HUMAN CAPITAL AND GROWTH IN INDIA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GOA BUSINESS SCHOOL

GOA UNIVERSITY

BY

Hussain Yasser Razak

GOA UNIVERSITY

GOA

JANUARY 2024

HUMAN CAPITAL AND GROWTH IN INDIA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GOA BUSINESS SCHOOL

GOA UNIVERSITY

BY

Hussain Yasser Razak

GOA UNIVERSITY

GOA

JANUARY 2024

DECLARATION

I, Hussain Yasser Razak, hereby declare that this thesis represents work which has been carried out by me and that it has not been submitted, either in part or full, to any other University or Institution for the award of any research degree.

Place: Taleigao Plateau
Date: 23-01-2024

Hussain Yasser Razak
Research Scholar
Goa Business School
Goa University

CERTIFICATE

I hereby certify that the above Declaration of the candidate, Hussain Yasser Razak is true and the work was carried out under my/our supervision.

Professor Pranab Mukhopadhyay, Ph.D
Research Guide
Goa Business School, Goa University

ACKNOWLEDGEMENT

All praise and glory be to The Almighty - 'Most Gracious , Most Merciful' for help and guidance in undertaking and completing this thesis. The completion of this thesis was possible thanks to the guidance, motivation, and patience of many individuals, and I would like to express my deepest gratitude to them.

I would like to extend my utmost gratitude to Professor Pranab Mukhopadhyay, my esteemed guide and mentor. His exceptional expertise, valuable time, and insightful guidance have been instrumental in my growth over the years. His unwavering motivation and boundless patience have kept me on the right track, and I am sincerely grateful to him for broadening my perspective and outlook on research as a whole.

I would like to extend my sincerest gratitude to the esteemed members of the monitoring committee, including Professor P.K. Sudarsan, Professor Rahul Tripathi and Professor Purva G. Hegde-Desai, for their most valuable input, constructive feedback, and timely guidance throughout the completion of this thesis.

I am grateful to the faculty and administrative staff at Goa Business School, and Goa University for their invaluable assistance and support during my research tenure.

I am grateful to the other research scholars at Goa Business School for their constant motivation, suggestions and the insightful discussions we had over the years. Their unwavering support and research expertise have been immensely helpful in my academic journey.

I am deeply grateful to the Inter-University Consortium of Political and Social Research (ICPSR), a branch of the Institute of Social Research (ISR) located at the University of Michigan. Their Data sharing for Demographic Research (DSDR) initiative has made vast datasets accessible to the public, for which I am thankful.

I would like to express my gratitude to the Management, Principal, Former Principals, faculty, staff, and students of M.E.S. Vasant Joshi College of Arts and Commerce, Zuarinagar for their valuable support and encouragement throughout the course of my thesis.

I take this opportunity to also thank my mentors at different points of time in my academic journey. I thank Dr. Meenakshi Bawa and Mr. Avinash Periera for their guidance and motivation in my academic pursuit.

To my dearest friends, near and far, who have always stood by me, I express my boundless gratitude for their unwavering support time and again.

Throughout the demanding process of completing this thesis, the unwavering patience, encouragement, and prayers of my grandmother, parents, sisters, uncles and aunts, cousins, niece and nephews were truly invaluable. I am immensely grateful for their support, which played a crucial role in enabling me to reach this important milestone.

Without the collective support of all these individuals, I would not have been able to complete this thesis in its current form, and for that, I am eternally grateful.

Hussain Yasser Razak
Research Scholar
Economics Discipline
Goa Business School, Goa University
January, 2024

DEDICATIONS

'Oh Lord! Increase Me in Knowledge'

This Thesis is a Dedication to

My Grandfather : Mr. Yusuf Hussain

My Grandmother: Ms. Bi Fathima

My Father: Abdul Razak

My Mother: Fathima Hussain Razak

HUMAN CAPITAL AND GROWTH IN INDIA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF CHARTS

CHAPTER 1: INTRODUCTION TO HUMAN CAPITAL AND GROWTH

1.1. INTRODUCTION

A nation's human capital can be viewed as a driving force for sustained economic growth and economic development. This implies a more significant investment in Education and skills and uses a targeted policy intervention to influence the growth of human capital and, thereby, economic growth and development. Quality education is a priority sustainable development goal (Angrist et al., 2021).

Human capital theory has been widely studied using either a microeconomic (Mincer, 1958; Mincer, 1974; Schultz, 1961) or macroeconomic (Barro, 1999; Romer, 1989) framework. In the microeconomics context, we find human capital theory finding its prominence from the late 50s and early 60s in Schultz and Mincer's contributions (Goldin, 2016; Psacharopoulos & Patrinos, 2018). The earliest mention of the term Human Capital is attributed to (Schultz, 1961). Other influential contributions (Becker, 1962; Mincer, 1994;1974) paved the way for human capital studies. These studies looked at investments in human capital in the form of education, health, on-the-job training, and Migration for jobs. Of these, one of the largest forms of investment in human capital is conceived to be education (Goldin, 2016). Education, measured as educational attainment or years of education/schooling, was among the initial measures of quantification of an individual's investment in herself/himself (Griliches, 1977; Mincer, 1974) and continues to be used in empirical studies (Agrawal, 2012; Blundell et al., 2001; Lemieux, 2006). The role of human capital was explored by (Lucas, 1988) and (Gregory et al., 1992) in the original Solow model. (Barro, 1999), who looked at differences in average education and labor skills to explain differences in growth rates across nations.

Approaches to human capital measurement can broadly be classified as indicator approach, cost approach, and income approach (Abraham & Mallatt, 2022). The most influential of these studies use the indicator approach (average years of schooling/ educational attainment) (Barro & Lee, 1993; Denison, 1962).

These seminal contributions of Schultz (microeconomic framework) (Schultz, 1961) and Denison (macroeconomic framework) (Denison, 1962) in the early 1960s triggered further research on both fronts (Sobel, 1978).

Consequent to Mincer (1958) and Schultz (1961), a large pool of literature emerged on human capital theory, testing the validity of the model specification across the world (Aakvik et al., 2010; Agrawal, 2011; Alba-Ramirez & San Segundo, 1995; Asadullah & Xiao, 2020; Byron & Manaloto, 1990; Card & Krueger, 1992a; Kikuchi, 2017; Leigh, 2008a). Research in this area has evolved with alternative frameworks and methodologies, contrasting findings across nations, and better estimation techniques. In the sphere of empirical estimations, the work of Heckman (1976, 1979) occupies a unique space as he pointed to the potential self-selection bias in earlier results.

We incorporate this methodology in our study. We restrict our study to education as a specific form of human capital investment. In doing this, we incorporate an analysis of returns to the most significant form of human capital investment at the individual level.

1.2. STATEMENT OF THE PROBLEM

The primary objective of measuring human capital investment lies in ascertaining the impact of these investments (such as on skills and abilities) to explain differences in gains (economic and social) for individuals across nations over time (Abraham & Mallatt, 2022; Schündeln & Playforth, 2014). Quality education has been a priority sustainable development goal; however, its progress toward global attainment raises questions (Angrist et al., 2021).

Private returns to Education have been extensively used in human capital literature (Bhattacharya & Sato, 2017; Bratsberg & Terrell, 2002; Card & Krueger, 1992; Mendiratta & Gupt, 2013). Estimations of these returns provide insight into an individual's decision-making to pursue additional years of education instead of entering the job market, bearing direct and indirect costs of such investment decisions (Harmon et al., 2003; Solmon, 1970). However, these estimated returns pose two significant questions in their use for policy recommendations. Firstly, researchers have questioned the methodology used to estimate these returns in terms of the structural specifications of the model used (Angrist & Krueger, 1992; Briggs, 2004; Card, 1999; Heckman, 1990; Heckman, 1976).

The other aspect widely discussed is the range of the estimates themselves. Typically, the estimates of private returns would vary across nations and regions, owing to several factors, including the quality of education and the incentive to spend more time in the system (Patrinos & Psacharopoulos, 2020; Psacharopoulos & Patrinos, 2018).

With this in hindsight, we first ascertain the suitable methodology and framework to estimate private returns to education and test the model for specification issues. However, studies have shown that heterogeneity in returns to education is observable across nations (Aakvik et al., 2010; Coady & Dizioli, 2018; Hu & Hibel, 2015; Mitra, 2019; Tsai & Xie, 2011). Estimates of private returns in earlier studies vary significantly across gender, ethnic groups, and other social hierarchies (Arabsheibani et al., 2018; Bhaumik & Chakrabarty, 2006, 2008).

In the Indian context, social constructs based on religion and caste have been known to impact earnings (Dhesi & Singh, 1989; Rani, 2013). Gender wage differentials are a well-studied aspect of the Indian labor markets (Duraisamy, 2002; Mendiratta & Gupt, 2013). It is anticipated that disaggregated estimates of private returns to education will provide insights into pertinent questions regarding the Indian labor market. For instance, will a girl in India secure a job by attending school? Do women receive comparable returns on education as men? To what extent do caste and religion influence wage determination? These questions remain at the forefront of household decision-making and are central to the ongoing debate on development policy in countries such as India (Hussain & Mukhopadhyay, 2023). Similar concerns are voiced in various regions globally, including developed nations, where discrimination has been observed based on factors such as economic status, caste, race, gender, and religion. This thesis presents a structured methodology to scrutinize the study of such issues and eventually provide estimates to address the prevalent questions in the Indian context.

