INDIAN INSTITUTE OF MANAGEMENT KOZHIKODE

Working Paper

Multigenerational Persistence and The Great Gatsby Relation for India

Chinmayi Srikanth[1]
Shubhasis Dey[2]

[1]Doctoral Student, Economics Area, Indian Institute of Management Kozhikode, IIMK Campus PO, Kunnamangalam, Kozhikode, Kerala 673 570, India; Email - srikanthc12fpm@iimk.ac.in

[2]Professor, Economics Area, Indian Institute of Management, Kozhikode, IIMK Campus PO, Kunnamangalam, Kozhikode, Kerala 673 570, India; Email: s.dey@iimk.ac.in, Phone Number (+91) 495 – 2809115

Multigenerational Persistence and The Great Gatsby Relation for India

Abstract

The purpose of this paper is to study aspects of identity that reinforce advantage and disadvantage in the educational sphere. Extant literature on multigenerational mobility in India establish the existence of a 'grandfather effect'. By using distinct models to measure intergenerational and multigenerational educational persistence, this paper contends that this effect can be direct, indirect, or non-existent depending on the identity group - caste or religion - to which the family belongs. The paper also studies *The Great Gatsby* relation for zonal councils within India using caste and religion-based identifiers. This study finds that *The Great Gatsby* relation holds for religious groups but not for caste groups, regardless of the model specification used to measure persistence. The study further explores the mechanism through which advantages and disadvantages manifest for each identity group by linking their educational persistence with the average educational levels of each generation. Additionally, segregation in educational categories is studied using a relative entropy measure.

1 Introduction

The existence of the caste system in India has led to the creation of a stratified society based on an underlying hierarchy that was primarily occupation-based. While this system of occupation-based hierarchy has been diluted by the growing complexity of economies and consequently jobs, the concept of caste and its inherent social hierarchy continues to perpetuate occupational and educational spheres.

Categorical social identities are those identities that people assume at birth. An inheritance of caste identity also means an inheritance of the disadvantages or advantages that come with it. Such inheritance are indeed a multigenerational, rather than an intergenerational phenomenon. Studying this phenomenon is particularly challenging given the paucity or even absence of data at a multigenerational level (Güell, Rodríguez Mora, & Solon, 2018), especially in the Indian context. Kundu and Sen's (2022) study is the only exception. The study finds evidence of a 'grandfather effect' for India, where the effect of the grandfather's outcome on the grandson's outcome is independent of the father's outcome, where outcome refers to education and occupation. The study, however, makes an *à priori* assumption about the model specification for multigenerational mobility. Whether the grandfather affects the grandson directly or indirectly through the father is a proposition that needs to be tested to assess the persistence of categorical affiliations over generations. The purpose of this study is to test whether the 'grandfather effect' is direct, indirect, or even non-existent for each caste and religious group. This is based on the premise that each identity group has unique cultural influences that affect the future generations in ways that cannot be generalized using a straitjacket model specification.

From a policy perspective, understanding the relationship between inequality and intergenerational persistence is crucial. *The Great Gatsby* curve establishes that countries with high inequality have high intergenerational persistence and therefore low social mobility (Corak, 2013). This study explores the same relation for each of the zonal councils in India-namely, North, North-East, Central, East, West, and South.

The study, therefore, focuses on three major aspects of categorical identity: its persistence, how differently it operates for different identity groups, and its implications for macroeconomic indicators such as inequality.

2 Review of Literature

Much of the literature on intergenerational mobility has been directed towards occupational (Corak, Lindquist, & Mazumder, 2014; Heath & Zhao, 2021; Li & Zhao, 2017; Wu & Treiman, 2007) and educational mobility (Alesina, Hohmann, Michalopoulos, & Papaioannou, 2021; Ferreira et al., 2012; Hertz et al., 2008; Narayan et al., 2018). Earnings mobility is not widely studied due to the unavailability of reliable data, especially for developing countries (Iversen, Krishna, Sen, et al., 2016).

Studies focusing on mobility have been carried out in India as well. Azam (2019) finds that intergenerational mobility follows a hierarchy that mimics one's position in the social (caste) hierarchy, with lower caste groups reflecting high persistence and lower mobility and higher caste groups showing lower persistence and therefore, higher mobility. Azam and Bhatt (2015) discover a decline in educational persistence in India, which can be attributed to public education policies.

Predicting long-term intergenerational persistence can be difficult with two generations of data alone. Multigenerational mobility is therefore more desirable where data is available. Lindahl, Palme, Massih, and Sjögren (2015) use Swedish data for four generations of educational attainment and three generations of earnings. Their study establishes that long-run intergenerational persistence is underestimated when only two generations are taken into account. As far as developing countries are concerned, studies on multigenerational mobility are limited (Solon, 2018). Kundu and Sen (2022) make use of data for three generations and is the only study for India on multigenerational persistence and establishes the existence of a 'grandfather effect' for India. Multigenerational mobility literature has established specifications that either assume a direct effect of the grandfather's outcomes on the grandson's outcomes (Solon, 2014) or an indirect effect that operates through the father's outcomes (Boserup, Kopczuk, & Kreiner, 2013; Lindahl et al., 2015). This study is motivated by the need to explore whether the 'grandfather effect' operates directly or indirectly.

Persistence of outcomes are better assessed by studying how different social groups are actually

represented in educational attainment categories. We study educational segregation using a relative entropy-based measure called the 'Divergence Index' (Roberto, 2015).

At the macroeconomic level, the relationship between inequality and persistence has been established by *The Great Gatsby Curve*, which finds countries with high income inequality to have low social mobility (high intergenerational persistence), while countries with low income inequality have high social mobility (low intergenerational persistence) (Corak, 2013). We also test whether the relationship holds for zonal councils within India.

3 Data

This study uses household-level data from the India Human Development Survey (IHDS), which is a panel survey conducted by the National Council for Applied Economic Research (NCAER) and University of Maryland. The first wave of IHDS (2004-05) comprises 41,554 households, while the second wave (2011-12) includes 42,152 households of which 83% of the households were re-interviewed from 2004-05 (Desai & Vanneman, 2010, 2015).

For the purpose of this study, we use education data from the second wave. The household questionnaire collects information on the education of the head of the household, the father of the household head while the individual questionnaire collects information on education for each member of the household and their children. This study uses data only for male-headed households and studies three generations of males since traditionally in Indian society, women once married are considered a part of their husbands' household. Using male members in three generations makes it easier to compare across households since the within-household environment are similar for all three generations.

4 Methodology

The use of distinct model specifications to ascertain the nature of persistence for each identity group, forms the crux of this study. This also serves the purpose of testing the 'grandfather effect' for India using two specifications - one that assumes a direct and independent effect of grandfather's outcome on the grandson's outcome, and one that assumes an indirect effect of the grandfather's outcome on the grandson's outcome. A third model specification does away with the existence of the 'grandfather effect' which reduces it to an intergenerational model wherein the father's education alone affects the son's education.

The three models are presented in an increasing order of complexity. First is the model with a direct effect having the following specification:

$$y_{it} = \beta_0 + \beta_1 y_{i,t-1} + \epsilon_{it}$$

$$(1)$$

where β_1 is the father's effect on the son's education.

To test the 'grandfather effect' for India, we use two regression specifications- to test the direct effect (Model II), we use the following specification:

$$y_{it} = a_1 + b_1 y_{i,t-1} + b_2 y_{i,t-2} + \epsilon_{it}$$

$$(2)$$

where b_1 represents intergenerational persistence or the persistence of the father's educational outcome on the son's educational outcome, while b_2 is the persistence of the grandfather's outcome on the son's outcome.

