged as PER.

names). Sometimes, it extracts the names of places (partly because, in a few cases, it is difficult to differentiate between the names of places and people. For instance, from the name Gautam Buddha Nagar district, it picked Gautam Buddha, which can be the name of a person). We filter out the names of places using the names of districts, subdistricts, and cities. The algorithm also selected the judges' names (and their variants, given that it was translated using Google Translate), which we remove carefully. Further, we screen all non-names that the algorithm selected. Even after all these checks, some scope for error remains.

After extracting the names, we use the Nilabhra name2community algorithm to identify the defendants' religion. We discuss the efficiency of the Nilabhra algorithm in Subsection C3. In some cases, only nicknames are provided (e.g., Bhura and Kaalu). In such cases, we use the name/surname of the defendant's father to identify their religion.

iii) We use the text extraction method to extract information on the Acts[45] and the Indian Penal Code (IPC) sections from the judgment documents. Criminal cases are registered either through a First Information Report or a Complaint Register by the police. The IPC is the official criminal code of India. It provides a comprehensive list of offences and associated punishments and states whether the offence is bailable or not. In addition to the IPC, special Acts passed by the central and the state government guide the categorization of crimes. Appendix Table 3.32 and 3.33 provide the list of all the sections/Acts under each crime category. A case can be lodged under one or more IPC sections and/or under special Acts. For every case, we use the offence (IPC/Act section) carrying the maximum punishment to categorize the case. When there is a tie, we are indifferent and randomly pick one section. We create 11 crime categories: Arms and Explosives, Body Crime, Cow Slaughter, Electricity Theft, Gangster and Dacoity, Property Crime, Forgery, Criminal Intimidation, Public Tranquility, Public Health, and Other.

F.3. Accuracy of Algorithm in Classification of Religion. The religion assignment algorithm is crucial since we use it to assign the religion of the defendants and the judges. We do it in two steps; first, we use the free Nilabhra algorithm to assign the religion, and then, we check all the names (along with fathers' names) manually to correct

[45]We perform text extraction on these judgment files, searching for the word "Act" at the top of the document.

the religion. Next, we describe in detail this procedure and the accuracy of the algorithm:

i) Judges' religion assignment: We manually classify Urdu sounding names[46] of all the judges (using not only their names, but also their fathers' name) as Muslim following Bhalotra et al., 2014. To test the algorithm further, we run the algorithm on the judges' names. The error rate of the Nilabhra algorithm is 1.9% (20 out of 1,150 Hindu judges are incorrectly classified as Muslims by the algorithm, and four out of 83 Muslim judges are incorrectly classified as Hindu). This gives us confidence that the algorithm is relatively sound in predicting the religion from the names.

ii) Testing on a different dataset: We test our algorithm on a completely different dataset from Bhalotra et al., 2014, where names are categorized as Muslim and Non-Muslim. Our algorithm predicts with 6% error rate (880 out of 18,118 non-Muslim names are wrongly classified as Muslims by the algorithm, and 442 out of 3,791 Muslim names are wrongly classified as non-Muslims).

iii) Defendants' religion assignment : We use the Nilabhra algorithm first to assign a religion to the defendants using their names. In case the defendants' name is neutral, we use the fathers' name to categorize the defendants as Muslim or Hindu. We test for measurement error in the defendants' religion classification using a subsample of manually digitized entries. In the next section, we explain how we arrive at the subsample from which we retrieve information manually.

The measurement error in religion assignment for the defendants is attributable to two reasons: the improper extraction of names from the pdfs and incorrect religion classification. To obtain error rate, we compare the religion of the names extracted using the algorithm with the manually assigned religion. The overall error rate in the total manually extracted sample is 6% (570 out of 31,080 Hindu names are incorrectly classified as Muslim, and 1,535 out of 4,933 Muslim names are incorrectly classified as Hindu). The reduction in the sample is because we ignore the cases that have both Hindu and Muslim defendants are present.

[46]Most of the names have clear first names, such as Mohammad and Begum

F.4. Selection of Sample for Manual Entry. We manually extract all the aforementioned variables from a subsample of cases to compute approximate error rates. We include all cases (30,000) handled by all Muslim bail judges in our analysis sample. Muslim judges comprise around 7.6% (51 judges out of 668 judges) of the total number of judges and handle 9.43% (30,727/325,944) cases. We randomly select an equal number of cases, that is, 30,000 cases, handled by Hindu judges. To arrive at our random sample of cases handled by Hindu judges, we first randomly select Hindu judges and then randomly choose cases handled by them.

In selecting Hindu judges, our objective is to retain judges similar to Muslim judges along key covariates. First, we ensure that the Hindu judges are from the same pool of home-districts as the Muslim judges. Second, we restrict the age range of the pool of Hindu judges to the same age range (30–63 years) as that of Muslim judges. We stratify the random sample by crime types of the bail cases. Within each crime category, we select a sample of Hindu judges, such that Hindu judges' median experience is the same as that of Muslim judges.

After randomly selecting the Hindu judges based on these conditions, we randomly select the cases for every judge to ensure that the total number of cases is approximately 30,000. It should be noted that the text extraction of IPC codes and Acts yielded an error rate of approximately 22%.[47] However, the measurement error for the crime category is random, and therefore, the sample of cases handled by Hindu judges that is selected for manual digitization is random.

F.5. No classical measurement error. The overall measurement error on bail outcomes is 5% (1,525 bail cases are incorrectly classified as denied out of 38,233 bail cases granted, and similarly 1,358 cases are incorrectly classified as granted out of 19,138 bail cases denied.).

In this subsection we test whether the bail outcomes and religion assignment of defendants are correlated with the exposure to riots variable. The measurement error is defined as the deviation of the text extraction values from the manual data entry is measurement. It is a dummy variable that takes value 1 when there is measurement error and is 0 otherwise. We test it through the following regression.

[47]This error rate was computed using the manually digitized data set.

$$ME_{j,d,t} = \alpha + \eta_{d,t} + \delta_{h,t} + \mathcal{F}.\mathcal{E}_b + \beta \times kid[0-6]_j + \sigma X_j + \sum_{k=7}^{9} \gamma(k) \times exposure(k)_j + \epsilon_{j,d,t} \tag{18}$$

The β coefficients as reported in the table 3.35 are close to zero and insignificant which indicates that the measurement errors are orthogonal to our main explanatory variable.