Since we establish the crucial role of education in determining earnings, we then ask about the impact of natural shocks on an individual's educational outcomes. This part of the thesis explains the magnitude and direction of the causal relationship between educational outcomes and natural disasters. Given the onset of climate change with the possibility of recurring natural disasters and recent commitments to targets for sustainable goals, it would be interesting to study this relationship. Studies in this regard have focussed mainly on specific regions or types of natural disasters. This thesis contributes by providing estimates using a nationally representative data set at the village level. We estimate the impact of natural disaster intensity on educational attainment in India.

1.3. RESEARCH QUESTIONS

Our objectives in this thesis are to cover two interrelated aspects of human capital investment in education. In doing so, we aim to address several research questions pertaining to private returns to education in India and the impact of natural disasters on educational outcomes in rural India.

Firstly, we seek to explore the extent to which an additional year of education translates to private returns in India and how these returns vary across different economic distributions, castes, religions, and genders. To achieve this, we will delve into various research inquiries, such as how much impact does an extra year of education have on private returns and how these returns differ among diverse economic distributions, castes, religions, and genders within the country.

In the Indian context, we ask questions like: What impact do private human capital investments in education have on individual earnings? How much do these private returns vary across economic distributions? Do these private returns to education differ significantly for women as compared to men? Do individuals from marginalized castes realize similar private returns to an additional year of education as non-marginalized? How much do these private returns vary across religions?

Secondly, we also intend to investigate the impact of natural disasters on educational outcomes in India. We seek to answer questions like How much do natural disasters impact educational outcomes in India? How extreme are the consequences for vulnerable sections of society?

1.4. RESEARCH GAP

Education is widely considered an investment in human capital, which provides a pathway to overcome such barriers. Accordingly, the human capital framework has been used to explain the earnings of individuals, treating it as a return on investment

(Becker, 1962; Mincer, 1958; Schultz, 1961). Extensive empirical literature has emerged based on this framework (Acemoglu & Autor, 2012; Becker, 1994). However, conventional wage determination models using ordinary least squares (OLS) methods may suffer from inconsistency and bias – selection and endogeneity (Heckman et al., 2006; Wooldridge, 2018). Heckman (1974) provided a mechanism to overcome the problem of selection bias in Mincer's original formulation (Mincer, 1974). Apart from cross-country studies, the Heckman selection model has been used to estimate returns on education in different parts of the world, like the USA, Europe, and China.

Literature on returns to additional years of education in the Indian context has used an extended form of Mincer's earnings function, highlighting the nature and magnitude of returns to education and experience (Bhaumik & Chakrabarty, 2008; Dutta, 2006; Jacob, 2018). Most of these studies have relied on data from the National Sample Survey Organisation (NSSO) (Mohanty, 2021) or the India Human Development Surveys (IHDS) (Azam et al., 2013). While the NSSO datasets are rich in information, they pose one problem for the researcher – each round interviews different households and, therefore, is not amenable to panel data analysis directly. Almost all earlier studies are based on cross-sectional analysis, which has numerous limitations as they fail to observe changes in an individual over time (Wooldridge, 2018).

A few studies have created a regional pseudo-panel dataset using district identifiers as proposed by Deaton (1985) and thus overcome the problems posed by estimates of cross-section data. Bhattacharya and Sato (2017) used this technique to analyze the effect of socioeconomic factors on the real wage rate of male workers in India. They found that caste and religion were not significant determinants of wages. This was contrary to the understanding of the Indian labor market, where gender, caste, and religion were found to influence wages (Agrawal, 2014; Sengupta and Das, 2014).

Many factors could determine the years a child spends in school. Some of these are individual ability-related (Siddhu, 2011), school-related (including the number of teachers, medium of instruction, and the number of toilets) (Chatterjee et al., 2018), or family-related (Choudhury, 2006).

1.5. OBJECTIVES

1. To estimate the nature and impact of human capital investment (education) on earnings in India. (Chapter 4)

2. To study trends and differences in returns to human capital investment across class (Chapter 5)

3. To study trends and differences in returns to human capital investment across gender (Chapter 6)

4. To study trends and differences in returns to human capital investment across castes (Chapter 7)

5. To study trends and differences in returns to human capital investment across religion (Chapter 8)

6. To estimate the impact of natural disasters (with differential intensity) on educational outcomes in rural villages (Chapter 9).

1.6. CHAPTER SUMMARY

Chapter 1: Introduction

This chapter introduces the concept of investment in Human Capital and its relationship to Economic growth and development. The chapter discusses the origin of human capital in economic studies. The constituents/ forms of human capital investment are discussed to give the reader a sense of what comprises human capital investment. A broad outline is also provided about the scope of the subject in the macroeconomic and microeconomic domains. The introduction also highlights the pioneering work in these respective fields. We then approach the idea of returns to investment in human capital. The subject of estimating and interpreting returns to human capital investment in the form of earnings has been of interest to researchers for well over seven decades. Apart from providing a conceptual background to the subject of this thesis, the chapter also highlights the problem statement, the gap in empirical research, and such frames the study's objectives. The chapter concludes with a chapter summary.

Chapter 2: Literature Review

The literature review chapter in the thesis is divided into two parts. Part a) Deals with the extensive literature and studies on investment in human capital and the estimation of private returns to Education as a form of investment in Human capital. This sub-section of the chapter is a humble attempt to organize International and National literature on returns to human capital investment. Education is considered an investment in human capital, which provides a pathway to overcome socioeconomic and identity-based barriers. Accordingly, the human capital framework is often used to explain the earnings of individuals as a return on investment following the seminal contributions of Mincer (1958), Schultz (1961), and Becker (1962). Extensive empirical literature has emerged based on this framework (Acemoglu, 2012; Angrist & Krueger, 1992; Card & Krueger, 1992; Goldin, 2016; Heckman, 1979; Patrinos & Psacharopoulos, 2020).

Firstly, this literature review covers the origin and subsequent advances in human capital studies. Secondly, we look at the emergence of focused studies on the impact of Education as a human capital investment and its returns. Thirdly, the chapter highlights the crucial developments in the methodological framework and empirical testing for better-estimating returns to Education. The chapter also highlights the conclusions derived from these research studies, which have been critical in identifying the research gaps and conceptualizing our objectives for this thesis. Based on our objectives, we dwell on the literature on observed variations in private returns to Education across socioeconomic groups in national and international literature. This would allow the reader to understand the narrative in the following main chapters.

The second sub-section, Part b) Deals with literature concerned with educational outcomes and natural calamities. This part of the literature is concerned with chapter 9 in the thesis. Natural disasters have varied direct and indirect effects on human well-being, and some may have immediate consequences and long-term effects. A growing international literature has examined the causal relationship between natural disasters and educational outcomes. One of the long-term effects is the impact on human capital, specifically Education (Baez et al., 2010a, 2010b; Butz et al., 2014; Nguyen & Minh Pham, 2018). Thus, an entire chapter is dedicated to establishing the impact of natural disasters on educational outcomes in the Indian context. The literature in this regard broadly consists of studies that analyze the impact of these disasters on educational outcomes in different regions, for specific types of natural disasters, the vulnerability among individuals living in areas prone to natural disasters, and its impact on intergenerational educational attainment leading to lower human capital accumulations. Lastly, this chapter also bridges the extensive human capital literature focussed on returns to Education with disruptions to educational attainment in the form of natural disasters.

Chapter 3: Data and Methodology

This chapter provides insight into the data and methodology utilized to attain the objectives in the introduction. The thesis uses cross-sectional and panel data to attain the set objectives. This chapter provides an elaborate insight into the nature, structure, and construction (panel data) of the data sets to be used. We use Individual-level data from the Indian Human Development Survey (IHDS) (Desai et al., 2008; Desai & Vanneman, 2015) for two cross-sectional rounds, IHDS I (2004-05) and IHDS II (2011-12). These two rounds of cross-sectional data are then methodically merged and converted to form panel data for 150,988 individuals interviewed in both rounds.

It is also the only national survey that provides longitudinal information on natural disasters and a wide range of socioeconomic characteristics with individual, household, and village-level identifiers. This survey provides information on natural disasters in India at the village level (rural) for 2006 and 2012.

The chapter also describes the available and constructed variables used in the analysis of the subsequent chapters. These variables consist of individual characteristics such as hourly earnings (returns), years of Education (human capital), potential Experience, gender; household characteristics such as networking intensity (social capital) which is social intensity of people's informal networks, social status, economic status, and number of children among others and village characteristics such as natural disaster intensity. The description of these variables are discussed at length in Chapter 3.

We also use the chapter to establish the sub-sample used in the panel data analysis. Since we focus on the earnings of eligible persons in the labor force, we restrict our analysis to individuals belonging to the working age group (15–64 years)—the age group considered eligible to participate in the labor force (The World Bank, n.d.). To study the variation in earnings and Education between the two rounds, we exclude (a) those who were not interviewed in both rounds, (b) those who were still enrolled in school or college during IHDS II and thus ineligible to be in the labor market, and (c) those whose education level did not change between the two rounds.

For chapter 9, which discusses the impact of natural disasters on educational outcomes, we restrict our analysis to rural India (because natural disaster data is available at the village level), featuring individuals and households that were in both rounds of IHDS (IHDS I, 2004-05 and IHDS II, 2011-12) and were below 23 years of age (to restrict the analysis to probable students who were still in the educational system or just exited). The methodology section in the chapter provides a detailed account of methods and techniques used to analyze data for each of the main chapters. It is the nature and empirical basis of the estimated models and econometric tools utilized in those models for each type of data used. We also use econometric techniques such as the Ordinary Least Squares method, Heckman Selection Method, Generalized Least Squares method, and other model specification tests to verify the efficiency of the estimates derived from these models. The methods used in this thesis have a strong base in literature; however, some methods that earlier required manual evaluations are now computable in advanced software. One such estimation model used in this thesis is the Heckman Fixed Effect Model to account for selection bias and endogeneity.