If the effect of the grandfather's outcome on the grandson's outcome is presumed as indirect, we use an instrument variable specification where the grandfather's education affects the grandson's education only through the father. The specification for Model III as follows:

$$y_{it} = a_2 + b_3 y_{i,t-1} + \epsilon_{it}$$

(3)

where $y_{i,t-1}$ is instrumented with $y_{i,t-2}$.

To arrive at the best model specification, we first test for endogeneity and correct for it by using grandfather's education as an instrument for father's education. In other words, Model III is chosen as the ideal model where endogeneity exists. If the father's education is an exogenous variable, a Hausman specification test is used to determine the choice between Models I and II.

To measure educational segregation, we use a measure of relative entropy based on Kullback-Leibler divergence (Solomon, 1987) called the 'Divergence Index' (Roberto, 2015).

$$D = D_0 + \sum_{j=1}^{J} \frac{T_j}{T} D_j$$

(4)

where

$$D_j = \sum_{i \in S_j} \frac{\tau_i}{T_j} \sum_{m=1}^{M} \pi_{im} log \frac{\pi_{im}}{\pi_{jm}}$$

(5)

$$D_0 = \sum_{j=1}^{J} \frac{T_j}{T} \sum_{m=1}^{M} \pi_{jm} log \frac{\pi_{jm}}{\pi_m}$$

(6)

The Divergence Index can be decomposed into *between* and *within* zone components. This index is being used to measure educational divergence - D, for India, wherein D_j represents the divergence index for each residential zone[1] and D_0 represents the divergence between zones. For educational segregation, the divergence index is calculated for each category of educational attainment[2] separately based on the actual representation of each of the groups m, with respect to their respective proportions at the state level. In this study, the subscript i represents each state within a zone, j represents each zone within the country, and m denotes the social groups - caste or religion. The term π_{im} refers to the proportion of group m in each state i for a particular educational category. The term τ_i refers to the total population in that educational category in each state, while T_j is the total number of people in a particular educational category for each zone and T represents the same for the country as a whole. This measure of segregation studies how surprising it is to observe an empirical distribution (P) given a theoretical distribution (Q).

5   Models of Persistence

Table 1, Table 2, Table 3, and Table 4 report diagnostic tests associated with each of the models to arrive at the ideal model specification. Table 1, Table 2, and Table 3 determine those categories for which persistence must be measured using an IV specification due to endogeneity of the independent variable - *father's education*. Endogeneity is measured using two tests: Wu-Hausman test and Durbin-Wu-Hausman test. Where endogeneity is detected, additional identification tests such as Anderson-Rubin and Cragg-Donald Wald Tests are used to check for whether the instrument variable - *grandfather's*

---

[1]Which is the sum of segregation in each state within the zone.

[2]The educational categorization is based on Iversen et al. (2016) and Asher, Novosad, and Rafkin (2020) respectively, the details of which are discussed in Appendix A.2.

education - is a good instrument for father's education. Where there is no endogeneity, the choice is between Model I and Model II as shown in Equation (1) and Equation (2). This choice is guided by the Hausman test results as shown in Table 4.

Table 5 and Table 6 report β coefficients of persistence associated with the model that best explains intergenerational/multigenerational education mobility for each of the caste and religious categories under each zone. This reveals that persistence is indeed a cultural phenomenon that affects each identity group differently. The use of a predetermined model specification tends to assume away such differences and, therefore, compromises the study of the underlying mechanisms that influence educational mobility.

A positive and large β coefficient signifies high persistence and low educational mobility and a negative β coefficient is associated with lack of persistence and high mobility. Whether persistence is good or bad varies from one identity group to another. A group with a higher social standing is likely to benefit from high intergenerational persistence of outcomes, since that would imply the inheritance of advantages. Conversely, a group with a low social standing will only benefit from higher mobility or lower intergenerational persistence. It is also important to consider how the three generations within each religious group have fared, on average, as far as educational outcomes are concerned, so that we are better equipped to comment on their associated persistence coefficients.

Table 1
Diagnostic Tests for North and North-Eastern Zones

| | North | | | | | North-East | | | | |
| | Endogeneity Tests | | Endogenous | Identification Tests | | Endogeneity Tests | | Endogenous | Identification Tests | |
	Wu-Hausman	Durbin-Wu-Hausman		Anderson	Cragg-Donald	Wu-Hausman	Durbin-Wu-Hausman		Anderson	Cragg-Donald
General	2.405	2.406	No			4.630	4.615	Yes	151.074	214.706
	0.121	0.121				0.032	0.032		0.000	0.000
OBC	2.891	2.891	No			0.240	0.244	No		
	0.089	0.089				0.625	0.621			
SC/ST	21.564	21.371	Yes	256.506	292.704	1.980	1.984	No		
	0.000	0.000		0.000	0.000	0.160	0.159			
Hindu	6.139	6.135	Yes	955.697	1186.426	0.059	0.059	No		
	0.013	0.013		0.000	0.000	0.809	0.808			
Muslim	2.360	2.362	No			4.375	4.351	Yes	42.763	51.134
	0.125	0.124				0.037	0.037		0.000	0.000
Christian	0.026	0.028	No			2.171	2.180	No		
	0.873	0.868				0.142	0.140			
Other	0.103	0.104	No			0.389	0.405	No		
	0.748	0.748				0.535	0.524			

Notes. The table reports the results of endogeneity tests and consequently, identification tests for instrumenting the father's education level with that of the grandfather for the Northern and North-Eastern regions. Endogeneity is tested for using the *Wu-Hausman F test* and the *Durbin-Wu-Hausman chi^2 test*. Where endogeneity has been detected, identification tests such as *Anderson-Rubin Test* and *Cragg-Donald Wald Test* are used to test the validity of the instrument (grandfather's education).

Table 2

Diagnostic Tests for Central and Eastern Zones

| | Central | | | | | East | | | | |
| | Endogeneity Tests | | Endogenous | Identification Tests | | Endogeneity Tests | | Endogenous | Identification Tests | |
	Wu-Hausman	Durbin-Wu-Hausman		Anderson	Cragg-Donald	Wu-Hausman	Durbin-Wu-Hausman		Anderson	Cragg-Donald
General	1.949	1.950	No			0.968	0.969	No		
	0.163	0.163				0.325	0.325			
OBC	14.389	14.329	Yes	432.383	512.504	0.007	0.007	No		
	0.000	0.000		0.000	0.000	0.931	0.931			
SC/ST	6.013	6.002	Yes	140.628	153.325	0.568	0.569	No		
	0.014	0.014		0.000	0.000	0.451	0.451			
Hindu	2.685	2.685	No			7.982	7.971	Yes	1083.761	1554.359
	0.101	0.101				0.005	0.005		0.000	0.000
Muslim	0.111	0.111	No			4.375	4.351	Yes	42.763	51.134
	0.740	0.739				0.039	0.038		0.000	0.000
Christian						0.008	0.009	No		
						0.927	0.925			
Other						6.976	6.675	Yes	26.381	36.82
						0.010	0.010		0.000	0.000

Notes. The table reports the results of endogeneity tests and consequently, identification tests for instrumenting the father's education level with that of the grandfather for the Central and Eastern regions. Endogeneity is tested for using the *Wu-Hausman F test* and the *Durbin-Wu-Hausman chi^2 test*. Where endogeneity has been detected, identification tests such as *Anderson-Rubin Test* and *Cragg-Donald Wald Test* are used to test the validity of the instrument (grandfather's education).

The table does not report results for *Christian* and *Other* religious groups in the Central zone since the number of observations was below 30.

Table 3
Diagnostic Tests for Western and Southern Zones

	West					South				
	Endogeneity Tests		Endogenous	Identification Tests		Endogeneity Tests		Endogenous	Identification Tests	
	Wu-Hausman	Durbin-Wu-Hausman		Anderson	Cragg-Donald	Wu-Hausman	Durbin-Wu-Hausman		Anderson	Cragg-Donald
General	0.060	0.060	No			11.616	11.504	Yes	261.896	371.132
	0.807	0.806				0.001	0.001		0.000	0.000
OBC	0.035	0.035	No			1.405	1.406	No		
	0.851	0.851				0.236	0.236			
SC/ST	0.641	0.643	No			4.404	4.400	Yes	250.288	302.479
	0.424	0.423				0.036	0.036		0.000	0.000
Hindu	1.666	1.667	No			12.618	12.591	Yes	1030.688	1339.953
	0.197	0.197				0.000	0.000		0.000	0.000
Muslim	1.666	1.674	No			3.810	3.805	No		
	0.198	0.196				0.051	0.051			
Christian	0.266	0.285	No			0.288	0.290	No		
	0.609	0.594				0.592	0.590			
Other	0.035	0.036	No			0.185	0.200	No		
	0.851	0.849				0.670	0.654			

Notes. The table reports the results of endogeneity tests and consequently, identification tests for instrumenting the father's education level with that of the grandfather for the Western and Southern regions. Endogeneity is tested for using the *Wu-Hausman F test* and the *Durbin-Wu-Hausman chi^2 test*. Where endogeneity has been detected, identification tests such as *Anderson-Rubin Test* and *Cragg-Donald Wald Test* are used to test the validity of the instrument (grandfather's education).

Table 4

Hausman Test for Caste and Religious Groups in each Zone

	North	North-East	Central	East	West	South
General	2.93 0.087 I		1.95 0.162 I	1.04 0.309 I	0.23 0.635 I	
OBC	4.21 0.040 II	0.03 0.855 I		0.16 0.691 I	0.01 0.928 I	2.23 0.135 I
SCST		3.1 0.078 I		2.46 0.117 I	0.19 0.664 I	
Hindu		0.59 0.443 I	2.89 0.089 I		2.26 0.133 I	
Muslim	1.47 0.225 I		0.11 0.740 I		1.87 0.172 I	2.39 0.122 I
Christian	0.02 0.877 I	2.67 0.102 I		0.01 0.930 I	0.4 0.527 I	0.49 0.484 I
Other	0.1 0.749 I	0.36 0.550 I			0.14 0.710 I	1.77 0.183 I

Notes. The table reports the results for the Hausman specification test and the corresponding p-values, based on which the ideal model for explaining educational persistence is chosen. The choice is made between two models- one that assumes intergenerational mobility where the father's educational outcome affects the son's outcome, and one that assumes multigenerational mobility wherein the father's and grandfather's education affects the son/ grandson's outcome independent of one another.

6   Four Emergent Scenarios

Table 7 and Table 8 report the average years of education of each of the three generations for caste and religious groups under each zone. When studied in conjunction with Table 5 and Table 6, these tables reveal the emergence of four distinct scenarios with respect to persistence, average education, and position in the social hierarchy[3].

1. **High persistence, favorable social position, and high education levels**

   General category comprises of the *Brahmins*, *Forward Castes*, and other caste categories that do not belong to any of the caste categories that benefit from affirmative action or reservation policy. Thus, they represent that part of society that are advantaged (or at least not disadvantaged) by birth. In other words, their categorical social identity is such that they are right on top of the social hierarchy. This, combined with a higher average education that generations have benefited from[4] and a high educational persistence ensure that there is a reinforcement of advantage in the educational sphere.

2. **Low persistence, unfavorable social position, and low education levels**

   The Scheduled Castes (SC) and the Scheduled Tribes (ST) are two groups that form part of the bottom-rung of the caste hierarchy. Added to their inherent disadvantage is the low average education for all three generations across zones as observed in Table 7 and Table 8. However, their low educational persistence makes this disadvantage less vigorous, pointing towards a probable disintegration of disadvantage.

3. **High persistence, unfavorable social position, and low education levels**

   Among religious groups, Muslims' disadvantage in society cannot be attributed to their minority status alone; since there are minority religious groups that do not face stigma the way Muslims do. Their stigmatized identity, low educational levels, and high intergenerational persistence, makes them particularly vulnerable, which points to a reinforcement or even an aggravation of their disadvantage.

4. **Low persistence, unfavorable social position, and high education levels**

   Among religious groups, Christians have the lowest educational persistence across zones and have reasonably high levels of average education except in the Northern zone. Christians are a minority everywhere except in the North-eastern states and Kerala. The lack of generalisability regarding the social status of Christians makes it rather difficult to categorically state the outcomes associated with low persistence and high education levels. It is however, useful to note that regardless of social position, high education combined with low educational persistence could mean that the educational achievements for each generation among Christians are the outcome of individual decisions in favor of education, rather than an inheritance of education advantage from their previous generations. This scenario is unique in identifying an identity group that has come up despite their historical disadvantage of being a minority but has consistently shown educational progress over generations.

These scenarios are general observations that hold across the zones under study. But the persistence, and the average education levels for three generations across caste and religious categories vary depending on their location. Overall, the Eastern zone has the highest educational persistence while the Western zone has the lowest educational persistence spanning across identity groups. The role of geography is less straightforward when we consider examples from our own analysis, i.e, Christians are better-off than even Hindus in the Western, Eastern, and Southern parts of the country, while they are much worse-off than even Muslims in the North. Similarly, Muslims of the South are better-off than the Muslims anywhere else in the country. Having said that, it is rather difficult to say whether advantage/ disadvantage is a feature of the land itself, its people or its history. For instance, little can be said of whether a Muslim migrant from North would benefit from the same reception as a native Muslim of the South, which is beyond the scope of this study.

---

[3]Of course, religious affiliation does not subscribe to a hierarchy, but it is interpreted in terms of minority and majority.
[4]Note that, among caste groups in Table 7 and Table 8 all three generations of General category have on average, the highest educational attainment.

Table 5

Educational Persistence among Caste and Religious Groups in the North, North-East, and Central Zones

	North				North-East				Central			
	β	Model	$R^2/AdjR^2$	N	β	Model	$R^2/AdjR^2$	N	β	Model	$R^2/AdjR^2$	N
General	0.138***	I	0.0202	2342	0.439***	III		505	0.291***	I	0.0808	1219
	(0.020)				(0.074)				(0.028)			
OBC	0.254***	II	0.059	1990	0.246**	I	0.0523	175	0.007	III		2755
	(0.023)				(0.080)				(0.047)			
SC/ST	-0.184***	III		2060	0.077	I	0.0068	500	-0.127	III		1676
	(0.064)				(0.042)				(0.083)			
Hindu	0.118***	III		4906	0.172***	I	0.0294	685	0.171***	I	0.0315	4841
	(0.031)				(0.038)				(0.014)			
Muslim	0.250***	I	0.0563	798	0.487***	III		251	0.288***	I	0.0874	777
	(0.036)				(0.137)				(0.033)			
Christian	0.265	I	0.0643	50	0.045	I	0.002	197				
	(0.146)				(0.071)							
Other	0.232***	I	0.0585	675	0.137	I	0.0201	66				
	(0.036)				(0.120)							

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Notes. The table reports the persistence associated with each caste and religious group for the North, North-East and Central zones, respectively. Persistence is calculated using three model specifications, of which the best model specification and its associated β coefficient is reported for each of the groups. Model I is an intergenerational model, Model II is a multigenerational model which assumes that the grandfather's effect of the grandson's educational outcome is independent of the father's effect, and Model III measures long-term intergenerational mobility by instrumenting the father's educational outcome with the grandfather's education, which is based on the premise that the grandfather's educational outcome affects the grandson's outcome indirectly through the father.

The table does not report results for *Christian* and *Other* religious groups in the Central zone since the number of observations was below 30.