We discuss Mincer's earnings function and the Ordinary Least squares method of estimating using the following function and the derived econometric equation (1.1):

$$\begin{aligned} &\boldsymbol{lnHourly\_Earnings} \\ &= \boldsymbol{f}\ (\boldsymbol{Education}, \boldsymbol{Experience}, \boldsymbol{Experience\_square}) \qquad (\mathbf{1.1}) \end{aligned}$$

The early econometric method for estimating the Mincer earnings function used the ordinary least squares (OLS) equation.

The econometric formulation of (1.1) would then be stated as:

lnHourly_Earningsit = β0 + β1 Educationit + β2 Experienceit +

β3 Experience_squareit + eit (1.2)

Where eit = stochastic error term and β1 = the average private returns to an additional year of Education.

We then discuss that OLS estimates suffer from at least two sources of bias— selection and ability bias (Angrist & Krueger, 1991). Therefore, selecting only those individuals who reported earnings to estimate the returns to an additional year of Education could overestimate the effect of Education. This estimation error, using only the sub-sample of individuals who reported earnings, is called selection bias (Heckman, 1974; 1976; 1979). The issue of endogeneity is discussed as another common bias in these estimations. Individuals may embody innate abilities or genetic traits that aid productivity over and above additional years of Education.

We then discuss the alternative model proposed by Heckman, which is a two-staged model, as follows:

In the first stage (selection equation), we predicted participation in the workforce with instruments such as Married, Nchildren, and Nadults apart from the explanatory variables of the outcome equation (Semykina & Wooldridge, 2010). The selection equation (1.3) is stated as:

Labour Force Participationit = α0 + α1Marriedit + α2NChildrenit + α3Nadultsit + α4(X)it + εit **(1.3)**

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 1.4 and ε = stochastic error term.

In the second stage of the Heckman procedure (outcome equation), we used a semi-log polynomial model to allow second-order covariates for Experience. The outcome equation (1.4)is stated as follows:

lnHourly_Earningsit = β0 + β1 Educationit + β2 Experienceit +

β3 Experience_squareit + β4 Networking Intensityit +

β5 IMRit + β6 Zit + eit (1.4)

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (1.3) (Married, NChildren, and Nadults), e = error term, i = ith individual, and t = year (2005 or 2012).

For the chapter on natural disasters and educational outcomes, our methodology discusses the multiple regression model for panel data. The functional relationship that we wish to explore is as follows:

Education = f (Natural disasters, Household characteristics, Economic cost of Education, Institutional type) (1.5)

Education is the dependent variable measured in an individual's years of completed Education, which depends on natural disasters, Household characteristics, Economic cost of Education, and Institutional type (explanatory variables trying to capture the impact they may have on educational outcomes).

Our effort is to explain variations in educational outcomes for individuals between 2004-05 and 2011-12. We use a panel data approach to estimate the impact of natural disasters on educational outcomes.

The specification equation is stated as (equation 1.6):

Education it = α0 + α1 Natural Disaster Intensity it + α2 Number of children in Household it + α3 Number of Household Assets it + α4 School Fees it + α5 Government School it + α5 Aided School it + α5 Private School it + εit (1.6)

Where,

Education = Years of Education

Natural Disaster Intensity = Ordered rank variable (range lies between 0 to 18)

Number of children in household = Number of children in the household

Number of Household Assets = Number of household assets owned by the household

School Fees = School fees paid by households (in terms of 2011-12 values)

School Type = Type of educational institution (Others (1), Government (2), Aided (3), Private (4))

Overall, this chapter provides the reader with the data context, methods, sub-samples, and techniques used in subsequent chapters.

Chapter 4: Earnings, Education, and Experience in the Indian Context: An Overview

This is one of three of the five major chapters of this thesis. In this chapter, we analyze and present results for Individual level data from the IHDS (Indian Human Development Survey) in 3 sections. Namely, cross-sectional analysis for the period 2004-05 (IHDS I), cross-sectional analysis for the period 2011-12 (IHDS II), and Panel Data (IHDS I & II). The chapter focuses on estimating returns to human capital investment (Years of Education and Potential Job market experience) at the individual level using Mincer's Earnings function.

Returns to human capital investment are estimated using model specifications with the help of Ordinary Least Square. Firstly, the analysis in this chapter makes it certain that the traditional OLS estimates suffer from Selection bias in cross-sectional and panel data. The estimates derived there would be less robust and precise. Alternatively, researchers have used the instrumental variables approach and Heckman selection bias approach to address this issue. We thus resort to using the Heckman-type model in the following chapters. Secondly, the analysis of cross-sectional data and panel data has provided evidence that Education, Experience, and Networking Intensity are significant in explaining the change in the Hourly Earnings of individuals. Lastly, we also learned that the panel data analysis suggests the fixed effects model compared to the random effects model. Thus, the Heckman selection model with fixed effects is more suitable.

We compare the estimates from an extended Mincer function using two methods, namely, OLS and Heckman's Selection Bias Correction method, to give the reader an idea about the range of these comparative estimates.

Results from this chapter are summarised as follows:

Analysis for IHDS I (2004-05) reveals that the traditional OLS method to estimate private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) (0.22) being highly significant. The coefficients for Education were positive and highly significant in the OLS model (0.106) and Heckman Selection model (0.105). The coefficient for Education was marginally higher in the OLS. The coefficient for Experience was positive and highly significant in both models; however, the estimates were underestimated in the OLS model (0.0484) compared to the Heckman selection model (0.055). The coefficient for Experience Square was negative and highly significant in both models, suggesting an inverted U-shape relationship. The Networking Intensity variable was not significant but jointly significant in the overall model.

Analysis for IHDS II (2011-12) reveals that the traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) (0.22) being highly significant. The coefficients for Education were positive and highly significant in the OLS model (0.0835) and Heckman Selection model (0.0826). The coefficient for Education was marginally higher in the OLS. The coefficient for Experience was positive and highly significant in both models; however, the estimates were underestimated in the OLS model (0.0334) compared to the Heckman selection model (0.034). The coefficient for Experience Square was negative and highly significant in both models, justifying an inverted U-shape relationship. The variable Networking

Intensity was positive and highly significant in the OLS model (0.0307) and the Heckman model (0.029), and again, the coefficient was marginally higher for the OLS model.

Analysis for Panel Data IHDS I & II (2004-05 & 2011-12) We first estimate the random and fixed effects models using the xtreg command. We then use the Hausman test to decide on a suitable model. With a Chi-squared value of 186.29 (p-value 0.0000), we conclude that the fixed-effects model is a more suitable fit.

We then proceeded to estimate the Heckman selection model with fixed effects using the xtheckmanfe command. We found that the within estimator (OLS) suffers from selection bias, with the coefficient for the Inverse Mills Ratio for 2005 (-0.2213) and 2012 (-0.3799) being highly significant. The coefficients for Education were positive and highly significant in the within estimator (OLS) (0.0666) and Heckman Selection model (0.0348). This illustrates that the within estimator (OLS) overestimated the coefficients for Private returns to an additional year of Education compared to the Heckman model with fixed effects. The coefficient for Education was nearly twice as significant in the within estimator (OLS). The coefficient for Experience was positive and highly significant in both the models as well; however, the estimates were overestimated in the within estimator (OLS) (0.0608) as compared to the Heckman selection model (0.012) with fixed effects. The coefficient for Experience Square was negative and highly significant within the estimator (OLS) but not significant in the Heckman model. The variable Networking Intensity was positive and highly significant in the within estimator (OLS) (0.0174) and the Heckman model (0.0164); the coefficient was marginally higher for the OLS model. The selection instrument was highly significant in the Heckman-type model. The Overall R square for the within estimator (OLS) was 0.207, and the Heckman-type model does not report an R square.

The number of observations in the within estimator (OLS) is reported as 21291, and in the Heckman model is reported as 51989.

This is an extensive chapter, covering our First objective while setting the tone for the chapters to follow.

Chapter 5: Trends in Returns to Human Capital Investment Across Class

In Chapter 5, we use the models presented in the previous chapter to study the trends in the estimates from OLS and Heckman models across class. We create six consumption categories to explore trends in returns to an additional year of human capital investment in the form of Education and Experience across these categories. The categories consist of one group of BPL (Below the poverty line) and five quintiles of APL (Above the poverty line (APL1 to APL5)). This chapter uses cross-sectional and panel data analysis.

Results from this chapter are summarised as follows:

Analysis for IHDS I (2004-05): In the Heckman Selection model, the coefficients for Education were positive and highly significant across all consumption quintiles, namely, BPL (0.0517), APL1 (0.0457), APL2 (0.0563), APL3 (0.0764), APL4 (0.0996) and APL5 (0.129). The coefficients for Experience were positive and highly significant across all consumption quintiles, namely, BPL (0.0313), APL1 (0.0255), APL2 (0.0351), APL3 (0.0398), APL4 (0.0587), and APL5 (0.0638). The coefficients for the Experience square were negative and significant for all consumption quintiles, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for BPL (0.0327) except APL1, whereas the estimates were negative and highly significant for APL3 (-0.0247), APL4 (-0.0275) and APL5 (-0.0473) except APL2.