Table 6

Educational Persistence among Caste and Religious Groups in the Eastern, Western, and Southern Zones

	East				West				South			
	β	Model	$R^2/AdjR^2$	N	β	Model	$R^2/AdjR^2$	N	β	Model	$R^2/AdjR^2$	N
General	0.408*** (0.024)	I	0.1742	1384	0.096*** (0.025)	I	0.0097	1485	0.267*** (0.064)	III		885
OBC	0.210*** (0.025)	I	0.0463	1497	0.147*** (0.027)	I	0.0204	1464	0.065*** (0.019)	I	0.0039	3081
SC/ST	0.199*** (0.025)	I	0.0404	1517	0.013 (0.032)	I	0.0002	857	0.113 (0.066)	III		1441
Hindu	0.354*** (0.028)	III		3575	0.123*** (0.017)	I	0.0162	3314	0.152*** (0.031)	III		4459
Muslim	0.486*** (0.074)	III		666	0.123* (0.057)	I	0.0162	280	0.085 (0.043)	I	0.0063	605
Christian	0.070 (0.136)	I	0.0053	51	-0.007 (0.157)	I	0	44	0.100 (0.077)	I	0.0058	289
Other	0.543** (0.177)	III		88	-0.016 (0.082)	I	0.0002	166	0.248 (0.146)	I	0.074	38

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Notes. The table reports the persistence associated with each caste and religious group for the Eastern, Western, and Southern zones, respectively. Persistence is calculated using three model specifications, of which the best model specification and its associated β coefficient is reported for each of the groups. Model I is an intergenerational model, Model II is a multigenerational model which assumes that the grandfather's effect of the grandson's educational outcome is independent of the father's effect, and Model III measures long-term intergenerational mobility by instrumenting the father's educational outcome with the grandfather's education, which is based on the premise that the grandfather's educational outcome affects the grandson's outcome indirectly through the father.

Table 7

Average Education among Caste and Religious Groups in the Northern, North-Eastern, and Central Zones

	North			North-East			Central		
	Son	Father	Grandfather	Son	Father	Grandfather	Son	Father	Grandfather
General	8.98	8.08	3.51	7.23	7.10	3.57	7.58	8.07	4.14
OBC	7.00	5.94	1.73	8.75	8.97	4.03	6.07	5.34	1.82
SC/ST	6.32	5.04	1.37	7.25	7.02	2.76	5.30	4.13	1.11
Hindu	7.65	6.71	2.33	8.38	8.32	3.97	6.41	5.77	2.15
Muslim	6.61	5.18	2.15	5.20	4.46	1.46	4.66	4.04	1.74
Christian	5.63	4.12	1.63	7.25	7.64	3.41	6.18	12.06	7.65
Other	7.81	6.27	2.24	7.09	6.50	2.16	9.76	7.47	3.40

Notes. The table shows the average education of each generation (son, father, and grandfather) for each caste and religious group in the Northern, North-Eastern, and Central zones, respectively.

Table 8

Average Education among Caste and Religious Groups in the Eastern, Western, and Southern Zones

	Eastern			Western			Southern		
	Son	Father	Grandfather	Son	Father	Grandfather	Son	Father	Grandfather
General	7.78	7.59	4.68	8.82	8.12	3.66	8.55	7.85	3.80
OBC	6.46	5.57	2.52	7.63	6.20	2.49	7.92	6.26	2.65
SC/ST	5.62	4.06	1.46	7.24	5.49	1.88	7.14	4.19	1.32
Hindu	6.85	5.99	2.99	8.09	6.88	2.85	7.81	5.88	2.38
Muslim	5.08	4.26	2.31	6.37	5.38	1.80	7.40	5.78	2.42
Christian	7.17	7.50	2.75	9.14	7.86	5.98	8.90	8.68	4.97
Other	6.61	4.02	1.68	8.57	6.77	2.55	7.27	6.16	1.89

Notes. The table shows the average education of each generation (son, father, and grandfather) for each caste and religious group in the Eastern, Western, and Southern zones, respectively.

7 The Great Gatsby Relation

The Great Gatsby curve was first used by Corak (2013) to show that countries with high income inequality had high generational earnings elasticity or persistence, while those with low income inequality had low persistence. *The Great Gatsby* relation establishes a direct relation between inequality and generational persistence. By being the first to test this relation for areas within a country, this study aims at exploring if *The Great Gatsby* relation exists for zones within India.

Figure 1 - Figure 8 analyze the relationship between education inequality in the father's generation and persistence for each of the caste/ religious groups in each zone to determine whether the Great Gatsby relation holds for identity-zone cells in India. We also test if this relation holds regardless of the model specification employed to measure persistence.

It may be observed from Figure 1, Figure 2, Figure 3, and Figure 4 that when the identity-zone cell considered is based on caste identity, *The Great Gatsby* curve is negatively sloped. In other words, caste-zone cells with high educational inequality (measured by the Gini coefficient) have low generational educational persistence (β coefficients) and therefore, high generational educational mobility. This negative relation holds regardless of how the β coefficients are measured. Another noteworthy pattern is that of how General category corresponds to high persistence and relatively low educational inequality. This means that an individual who belongs to the General category is likely to have high educational attainment, that is most likely because his father is also highly educated; and since educational inequality is low among General category individuals, it can be inferred that most individuals with this caste affiliation are highly educated. There is a cluster of SCs and STs on that part of the graphs that correspond to high inequality and low generational persistence, which means that within this group, there may be a few individuals with high education and high mobility ensures that there is ample scope for breaking out of a predominantly low education trap.
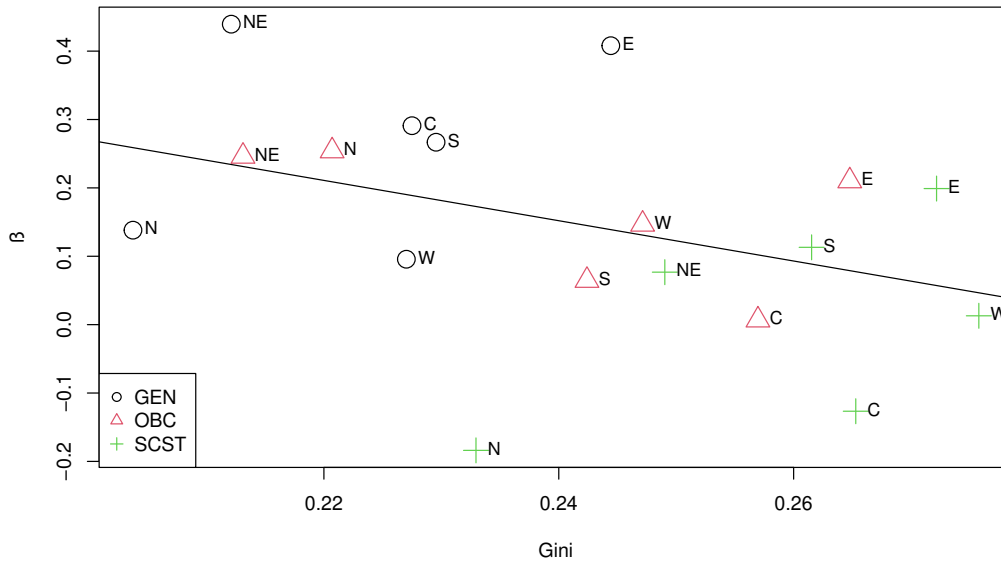