The private returns to an additional year of Education in the OLS model range between the lowest for APL1 (5.3 %) and highest for APL5(12.9%). The private returns to an additional year of Education in the Heckman model range between the lowest BPL (5.17%) and highest for APL5(12.9%). The private returns to an additional year of Experience in the OLS model range between the lowest for APL1 (4.5%) and highest for APL5(12.9%). The private returns to an additional year of Experience in the Heckman model range between the lowest for APL1 (2.6%) and the highest for APL5(6.4%). The estimates for Networking Intensity in the OLS model range between the lowest APL5 (- 4.6%) and the highest for BPL(4%). The estimates for Networking Intensity in the Heckman model range between the lowest APL5 (- 4.7%) and the highest for BPL (3.25%).

The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for all quintiles.

Analysis for IHDS II (2011-12): In the Heckman Selection model, the coefficients for Education were positive and highly significant across all consumption quintiles, namely, BPL (0.0412), APL1 (0.0389), APL2 (0.0494), APL3 (0.0636), APL4 (0.0762) and APL5 (0.123). The coefficients for Experience were positive and highly significant across all consumption quintiles, namely, BPL (0.022), APL1 (0.0189), APL2 (0.0248), APL3 (0.0297), APL4 (0.0332), and APL5 (0.0457). The coefficients for the Experience square were negative and significant for all consumption quintiles, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for all quintiles BPL (0.0691) and APL1 (0.0519), APL2 (0.0463), APL3 (0.0285), APL4 (0.0281), except APL5 (-0.0218) which was negative and highly significant.

The private returns to an additional year of Education in the OLS model range between the lowest for APL1 (3.91%) and the highest for APL5(12%). The private returns to an additional year of Education in the Heckman model range between the lowest for APL1 (3.89%) and highest for APL5 (12.3%). The private returns to an additional year of Experience in the OLS model range between the lowest for APL1 (1.73%) and the highest for APL5(4.54%). The private returns to an additional year of Experience in the Heckman model range between the lowest for APL1 (1.89%) and the highest for APL5(4.57%). The estimates for Networking Intensity in the OLS model range between the lowest APL5 (- 2.16%) and the highest for BPL (7.2%). The estimates for Networking Intensity in the Heckman model range between the lowest APL5 (- 2.18%) and the highest for BPL (6.91%).

The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for all quintiles.

Analysis for panel data IHDS I & II (2004-05 & 2011-12)

We found that education has a significant and positive impact on earnings across all quintiles. Experience, too, was positive and significant for APL 2 and APL 5. The squared effect was significant and negative for APL 5. Networking intensity was significant and positive for APL 1. The coefficients for IMR were significant for all quintiles (except BPL and APL 5) in 2005 and 2012, confirming the presence of selection bias. This also justifies Heckman's selection approach in our context.

Chapter 6: Trends in Returns to Human Capital Investment Across Gender

In Chapter 6, we use the models presented in Chapter 4 to study the trends in the estimates from OLS and Heckman models across Gender. We compare the returns to investment in human capital for males and females to highlight the direction and magnitude of these returns. This chapter uses cross-sectional and panel data analysis.

Analysis for IHDS I (2004-05): In the Heckman Selection model, the coefficients for Education were positive and highly significant across all genders, namely, Males (0.0919) and Females (0.078). The coefficients for Experience were positive and highly significant across all genders, namely, males (0.0603) and females (0.045). The coefficients for Experience Square were negative and significant for men and women, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for Males (0.0144) and negative and highly significant for Females (- 0.011).

The private returns to an additional year of Education in the OLS model range between the lowest of males (9.29%) and females (10.2%). The private returns to an additional year of Education in the Heckman model range between the lowest of females (7.8%) and males (9.19%). The private returns to an additional year of Experience in the OLS model range between the lowest of females (3.95%) and males (5.3%). The private returns to an additional year of Experience in the Heckman model range between the lowest of females (4.5%) and males (6%). The estimates for Networking Intensity in the OLS model range between the lowest of females (- 0.8%) (not significant) and males (1.65%). The estimates for Networking Intensity in the Heckman model range between the lowest of females (- 1.1%) and males (1.4%).

The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for both genders.

Analysis for IHDS II (2011-12)

In the Heckman Selection model, the coefficients for Education were positive and highly significant across all genders, namely, Males (0.0721) and Females (0.0790). The coefficients for Experience were positive and highly significant across all genders, namely, Males (0.0354) and Females (0.0355). The coefficients for Experience Square were negative and significant for men and women, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for Males (0.0511) and negative for Females (0.0122).

The private returns to an additional year of Education in the OLS model range between the lowest of males (7.3%) and females (7.9%). The private returns to an additional year of Education in the Heckman model range between the lowest of males (7.2%) and females (7.9%). The private returns to an additional year of Experience in the OLS model range between the lowest of males (3.4%) and females (3.5%). The private returns to an additional year of Experience in the Heckman model range between the lowest of males (3.54%) and females (3.55%). The estimates for Networking Intensity in the OLS model range between the lowest of females (1.34%) (not significant) and males (5.22%). The estimates for Networking Intensity in the Heckman model range between the lowest of females (1.22%) and males (5.11%).

The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for both genders.

Analysis for panel data IHDS I & II (2004-05 & 2011-12)

Education was positive and significant for both males and females. The Experience was positive and significant for females and jointly significant for males. The square term (Experience square) was negative. This suggests an inverted U-shaped relation between earnings and Experience for females. The implication is that while women's earnings increase with Experience, the rate of increase in earnings—caused by a rise in Experience—occurs at a decreasing rate. Networking intensity was positive and significant for males. The IMR was significant for both groups in 2005 and for males in 2012. Returns to females (5.2%) are much higher than to males (3.7%).

The received literature on gender differentials in earnings has universally found that women earn less than men for a given level of Education and Experience. We did not find evidence contrary to this. However, we did find a tendency towards convergence in earnings between males and females for given levels of Education and Experience. This has been noted by other authors as well, and it can be attributed in part to structural and technological change (Cortes et al., 2020). As the workforce has moved from manufacturing to service-oriented employment, women are less likely to be disadvantaged in such work situations.

Moreover, their soft skills at the workplace may be more sought after than those of men. Our findings provide evidence of convergence in the long run in the Indian context. We found that the initial value of earnings (intercept value) for males is much higher than that of females. However, the marginal returns to an additional year of Education are much higher for women than men.

Chapter 7: Trends in Returns to Human Capital Investment Across Caste

In the chapter titled 'Trends in Returns to Human Capital Investment (Education) across Castes, we estimate returns for human capital investment for four caste categories, namely General, Other Backward Castes (OBC), Scheduled Castes (SC), and Scheduled Tribes (ST). The scope of the chapter is restricted to one social group in the Indian context, namely, Caste. We compare the returns to investment in human capital for this social groups to highlight the direction and magnitude of these returns. These disaggregated private returns provide a deeper understanding of the trajectory their marginal earnings follow, given their level of Education and Experience. We are also interested in observing the role of social capital in determining earnings across these social groups.

Results from this chapter are summarised as follows:

Analysis for IHDS I (2004-05): In the Heckman Selection model, the coefficients for Education were positive and highly significant across all caste groups, namely, General (0.0996), OBC (0.0753), SC (0.0837), and ST (0.114). The coefficients for Experience were positive and highly significant across all caste groups, namely, General (0.0493), OBC (0.0455), SC (0.0527), and ST (0.0568). The coefficients for the Experience square were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for all caste groups, namely, General (0.0128), SC (0.0322), and negative and significant for ST (- 0.0574) except OBC (-0.00807), which was not significant.

The private returns to an additional year of Education in the OLS model range between the lowest for OBC (8.5%) and highest for ST (11.5%). The private returns to an additional year of Education in the Heckman model range between the lowest for OBC (7.6%) and highest for ST (11.4%). The private returns to an additional year of

Experience in the OLS model range between the lowest for OBC (4.25%) and highest for ST (5.45%). The private returns to an additional year of Experience in the Heckman model range between the lowest OBC (4.6%) and highest for ST (5.7%). v OLS model range between the lowest ST (-5.5%) and highest for SC (4.7%). The estimates for Networking Intensity in the Heckman model range between the lowest for ST (- 5.7%) and highest for SC (3.2%).

The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for all quintiles.

Analysis for IHDS II (2011-12) In the Heckman Selection model, the coefficients for Education were positive and highly significant across all caste groups, namely, General (0.0772), OBC (0.0633), SC (0.0848), and ST (0.0894). The coefficients for Experience were positive and highly significant across all caste groups, namely, General (0.0335), OBC (0.0263), SC (0.0424), and ST (0.0357). The coefficients for the Experience square were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for all caste groups, namely, General (0.0374), OBC (0.0382), SC (0.0308), and ST (0.0281).

The private returns to an additional year of Education in the OLS model range between the lowest for OBC (6.4%) and highest for ST (8.99%). The private returns to an additional year of Education in the Heckman model range between the lowest OBC (6.3%) and highest for ST (8.9 %). The private returns to an additional year of Experience in the OLS model range between the lowest for OBC (2.35%) and the highest for SC (3.95%). The private returns to an additional year of Experience in the Heckman model range between the lowest OBC (2.63%) and highest for SC (4.2%).

The estimates for Networking Intensity in the OLS model range between the lowest ST (2.85%) and the highest for OBC (3.9%). The estimates for Networking Intensity in the Heckman model range between the lowest for ST (2.8%) and highest for OBC (3.8%).

The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for all quintiles.

Analysis for panel data IHDS I & II (2004-05 & 2011-12)

We found that Education has a positive and significant impact on all castes except SC. The Experience was positive and significant for general castes. Experience square was negative and significant for general. This suggests a U-shaped relation between earnings and Experience for the group (general). Networking intensity was not significant. The IMR was significant for most of the values, indicating the presence of a selection bias.

Chapter 8: Trends in Returns to Human Capital Investment Across Religion

In the chapter titled 'Trends in Returns to Human Capital Investment (Education and Experience) across Religion, we estimate returns for human capital investment for 3 Religion groups, namely Hindus, Other Minorities (Christian, Jain, Sikh, Buddhist, and others), and Muslims. We compare the returns to investment in human capital for these religious groups to highlight the direction and magnitude of their returns. These disaggregated private returns would provide a deeper understanding of the trajectory their marginal earnings follow, given their level of Education and Experience. We are also interested in observing the role of social capital in determining earnings across these religious groups.

Results from this chapter are summarised as follows:

Analysis for IHDS I (2004-05):

The coefficients for Education were positive and significant across all religious groups, namely, Hindus (0.0103), Other minorities (0.116), and Muslims (0.0778). The coefficients for Experience were positive and significant across all religious groups, namely, Hindus (0.0536), Other minorities (0.04), and Muslims (0.0469). The coefficients for Experiencesqr were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were negative and significant only for Other minorities (-0.0408).

The private returns to an additional year of Education in the OLS model range between the lowest for Muslims (8.1%) and highest for Other Minorities (11.5%). The private returns to an additional year of Education in the Heckman model range between the lowest for Muslims (7.78%) and highest for Other Minorities (11.6%). The private returns to an additional year of Experience in the OLS model range between the lowest for other minorities (3.6%) and highest for Hindus (4.93%). The private returns to an additional year of Experience in the Heckman model range between the lowest for other minorities (4%) and highest for Hindus (5.36%). The estimates for Networking Intensity in the OLS & Heckman model only significant for Other Minorities (- 4.1%). The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for all religions.

Analysis for IHDS II (2011-12)

In the Heckman Selection model, the coefficients for Education were positive and highly significant across all religious groups, namely, Hindus (0.0784), Other minorities (0.104), and Muslims (0.0651). The coefficients for Experience were positive and highly significant across all religious groups, namely, Hindus (0.0324), Other minorities (0.0340), and Muslims (0.0469). The coefficients for the Experience square were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for Hindus (0.0357) and Muslims (0.0314) and were not significant for other minorities (0.0074).

The private returns to an additional year of Education in the OLS model range between the lowest for Muslims (6.5%) and highest for Other Minorities (9.7%). The private returns to an additional year of Education in the Heckman model range between the lowest for Muslims (6.5%) and highest for Other Minorities (10.4%). The private returns to an additional year of Experience in the OLS model range between the lowest for Hindus (3.02%) and highest for Muslims (4.69%). The private returns to an additional year of Experience in the Heckman model range between the lowest for Hindus (3.24%) and highest for Muslims (4.7%). The estimates for Networking Intensity in the OLS model were lowest for Muslims (3.2%) and highest for Hindus (3.76). The estimates for Networking Intensity in the Heckman model were lowest for Muslims (3.14%) and highest for Hindus (3.57).

The traditional OLS method to estimate Private returns to an additional year of Education and Experience suffers from selection bias, with the coefficient for the Inverse Mills Ratio (IMR) being highly significant for all quintiles.

Analysis for panel data IHDS I & II (2004-05 & 2011-12)

Education was significant and positive for all three religious groups. Experience, on the other hand, was significant and positive for Hindus and Other minorities but not for Muslims. Networking intensity was significant and positive for Hindus only.

Chapter 9: Impact of Natural Disasters on Educational Attainment in India: A Panel Data Analysis

This chapter focuses on India, which is highly vulnerable to climate change-induced natural disasters (Anees et al., 2020; BMTPC, 2006). According to some estimates, almost 85 percent of its geographical area is vulnerable to natural disasters (Paik, 2018). Natural disasters have been found to impact educational outcomes in India significantly negatively (Parida et al., 2021).

Studies have found that natural disasters disrupt welfare through widening inequalities, increased disparities, and long-term impacts on Education. This impact on Education can lead to human capital loss and adversely affect future growth and development. The research in this area has either focussed on the role of educational attainment in reducing vulnerability to natural disasters (Fuchs & Thaler, 2018; Muttarak & Lutz, 2014) or the impact of natural disasters on educational attainment (Baez et al., 2010; De Vreyer et al., 2015). We follow the latter by estimating the impact of natural disasters on educational attainment in India.

Most studies are region or disaster-specific, like floods (Thamtanajit, 2020), droughts (Garbero & Muttarak, 2013), and earthquakes (Di Pietro, 2018; Paudel & Ryu, 2018), among others. Some studies based on cross-section data (Joshi, 2019) have the limitation that they do not observe the same households over time. Other studies have been region-specific (Suar & Kar, 2005) and, therefore, lack the coverage of a national-level study. All these studies fail to account for the impact on Education based on the frequency of natural disasters at a country level using household longitudinal data, especially in large countries like India. Our study bridges this research gap by incorporating the severity of natural disasters in a longitudinal dataset covering all states and union territories in India.

We use data from two of the latest rounds of the Indian Human Development Survey (IHDS) and employ a panel difference in difference regression model with continuous treatment fixed effects at the individual level. This allows us to examine the impact of natural disasters on education outcomes between 2004-05 and 2011-12 at the individual level. Our estimates improve upon all earlier studies on this theme that have relied on cross-section or pseudo-panel data at the district level. We provide first estimates of the impact of natural disasters on educational attainment disaggregated by social (including caste, religion, and gender) and economic groups (consumption quintiles).

We use a natural experimental study approach (Quasi-experimental design), using Natural Disasters as a continuous treatment to estimate the average treatment effect of natural disasters on Educational Attainment in India. A natural experiment is said to be a form of quasi-experimental design if an intervention where the researcher does not control the circumstances of implementation of the intervention itself (Leatherdale, 2019). Such approaches are most preferred in the absence of randomized controls in studies where researchers must evaluate the effectiveness of programs or interventions (Chen et al., 2020; González et al., 2021). (Tamuly & Mukhopadhyay, 2022) use this approach to estimate the impact of natural disasters on the well-being (measured by consumption expenditure) of households in India. We follow (Tamuly & Mukhopadhyay, 2022) in constructing and using a continuous measure of the frequency of occurrence of natural disasters as Natural Disaster Intensity (NDI), trust in public institutions as Confidence Intensity, membership in various socioeconomic groups as Membership Intensity (a measure of social capital) and other covariates used by them to estimate their impact on educational attainment. In addition, we include School fees, School Type (Government, Aided, Private), and Repeated (if an individual has ever repeated a class) as covariates to estimate their impact on Educational Attainment.

We find that natural disasters significantly and negatively affected Education for Women and three Consumption Quintiles (including those below poverty), Other Backward Castes, Scheduled Castes, Muslims, and other Minorities.

We also find that the number of household assets, the number of children in the household, school fees, school type, Confidence intensity, health insurance, and life insurance influence educational years.

Our results have significant implications in the context of quality education (SDG 4), gender equality (SDG 5), reduced inequalities (SDG 10), and climate change (SDG 13). The study evaluates the impact of the rising incidence of climate extremes vis-a-vis natural disasters as external shocks on educational outcomes. Educational attainment is directly linked to earnings and is well-established in the literature. We thus believe the negative impact of natural disasters on educational attainment, especially for the marginalized, may have long-term disruptive effects beyond immediate losses and may last through generations.

Natural Disaster Intensity (NDI) yields a negative and significant Average Treatment Effect on the Treated (exposed to natural disasters) on Educational Attainment. Our findings from disaggregated estimates add to these results by illustrating the severity of these losses to educational attainment for marginalized sections in rural India.

Our results are consistent for Females, three consumption quintiles (BPL, APL4, and APL5), two marginalized caste categories (OBC and SC), and religious minorities (Muslims and Other minorities).

We also find that the number of children in the household negatively and significantly impacts educational attainment for both genders, consumption quintiles (BPL, APL1, and APL5), social groups (OBCs), and all religions.

The number of household assets has a positive and significant impact on educational attainment for both genders, consumption quintiles (BPL and APL1), social groups (SCs), and religion (Hindus and Muslims).

We find that school fees have a positive and significant impact on educational attainment for consumption quintiles (APL1) and social groups (SCs) and a negative and significant impact on religion (other minorities). However, we did not test for changes in enrolment.

Apart from school fees, we also looked at school type since it captures additional indicators beyond school fees, including the learning atmosphere. In school type, our reference category was other schools. We find that Government and Aided schools had a significant positive impact on educational attainment across gender, consumption quintiles (except APL2, -ve), caste (SCs, but negative for STs), and religion (Hindus). On the other hand, Private schools negatively impacted educational attainment for both genders, APL2, Hindus, and Muslims.

Membership Intensity negatively and significantly affects STs. Confidence Intensity positively and significantly affects women, BPL, and Hindus, whereas it negatively affects STs. Financial resilience in the form of Health insurance positively and significantly affects APL3 and other minorities, whereas it negatively affects APL1 and APL4, OBCs, and STs. Financial resilience in the form of Life Insurance positively and significantly affects men, APL4, whereas it negatively affects SCs. Repeating in a class negatively and significantly affects both genders, BPL and APL4, general and OBCs, and Hindus and Other minorities.