Figure 1

Great Gatsby Curve for Caste Groups based on the Best Model Specification

Notes. The figure plots the Gini coefficient for education for each caste group in each zone on the x-axis and the corresponding β coefficient using the best model specification on the y-axis.
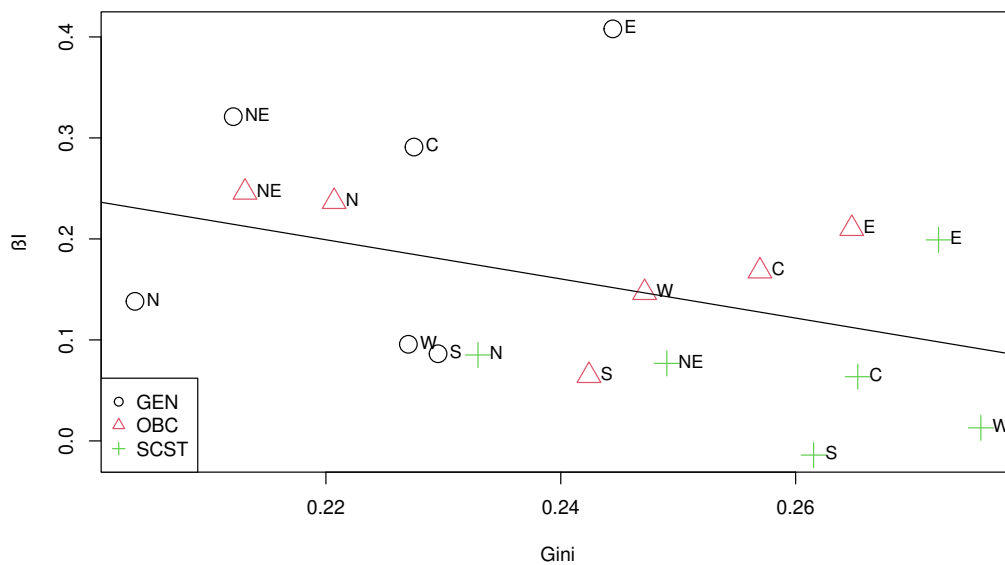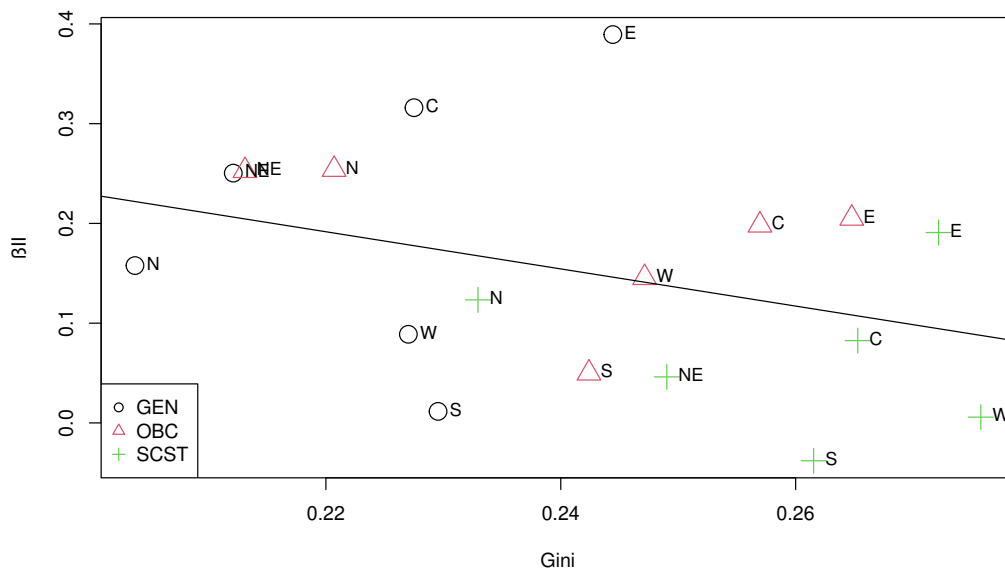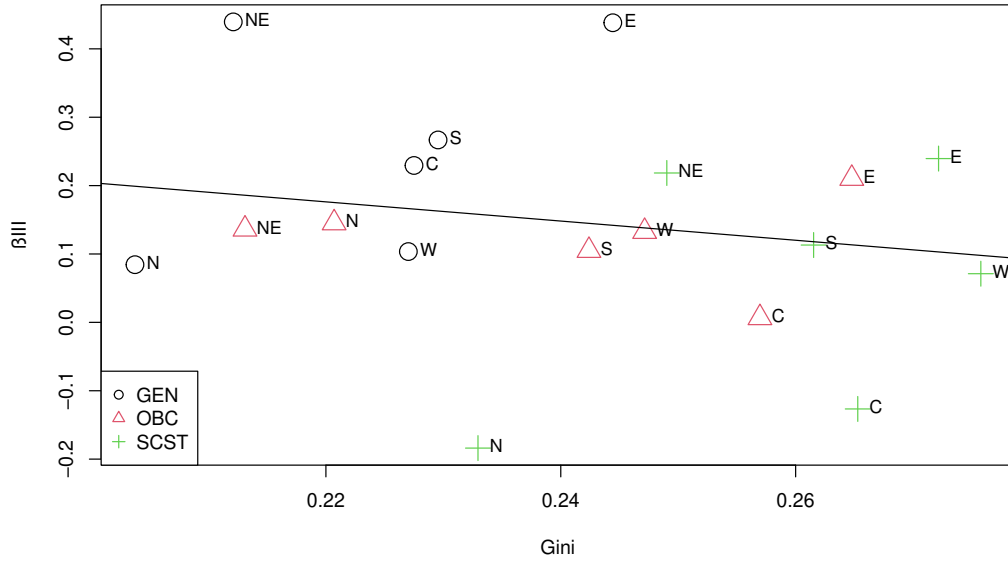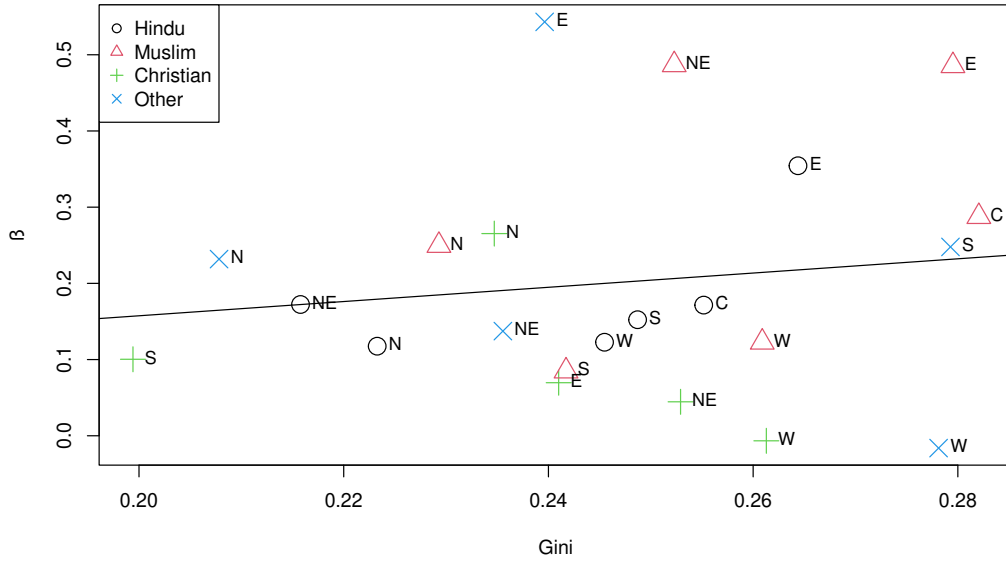
Figure 2

Great Gatsby Curve for Caste Groups based on Model I

Notes. The figure plots the Gini coefficient for education for each caste group in each zone on the x-axis and the corresponding β coefficient using Model I i.e, Equation (1) on the y-axis.