Our findings reaffirm that climate change (SDG13) vis a vis natural disasters have a significant adverse effect on educational attainment. We additionally conclude that the impact of natural disasters on educational attainment for marginalized sections (women, BPL, OBCs, SCs, minority religions) of the society is much more severe.

Chapter 10: Findings and Conclusions

In this chapter, we summarise our findings from the preceding chapters. The chapter also establishes the relevance of these findings. After that, we provide our conclusions and policy implications of the study.

While women have lower starting earnings, they have higher incremental earnings as they gain in Education and Experience. Studies have suggested that men gain more from additional years of Education than women in terms of wages. We find contrary evidence that even though women start with lower wages than men, their marginal increase for additional years of Education and Experience out-pace the growth of wages for men. To our understanding, no study has used a panel data set in the Indian context that has examined the role of Education and Experience in wage determination with a nationwide representative sample with a household-level panel thus far. Our results, therefore, update all earlier estimates in this context.

There is a wide range of variation in the reported private returns to Education in India. However, all earlier estimates in the Indian context have used cross-sectional data or a pseudo-panel approach. We improve on them by using a nationwide household-level panel with fixed effects for the Heckman selection model approach.

Our study also examined the role of social capital by examining the impact of networking intensity on different groups. We find that this variable had a positive impact on the earnings of males (among gender groups), Hindus (among religious groups), and APL1 (among consumption quintiles). Our results align with a study based in China (Liu, 2017), where individual-level social capital was found to have a more significant impact on men than women. This could be anticipated in India and might explain the wage gaps for women and marginalized social and religious groups (Deshpande & Khanna, 2021).

In conclusion, we need a targeted approach to ensure these groups stay in the education system because of their lack of ability to pay (Mishra & Ramakrishna, 2023). The Right of Children to Free and Compulsory Education (RTE) Act of 2009 obligates the government to ensure that every child in the age group of 6 to 14 years is provided free and compulsory Education (GoI, 2009). This is a step towards inclusive development (Das, 2020), although there is evidence of unintended outcomes (Chatterjee et al., 2020). Educational affirmative interventions based on caste and gender have been found to have positive impacts in India (Bagde et al., 2016). Our findings are similar to Chakraborty & and Bohara's (2021) study on the role of caste and religion in India and their recommendation for policy intervention for socially backward classes and Muslims to establish equity in the labor market.

In the chapter on Natural disasters and educational outcomes, we conclude that On one hand, lower educational attainment (SDG4) in disaster-prone areas could trigger a chain of disruptive welfare effects in the form of lower overall earnings, widening of gender differences (in attainment and earnings) (SDG5), and higher societal inequalities (SDG10). On the other hand, these lower educational attainments reduce the effectiveness and increase the cost of disaster management programs developed for mitigation from natural disasters and to reduce vulnerabilities. Studies have shown that higher education levels lead to better awareness levels and ease of dissemination of information and serve as a measure of development with direct and indirect effects on reducing fatality rates (Parida et al., 2021). The higher vulnerability combined with the increasing frequency of natural disasters in recent times and the inability to migrate leaves the marginalized in danger of intergenerational disruptions in well-being. This possibly creates a spiral of lower education, lower earnings, and lower welfare, inevitably pushing them into poverty. The impacts of climate change on human capital demand effective policy intervention to mitigate human capital losses to marginalized sections. The challenge then is to create a targeted response to more vulnerable sections of the population by 1) minimizing dropout rates disasters, 2) strengthening schooling mechanisms to resume educational programs rapidly, and 3) reducing the probability of household's decisions to switch from investment in human capital (Education) to rebuilding damages to physical assets by providing definitive monetary assistance in rehabilitation.

1.7. CONTRIBUTION OF THIS THESIS

We re-examine the issue of private returns to education earnings in India and contribute to this literature in multiple ways. We use data from two rounds of IHDS to construct a panel data set. To our understanding, ours is the first attempt using a panel data model with individual-level fixed effects. We follow the Heckman selection model and use a new and improved command (xtheckmanfe) in Stata16.1 to control for selection bias.

The thesis adds to the literature on the methodology applied, robust estimation, and coverage of interrelated topics between earnings, education, and natural disasters.

Firstly, we add to the literature by providing panel estimates using a Heckman-type selection model to the private returns to an additional year of education and experience. Secondly, we provide first estimates on the private returns to an additional year of education and experience disaggregated by consumption quintiles, caste, gender, and religion using a nationally representative dataset (Hussain & Mukhopadhyay, 2023). This improves Mitra's method (2019), which uses quintiles based on wages without separating people experiencing poverty from the non-poor. Our analysis provides greater disaggregation, which can help devise more effective and targeted policies for affirmative action. Previously, studies in India have employed a cross-section data set (Agrawal, 2014; Bairagya, 2020; Geetha Rani, 2014; Kingdon & Theopold, 2008; Mitra, 2019) or a quasi-panel (Bhattacharya & Sato, 2017; Mendiratta & Gupt, 2013) to study earnings in India.

Our panel data approach with individual fixed effects improves upon estimates of all earlier studies on this subject. Thirdly, the thesis explores the impact of social capital in the form of social networking, from membership of households to various economic and social associations.

Lastly, we use the natural calamities data in IHDS to estimate its impact on educational outcomes. Most studies relating to natural disasters are region or disaster-specific, like floods (Thamtanajit, 2020), droughts (Garbero & Muttarak, 2013), and earthquakes (Di Pietro, 2018; Paudel & Ryu, 2018), among others. Some studies based on cross-section data (Joshi, 2019) have the limitation that they observe different households over time. Other studies have been region-specific (Suar & Kar, 2005) and lack the coverage of a national-level study. All these studies fail to account for the impact on education based on the frequency of natural disasters at a country level using household longitudinal data, especially in large countries like India. Our study bridges this research gap by incorporating the severity of natural disasters in a longitudinal dataset covering all states and union territories in India.

Our objective is to estimate the impact of natural disasters (with differential intensity) on educational outcomes in rural villages. We provide estimates disaggregated by Gender (Male and Female), Consumption quintiles (Below the poverty line (BPL) and quintiles above the poverty line, from APL1 to APL5), Caste (General, Other backward castes, Scheduled castes and Scheduled tribes) and Religion (Hindu, Other minorities and Muslim).

This disaggregation would provide more robust estimates to analyze the impact of natural disasters on educational outcomes across sections of the sub-population.

1.8. LIMITATIONS

While several innovations have been implemented in our study, we acknowledge some limitations. We could not separate the impact of ability and inherent skills from the years of Education in wage determination (ability bias). The reporting of caste was non-uniform in the two rounds of IHDS. This could be due to the re-classification of castes, which was an ongoing process at the time of the survey. This could be further resolved when data from new ongoing survey rounds becomes available. Also, our study did not stratify the labor market by state and sector. Incorporating these classifications could yield new insights into a country with immense geographic and economic heterogeneity like India.

In Chapter 9, we used a panel regression model; using an instrumental variable approach was not possible due to the inability to find a suitable instrument. Secondly, the Natural disaster data was available only in the 2011-12 wave of IHDS, and we had to improvise methodologically to create a panel variable for Natural calamity intensity. This analysis is restricted to rural areas since data for natural calamities was a part of the village data.

1.9. POLICY IMPLICATIONS

Firstly, our findings have important policy implications and connect to multiple SDGs, such as poverty (SDG 1), quality education (SDG 4), gender equality (SDG 5), and reduced inequalities (SDG 10). The government of India has been committed to fulfilling the SDGs and has undertaken multiple social interventions. In the sphere of Education, policies have also been developed to enhance access and improve quality education. About a decade and a half ago, the Right of Children to Free and Compulsory

Education Act, or Right to Education Act, 2009, for free and compulsory schooling (GoI, 2009) was enacted to reduce the dropout rate in schools and ensure compulsory Education for up to 14 years. The New Education Policy announced in 2020 also attempts to upgrade quality education. Our findings provide evidence that education (SDG 4) helps achieve the goals of gender equality (SDG 5) and social justice for marginalized (caste and religious) groups (SDG 10).

The government has also implemented affirmative action policies in jobs and Education for specific groups mainly based on caste (Deshpande, 2013) and, more recently, economic backwardness (Nath, 2019). Despite Education positively impacting earnings, the less privileged groups would require affirmative action to achieve reduced inequalities (SGD 10) and escape poverty (SGD 1).

Our study addresses the issues of economic distribution, gender, caste, and religion in the context of larger sustainable developmental goals (Dhar, 2018; Pandey, 2019). Our findings of higher marginal returns to Education among females justify the need for greater investment in Education for women. The lower returns on Education among STs and Muslims indicate the need for affirmative action for these groups. Similarly, the positive and significant coefficient for all consumption quintiles suggests that anti-poverty programs in combination with educational opportunities for the less privileged (BPL) would meet the goals of social justice. Development policies that combine multiple objectives would have a more effective welfare outcome.

Secondly, given that (a) India is considered to be a highly vulnerable country to extreme events in the context of climate change (BMTPC 2006) and (b) large segments of the Indian population lack basic educational access (Jana, 2020), it is crucial to understand how natural disasters have impacted education attainment. This would be important for policymakers and communities that are constantly trying to mitigate and adapt to climate change. One policy implication that emerges from this is that regions facing natural disasters could increase educational access, which would help marginalized populations accumulate human capital. Expanding educational infrastructure could also provide temporary shelters during extreme weather events likely to intensify with climate change.