Figure 3

Great Gatsby Curve for Caste Groups based on Model II

Notes. The figure plots the Gini coefficient for education for each caste group in each zone on the x-axis and the corresponding β coefficient using Model II i.e, Equation (2) on the y-axis.

Figure 4

Great Gatsby Curve for Caste Groups based on Model III

Notes. The figure plots the Gini coefficient for education for each caste group in each zone on the x-axis and the corresponding β coefficient using Model III i.e, Equation (3) on the y-axis.

When it comes to studying the same relation for religion-zone cells, it is surprising to observe that the *The Great Gatsby* relation holds. Although the relation does not seem to hold true when the generational persistence is measured using Model II which corresponds to Equation (2), all other estimates of β seem to be consistent with the positive relationship between persistence and inequality. Those religion-zone cells with high educational inequality have high generational education persistence. Figure 5, Figure 6, and Figure 8 show a positive relationship while Figure 7 depicts no discernible relationship. Unlike the caste-zone curves, we do not observe clustering of religious groups in certain parts of the graphs.

Figure 5

Great Gatsby Curve for Religious Groups based on the Best Model Specification

Notes. The figure plots the Gini coefficient for education for each religious group in each zone on the x-axis and the corresponding β coefficient using the best model specification on the y-axis.
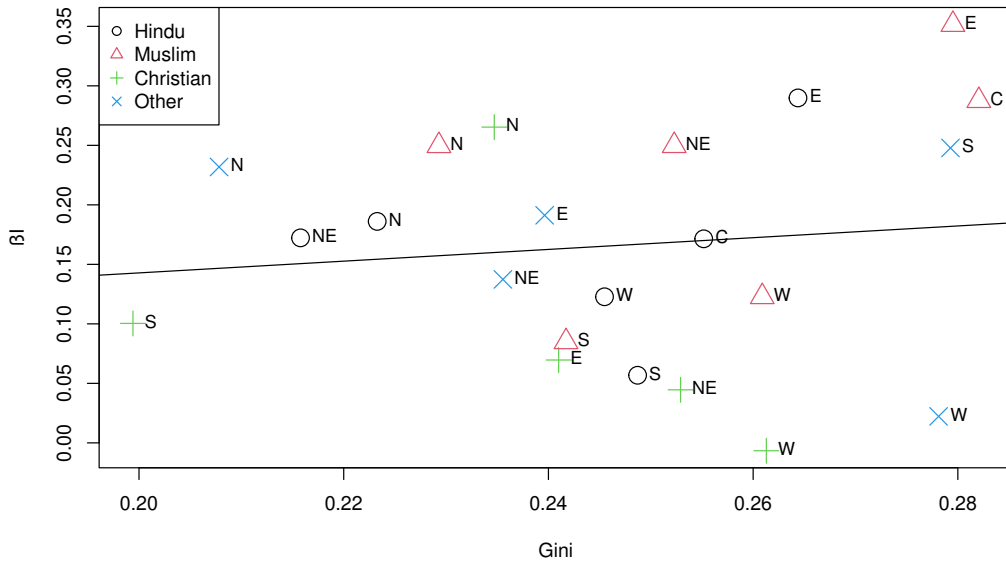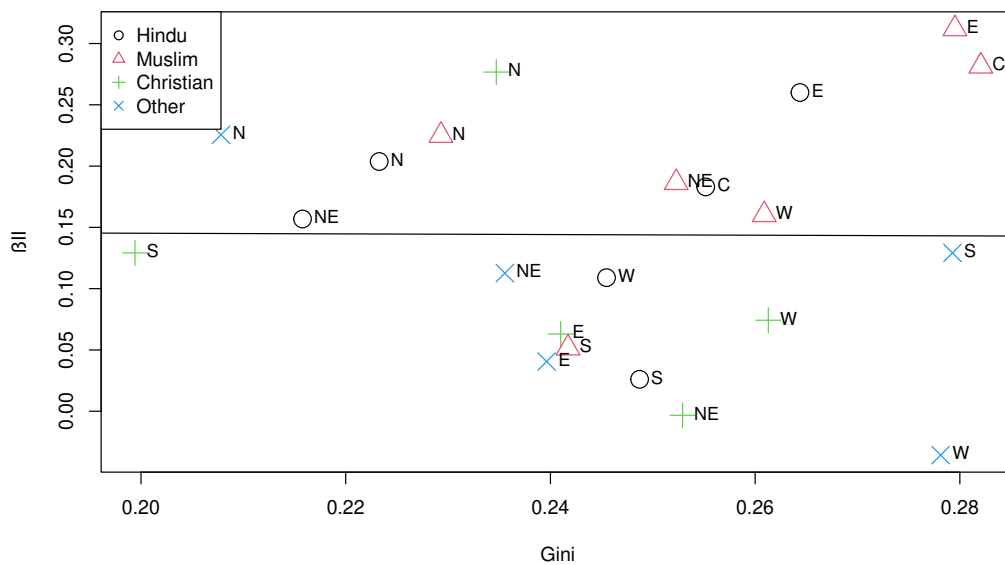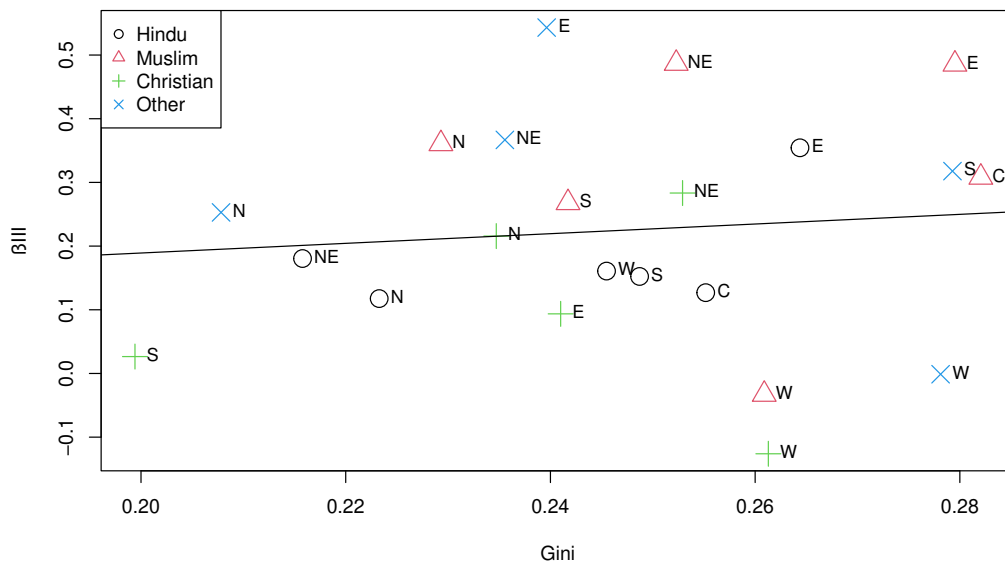

Figure 6

Great Gatsby Curve for Religious Groups based on Model I

Notes. The figure plots the Gini coefficient for education for each religious group in each zone on the x-axis and the corresponding β coefficient using Model I i.e, Equation (1) on the y-axis.

Figure 7

Great Gatsby Curve for Religious Groups based on Model II

Notes. The figure plots the Gini coefficient for education for each religious group in each zone on the x-axis and the corresponding β coefficient using Model I i.e, Equation (2) on the y-axis.

Figure 8

Great Gatsby Curve for Religious Groups based on Model III

Notes. The figure plots the Gini coefficient for education for each religious group in each zone on the x-axis and the corresponding β coefficient using Model I i.e, Equation (3) on the y-axis.