1.10. CONCLUSION

In this chapter we provided a synopsis view of the entire thesis. We introduced the concept of human capital and its link to growth and development. We discussed the research question and research gaps that thesis addresses. We framed the objectives based on our research questions. The chapter also provides the contribution of the thesis in adding to literature, while acknowledging the limitations and highlighting the policy implications from findings of the thesis.

In the next chapter we highlight prominent literature from across the world and in India pertaining to our study.

CHAPTER 2: LITERATURE REVIEW

Chapter 1 provided a synopsis view of the entire thesis. It introduced the link between human capital, growth, and development.

In this chapter, we provide a selective summary of literature that engages with human capital theory in the context of returns to education. The chapter is divided into 10 sections. Sections 2.1 to 2.8 pertain to literature on human capital theory and returns to education. Section 2.9 discusses an emerging area that links Education and Earnings with the environment, specifically natural disasters. Section 2.10 concludes the chapter.

2.1. INTRODUCTION

The conceptual background of investment in human capital relies primarily on three primary forms of investment: education, On-the-job training, and Health (Becker, 1962, 1993; Mincer, 1974; Schultz, 1961). The idea behind this is that an individual investing in himself/ herself in education, health, or job training leads to a sort of gain in returns to those investments in a future period (Goldin & Katz, 2020). In other words, individuals invest in themselves to maximize their returns, including higher earnings and other non-monetary gains (Flabbi & Gatti, 2018). Recent studies have classified these components of human capital into separate categories to study their impact on earnings. Many of these studies focus on individuals' returns to investment in education (Card & Krueger, 1992; Griliches, 1977; Heckman et al., 2018; Kanjilal-Bhaduri & Pastore, 2018; Mitra, 2019).

Human capital theory was initially frowned upon for considering people as capital assets that could be owned as resources (Deming, 2022; Goldin & Katz, 2020). Over the years, extensive literature on Human capital has emerged across the globe. An individual choosing to spend more time acquiring education would face an opportunity cost of not being in the labor force. As with any form of investment, investing in oneself (for example, in education) brings returns only in the future. Studies from across the world find that returns to education in the form of earnings are positively impacted by education (Asadullah & Xiao, 2020; Gao & Smyth, 2015; Harmon et al., 2003; Heckman et al., 2018; Leigh, 2008; Psacharopoulos & Patrinos, 2018; Wincenciak et al., 2022). Mincer (1958) is the first systematic effort to formalize the human capital model. There are also early references to contribution; Goldin (2016) points to the early work of Adam Smith (1776), who mentioned the acquisition of talents and skills. Reference is also made to Irving Fisher's work, which dates back to 1897. Schultz (1961) is considered to have triggered the onset of research in the area of Human capital theory, which then made way for more specific research interests in return to education. In 1962, Mincer tried to solve the paradox of work and leisure choices by including factors such as education.

According to Deming (2022), human capital theory brings forth four key points: Firstly, it explains the significant variations in labor earnings within and across countries. Secondly, it emphasizes the high economic returns associated with investing in childhood and young adulthood. Thirdly, while there is a clear understanding of the technology required to produce basic skills, the primary obstacle is the availability of resources. Finally, the production of higher-order skills such as problem-solving and teamwork remains an area where technology is lacking despite its significant economic value.

2.2. MACROECONOMIC AND MICROECONOMIC PERSPECTIVE ON HUMAN CAPITAL

Human capital theory has a strong foothold in both macroeconomic and microeconomic theory. Since research in human capital theory has emerged more prominently since the 1950s, we find two broad research areas.

In the macroeconomic framework, economists have always been intrigued by an unexplained part of the growth rates of nations (Denison, 1962; Romer, 1989). Various cross-country studies try to compare national output and average educational attainments to explain gaps in economic performances across nations (Barro, 1999; Becker et al., 1990; De la Fuente & Doménech, 2006).

A basic production function (eq 2.1) can be expressed as below:

$$\boldsymbol{Y = A\, f\, (L, K)} \qquad \textbf{(2.1)}$$

Where Y is the measure of national income (GDP), A measures technology, and L and K are factors of production ‘Labour and Capital,’ respectively. Assuming constant returns in Cobb Douglas production function, one can then outline that growth in national incomes can be generated by either technology (increase in innovation), an increase in capital stock, or by increasing labor participation (Flabbi & Gatti, 2018). However, this reduces the Labour component to homogenous units and does not account for differences within labor units for productivity or gains in productivity through additionally acquired skills. This led to the incorporation of human capital into production functions. These models have been highlighted in (Lucas, 1988; Romer, 1989), leading to what we now refer to as endogenous growth models.

A similar framework has been used to explain growth convergence among nations and estimate causal relations between human capital and national income (Cuaresma et al., 2013; Guaitoli, 2000; Soukiazis & Cravo, 2008). Over a couple of decades of research, it has been observed that education, as measured by attainment, positively impacts national income. However, estimates vary widely across data sets and studies. These have been attributed to several factors, such as quality of data and measurement errors, inability to incorporate a quality indicator of education, reverse causality, and methodological biases due to omitted variables (Flabbi & Gatti, 2018). Broadly, the research identifies three ways that human capital in the macro framework affects national income across countries: direct effect through human capital, indirect effect through interaction with physical capital, and its interaction with technology (endogenous models) (Flabbi & Gatti, 2018).

Micro-level studies in human capital theory emerged from the works of eminent economists like Jacob Mincer (1958) and Theodore Schultz (1961). Micro-level have an advantage over macro-level studies since they a) do not need aggregation of variables at the national level, b) account for heterogeneity, c) can use flexible specifications (across groups within the nation), and d) use additional control in estimation (Flabbi & Gatti, 2018). Studies at the micro level focus on different aspects, such as explaining changes in earnings of individuals based on investment in education and other forms of human capital (training and medical, though minimal), early childhood intervention, and intergenerational linkages between human capital investment and earnings (Flabbi & Gatti, 2018). We restrict our survey to specific studies that try to estimate the private returns to investment in human capital (additional years of educational attainment).

Despite the stated advantages of micro-level studies of human capital, they bring their challenges. Most prominently cited in the literature are the modeling and specification issues (Angrist & Krueger, 1999; Card, 1999; Griliches, 1977; Heckman, 1976, 1979). One such specification issue highlighted predominantly is the presence of selection

(Heckman, 1990; 1976) and ability biases (Belzil & Hansen, 2002; Card, 1993, 1999) in the estimation. It is now well-documented that methodological errors tend to overestimate or underestimate returns to private investment (Blundell et al., 2001; Griliches, 1977a; Montenegro & Patrinos, 2021). Other issues pertain to self-reported wages and proper measures of the variable for human capital (Heckman, 1976). Lastly, though the issues mentioned above can, to some extent, be resolved or addressed, the more significant issue of relating findings from micro studies to macro studies has been a challenge (Marginson, 2019; Tan, 2014). Flabbi and Gatti (2018) point out that looking at earnings or wages instead of national income may overestimate or underestimate the effect of human capital on growth. In other words, though suitable to measure private returns, the estimated earnings function might not fully capture the larger interest in social returns. Nevertheless, earnings are a significant part of national income.

2.3. HUMAN CAPITAL INVESTMENT AND PRIVATE RETURNS IN THE FORM OF EARNINGS

The human capital framework is often used to explain the earnings of individuals as a return on investment following the seminal contributions of Mincer (1958, 1974), (Schultz, 1961), and (Becker, 1962). This framework has resulted in the emergence of extensive empirical literature (Acemoglu & Autor, 2012; Angrist, Djankov, Goldberg, & Patrinos, 2021; Beam, Hyman, & Theoharides, 2020; Becker, 1994; Collin & Weil, 2020; England & Folbre, 2023; Faggian, Modrego, & McCann, 2019; Goldin, 2016; Patrinos & Psacharopoulos, 2020). This large pool of literature has significantly enhanced the understanding of the subject matter being discussed in this thesis. This sub-section highlights the most prominent contributions to estimating private returns to an additional year of human capital investment (specifically education). As pointed out above (in sub-section 2.2), these studies lie within the micro framework of human capital research.

The earliest attempt at conceptualizing investment in human capital into a modeling framework can be seen in the formulation of the earnings function. The most popular among these is the Mincerian Earnings function (Mincer, 1958, 1974). It estimated the private returns in the form of earnings to schooling and experience for individuals motivated by two distinct concepts (Heckman et al., 2006). Firstly, the compensating differences model (Mincer, 1958). Secondly, the accounting identity model (Mincer, 1974). Heckman et al. (2006) briefly explained these models.

Five major categories of human capital investment are discussed by Schultz (1961): 1) health facilities and services, 2) on-job-training, 3) formal education, 4) study programs for adults and 5) migration for job opportunities. Health involves not merely quantitative aspects, such as expenditure on medical care, but also qualitative aspects, such as quality of food intake, among others. This raises the issue of whether all expenditures on health should be considered human capital investment. As individuals become better off economically, better quality of food changes from being a producer good to a consumption good.

This argument has always existed in human capital literature, even for other forms of human capital investment. What is the point at which investment in human capital is no longer an investment but consumption? Investment in health as a form of human capital investment is studied using indicators such as height, weight, and BMI (body mass index) (Schultz, 2003), among others.