8 Educational Segregation

The purpose of studying segregation is to observe the concentration of certain groups in the various educational categories. This segregation is often reflective of the representation of certain groups at these education levels. The use of a relative entropy measure such as the 'Divergence index' conceptualizes segregation as a measure of surprise that takes into account the population composition of various caste and religious groups vis-a-vis their actual representation at the zone and state levels.
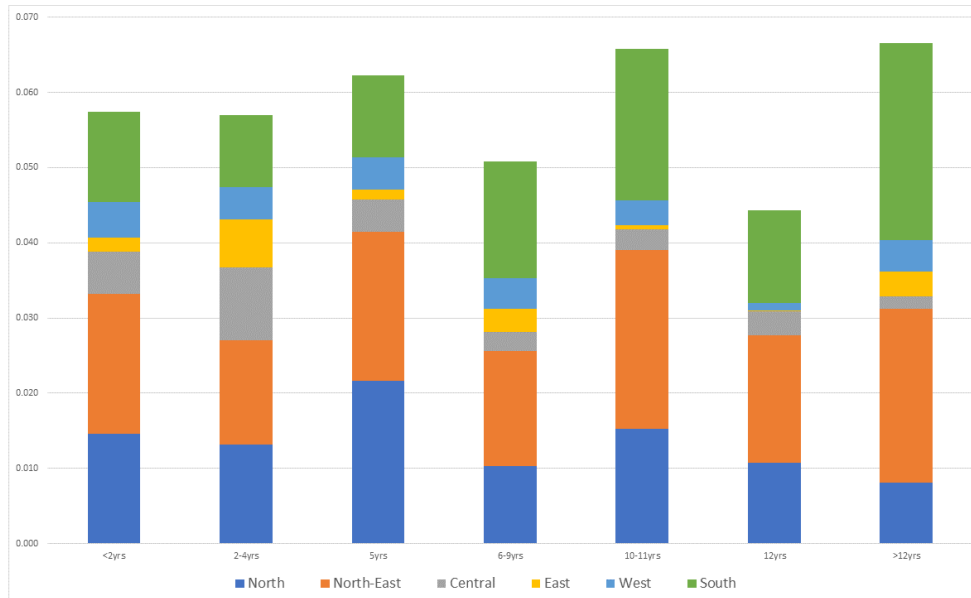
Figure 9

Between-Zone Caste Segregation

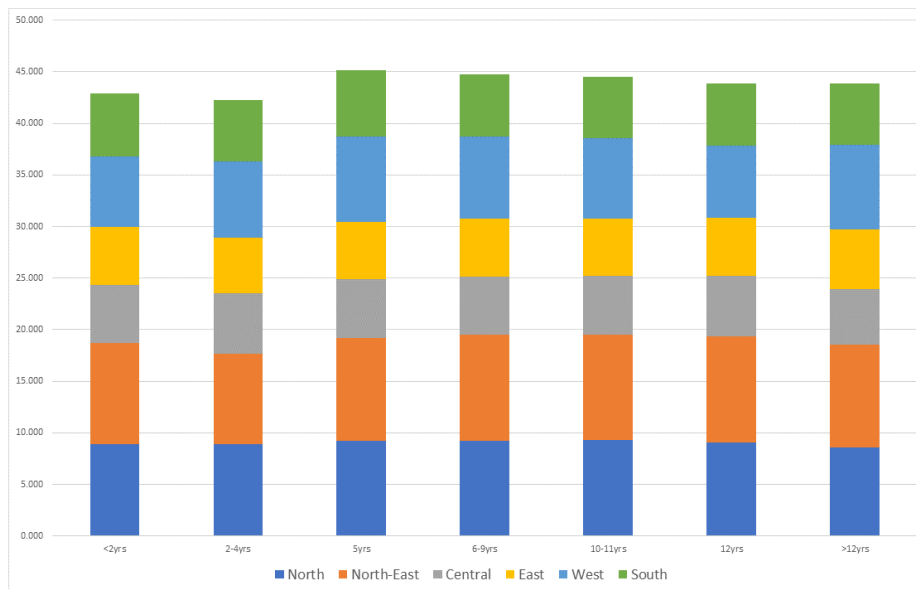*Notes.* The figure plots the between-zone caste segregation for each educational category.

Figure 10

Within-Zone Caste Segregation

Notes. The figure plots the within-zone caste segregation for each educational category.

Figure 9 and Figure 10 represent between and within-zone segregation based on caste categories measured separately for each domain of educational attainment. Overall, the between-zone segregation is highest in the post-secondary educational category, where the South and North-East zones display the largest deviations from the population proportions. The Divergence Index is consistently high in the North-East zone, while the East zone records the lowest deviations from population proportions as far as caste segregation is concerned.
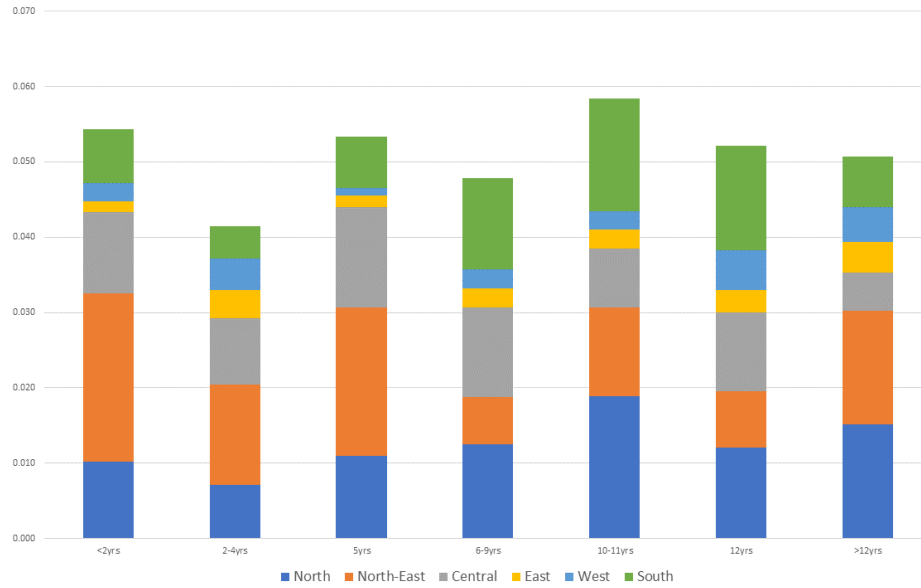

Figure 11

Between-Zone Religious Segregation

Notes. The figure plots the between-zone religious segregation for each educational category.
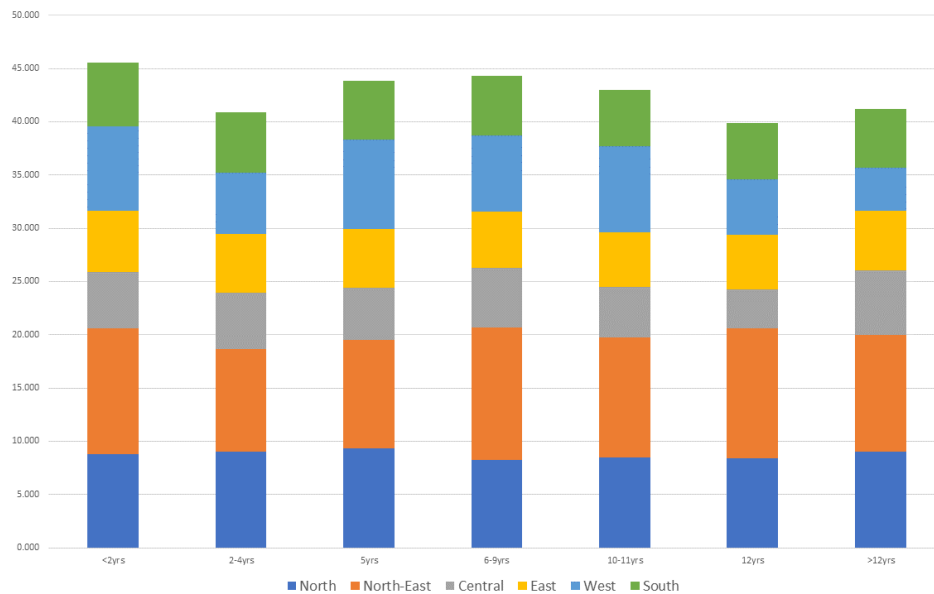

Figure 12

Within-Zone Religious Segregation

Notes. The figure plots the within-zone religious segregation for each educational category.

Figure 11 and Figure 12 show that between-zone and within-zone religious segregation is consistently

high in the North-East zone. It can also be observed that unlike in caste-segregation, the Central zone has relatively high divergence with respect to between-zone religious segregation. The between-zone religious segregation is highest in the secondary education category, with a major proportion of it being dominated by the North-East and South Zones.

9   Discussion & Conclusion

While this study analyzes caste and religious groups separately, they represent aspects of identity that do not exist in isolation from one another. In fact, the two identity groups overlap for all individuals, thereby creating an intersectionality of identities that reinforce advantages or disadvantages depending on their position in the caste hierarchy and the status of their religious affiliation (majority or minority). The study does not take for granted the fact that socially advantaged groups have better educational outcomes than those groups with an inherent disadvantage imposed by their identity. It seeks to further develop this notion in terms of how far these disadvantages or advantages manifest in the future generations, what is the average educational outcomes for each of the generations under study and what it means for their future progress (mobility) when the group has relatively high or low educational inequality. Another aspect that this study accounts for throughout, is the role of one's location in addition to their identity affiliations. It is apparent that the region in which an identity group is located also matters when it comes to not just their educational outcomes but also that of its persistence across generations.

The purpose of this study is manifold. On one hand, it seeks to study how generational persistence operates for each identity group. We find evidence for differential persistence mechanism for different identity groups in specific zones. For some groups there exist no grandfather effect and therefore persistence is a phenomenon that lasts two generations, for others, the grandfather's educational outcome affects the grandson's outcome independently while for some other groups, this effect operates indirectly through the father. On the other hand, this study ascertains the relationship between persistence and average education for each generation and analyzes it by considering the position of the identity group in the social hierarchy. In this manner, the study identifies four emergent scenarios and what it implies for a reinforcement of advantage or disadvantage. Another aspect of generational persistence this study covers is the relationship between persistence and inequality of education through *The Great Gatsby* curve, and how the relationship changes when the identity groups in question are altered. We find evidence for a relation that is in line with *The Great Gatsby* when religious identity is considered. The same relationship flips when caste groups are taken into account.

References

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2021). Intergenerational mobility in africa. *Econometrica*, *89*(1), 1–35.

Asher, S., Novosad, P., & Rafkin, C. (2020). Intergenerational mobility in india: New methods and estimates across time, space, and communities. *Unpublished manuscript. Available at: https://paulnovosad. com/pdf/anr-india-mobility. pdf*.

Azam, M. (2019). Household income mobility in india, 1993-2011. *Available at SSRN 2815008*.

Azam, M., & Bhatt, V. (2015). Like father, like son? intergenerational educational mobility in india. *Demography*, *52*(6), 1929–1959.

Boserup, S. H., Kopczuk, W., & Kreiner, C. T. (2013). Intergenerational wealth mobility: Evidence from danish wealth records of three generations.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, *27*(3), 79–102.

Corak, M., Lindquist, M. J., & Mazumder, B. (2014). A comparison of upward and downward intergenerational mobility in canada, sweden and the united states. *Labour Economics*, *30*, 185–200.

Desai, S., & Vanneman, R. (2010). National council of applied economic research, new delhi. india human development survey (ihds), 2005. inter-university consortium for political and social research [distributor]. *Ann Arbor, MI*, 06–29.

Desai, S., & Vanneman, R. (2015). *India human development survey-ii (ihds-ii), 2011-12*. Inter-university Consortium for Political and Social Research Ann Arbor, MI.

Ferreira, F. H., Messina, J., Rigolini, J., López-Calva, L.-F., Lugo, M. A., Vakis, R., . . . others (2012). *Economic mobility and the rise of the latin american middle class*. World Bank Publications.

Güell, M., Rodríguez Mora, J. V., & Solon, G. (2018). New directions in measuring intergenerational mobility: Introduction. *The Economic Journal*, *128*(612), F335–F339.

Heath, A., & Zhao, Y. (2021). Rethinking occupational mobility in developing countries. *Social Mobility in Developing Countries*.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2008). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy*, *7*(2).

Iversen, V., Krishna, A., Sen, K., et al. (2016). *Rags to riches? intergenerational occupational mobility in india* (Tech. Rep.). GDI, The University of Manchester.

Kundu, A., & Sen, K. (2022). Multigenerational mobility among males in india. *Review of Income and Wealth*.

Li, Y., & Zhao, Y. (2017). Double disadvantages: A study of ethnic and hukou effects on class mobility in china (1996–2014). *Social Inclusion*, *5*(1), 5–19.

Lindahl, M., Palme, M., Massih, S. S., & Sjögren, A. (2015). Long-term intergenerational persistence of human capital an empirical analysis of four generations. *Journal of Human Resources*, *50*(1), 1–33.

Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D. G., . . . Thewissen, S. (2018). *Fair progress?: Economic mobility across generations around the world*. World Bank Publications.

Roberto, E. (2015). The divergence index: A decomposable measure of segregation and inequality. *arXiv preprint arXiv:1508.01167*.

Solomon, K. (1987). Letter to the editor: The kullback-leibler distance. *The American Statistician*, *41*(4), 338–341.

Solon, G. (2014). Theoretical models of inequality transmission across multiple generations. *Research in Social Stratification and Mobility*, *35*, 13–18.

Solon, G. (2018). What do we know so far about multigenerational mobility? *The Economic Journal*, *128*(612), F340–F352.

Wu, X., & Treiman, D. J. (2007). Inequality and equality under chinese socialism: The hukou system and intergenerational occupational mobility. *American Journal of sociology*, *113*(2), 415–445.

A Appendix

A.1 Zonal Councils

S.No	Zone	States
1	North	Chandigarh, Delhi, Haryana, Himachal Pradesh, Jammu & Kashmir, Punjab, Rajasthan
2	North-East	Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, Sikkim
3	Central	Chhattisgarh, Madhya Pradesh, Uttarakhand, Uttar Pradesh
4	Eastern	Bihar, Jharkhand, Odisha, West Bengal
5	Western	Dadra & Nagar Haveli, Daman & Diu, Goa, Gujarat, Maharashtra
6	Southern	Andhra Pradesh, Karnataka, Kerala, Puducherry, Tamil Nadu, Andaman & Nicobar, Lakshadweep

Notes. The table includes only those states and union territories that are part of the IHDS data. Two union territories, *Andaman & Nicobar*, and *Lakshadweep* were excluded from the analysis.

A.2 Educational Categories

Categories	Description	Years of education completed
1	Illiterate with less than primary	<2
2	Literate with less than primary	2-4
3	Primary	5
4	Middle	6-9
5	Secondary	10-11
6	Higher secondary	12
7	Post-secondary	>12

Notes. Based on Asher et al. (2020).

Research Office

Indian Institute of Management Kozhikode

IIMK Campus P. O.,

Kozhikode, Kerala, India,

PIN - 673 570

Phone: +91-495-2809237/ 238

Email: research@iimk.ac.in

Web: https://iimk.ac.in/faculty/publicationmenu.php