On-job training as human capital investment has been explored widely in the past (Black & Lynch, 1996; Duncan & Hoffman, 1979; Mincer, 1962), as it still is in more recent times (Lyons, 2020; Tian et al., 2022). Schultz (1961) believed that the expansion of education had not discarded the need for job training and that more research on the subject was needed. This form of investment in human capital could be undertaken by the individual, the firm, or the government (public enterprises). There are two issues here, then; firstly, who benefits most from investment in job training? Moreover, who should undertake this investment? The issue with firms and public enterprises is the low retention rate of employees undertaking the training. Becker (1975) discusses two forms of job training: 1) general and 2) specific training.

A general form of training provides the individual with opportunities in firms' industries other than the one providing the on-the-job training. In other words, the increase in the worker's productivity is the same in the firm providing the training as in another firm that could use the worker with these skills (Becker, 1975). In comparison, specific-training increases a worker's productivity more in the firm providing the training than in other firms. Though measures of on-the-job training can be identified at particular points in time, the issue is capturing the stock of addition to human capital in a period. However, it is essential to understand that on-the-job training is integral to accounting for human capital investment.

In addition, the migration aspect of human capital investment has generated mixed perspectives. While some studies suggest that migration increases wages (Hunt & Kau, 1985; Yezer & Thurston, 1976), others argue that the impact on earnings for migrating individuals is only higher for the better educated. It can even be harmful to those who are unskilled (Korpi & Clark, 2015). Wage outcomes are also affected by whether the migration is motivated by an investment or non-economic reasons (Zhao, 1999).

Studies on Adult literacy programs and other extension activities have also reinforced the idea that investment in human capital does bear higher returns (de Baldini Rocha & Ponczek, 2011; Green & Craig Riddell, 2003; Tuijnman et al., 1988).

We now turn to formal education as a form of human capital investment. This is probably the most widely studied area within human capital theory. Schultz (1961) comments that higher education investment could explain a significant part of the rise in earnings. He also accounted for the direct cost of education and indirect costs that exist as foregone earnings. The opportunity cost of education and foregone earnings are integral to human capital theory. The earliest seminal work cited in the literature is Becker (1962), followed by subsequent contributions like Becker (1993). This illustrated that the private rate of return on human capital (education) investment is much more significant than other capital.

Private returns to education investment have been widely studied using the Mincerian regression developed by Jacob Mincer (Mincer, 1958, 1974). The Mincer earnings function typically estimates a semi-log model of earnings that depends on education and experience. This model includes a quadratic term for experience to capture any nonlinear relationships, among other variables. Ben-Porath (1967, 1970) has also been influential in explaining how human capital is produced and in tracking the life cycle of earnings.

Empirical work on human capital has examined different aspects, such as testing the validity of Mincer’s earnings function over time (Heckman et al., 2006), critically examining the specification issues (Heckman, 1979, 2003) (Heckman et al., 2003), testing the causal relation between education and other human capital variables (Heckman et al., 2018) and establishing intergenerational linkages between human capital and income (Black et al., 2005; Han & Mulligan, 2001; Huang, 2013).

The estimation of returns to education has posed several methodological challenges. Studies have shown that models estimating earnings from human capital (education) suffer from selection bias (Heckman, 1990;1976, 1979) and ability bias (Behrman &

Rosenzweig, 1999; Belzil & Hansen, 2002; Griliches, 1977a). These econometric issues and the suggested solutions are discussed in the following sub-section 2.4.

2.4. METHODOLOGICAL IMPROVEMENTS

Formal education has been used as an indicator of human capital. Some have taken individual characteristics such as years of educational attainment and learning-adjusted years of schooling (Barro & Lee, 1996, 1993; Filmer et al., 2020). While others have used school-level characteristics like student-teacher ratio, average term length, and percentage of schools with electrification and water (Bedi & Edwards, 2002; Card & Krueger, 1992b).

Returns to education can be perceived in different ways. It can include private returns, social returns, and returns from labor productivity. Private returns refer to individual returns bearing the cost of education, and social returns include the spillovers and externalities from investing in education to society. Returns to labor productivity provides an aggregated view of the same in terms of growth (Blundell et al., 2001). The estimation of returns to education has typically provided varying results across the world. Researchers have employed diverse frameworks and data sets across nations to estimate returns to education. This section highlights some of these frameworks and their contributions.

The issues recorded in the literature to estimate returns to education include basic elements of the earnings function like the theoretical basis, functional form, and suitability of measures of education and earnings. More complex issues deal with causal modeling like measurement errors, selection bias, ability bias, limitations of Instrumental variables, and heterogeneity of returns, among others (Card, 1999; Heckman, 1990).

The Mincerian earnings function (Mincer, 1974) is still perhaps the most widely used model for such estimation. However, conventional wage determination models using Mincer's earnings function with ordinary least squares (OLS) methods are likely to suffer unobserved individual heterogeneity (Flabbi & Gatti, 2018) in the form of both selection and endogeneity bias (ability bias) (Heckman, Lochner, & Todd, 2006; Wooldridge, 2018). This has led to an entire domain of research focussed on improving the methodology of estimating this return to provide a better picture of the real impact of education on earnings or incomes.

The issue of selection bias refers to the non-random selection of samples and may arise from two sources: self-selection or selection by analysts (Heckman, 1976, 1979). These issues are commonly seen in observational and experimental studies (Ellenberg, 1994). Heckman (1974) provides a mechanism to overcome the problem of selection bias. Apart from cross-country studies (Patrinos & Psacharopoulos, 2020; Peet, Fink, & Fawzi, 2015), the Heckman selection model has since been used to estimate returns on education in different parts of the world, such as the United States (Angrist & Keueger, 1991; Heckman and Polachek, 1974), Europe (Harmon, Walker, & Westergård-Nielsen, 2001), and China (Li, 2003). (Ahmed & Chattopadhyay, 2016) also suggest that Heckman's two-stage is better suited over the traditional OLS method.

Ability bias refers to a form of endogeneity; in the context of returns to education, it refers to the endogeneity in the human capital indicator of education. Ability bias arises from the correlation between ability and capabilities (unobserved) and the marginal cost of schooling (Card, 1999; Griliches, 1977a; Wooldridge, 2012) and may result in biased estimates. One way of overcoming ability bias is by incorporating measures of ability like math and verbal ability, amongst others (Cawley et al., 1998; Lemieux, 2014). The most widely used method is the instrumental variables approach. However, literature shows that identifying a strong instrument in itself is a difficult task (Bound et al., 1995; Card, 1993, 1999). The instrumental variable approach originates from natural experiments, examples of which can be found in Angrist & Krueger (1992). In recent years, we have also found methods such as RCT (randomized control trial) to surpass endogeneity issues in education (Coelli & Tabasso, 2019).

2.5. IMPORTANCE OF HUMAN CAPITAL (EDUCATION) IN RESEARCH

Education is an important form of investment in the human capital (Becker, 1994). Moreover, education forms a significant part of human capital theory. As society realizes that technological advances demand new production methods, machinery, etc., and the investment expenditure for them; likewise, investment in humans is required to gain from these advances (Weisbrod, 1962). Since then, human capital research has come a long way in providing empirical evidence of its relevance in almost every allied field. Human capital research has found a strong foothold in the macro and micro frameworks of economic research. The application of human capital theory has been extended to numerous branches within economics and other multidisciplinary areas such as education, health, migration studies, gender, socio-economic, and politics, to name a few.

Goldin and Katz (2007) explained U.S. educational wage differentials owing to an increase in demand for educated workers and fluctuations in skill supply. They argued that the rise in college wage premiums between 1980 and 2005 could be attributed to the slowdown in growth relative to the supply of college workers, which started in 1980. Acemoglu (2012) reviewed the work of (Goldin & Katz, 2007) and found that human capital plays a vital role in explaining economic growth of nations. Autor et al. (2020)

attribute the changes in wage structure during the twentieth century in the United States to a combination of increased demand for educated workers (arising from skill-biased technological changes) and variations in the supply of skills from better access to education.

In the micro context of human capital theory, the returns to education are highly significant for two major reasons. Firstly, they shed light on an individual's investment in themselves and their private returns to education, including direct monetary benefits (Jones et al., 2023; Patrinos & Montenegro, 2022). Secondly, society has a keen interest in measuring the social returns to investment in education, which can manifest in various indirect benefits such as reduced crime rates, improved literacy, increased productivity, better governance and politics, and more (which are largely non-monetary in nature) (Johnson & Stafford, 1973; Moretti, 2005). Blanden (2020) finds that educational opportunities play an important role in improving an individual’s chances in life. Recent evidence on returns to education from across the world also suggests that investment in human capital could lead to sustainable growth and poverty reduction (Andrews et al., 2023). These arguments and discussions highlight the crucial role of education as an indicator of human capital within the framework of human capital theory.

2.6. CRITICISMS OF HUMAN CAPITAL THEORY

Human capital theory has been under scrutiny for decades in comparative and international education (Klees, 2016). Even before it began to draw attention from mainstream economics, Schultz (1961) mentions the debate regarding the moral aspect of considering humans as assets. Klees (2016) broadly classifies human capital theory on two grounds: 1) the relation between income and education as measured by labor productivity and 2) the link between education and economic growth. Critics of human capital theory raised concerns over moral issues, the economic effect of education, issues of strong ability bias (concerning private returns), family background effect on earnings, less attention given to social and indirect returns, and socioeconomic factors, to name a few (Teixeira, 2014).

Drawing on the fact that research in returns to education using years of education to explain changes in earnings or to explain the impact of education on GNP has shortcomings in methodological and specification choices, yielding widely varying results. Klees (2016) argued that these estimates are not helpful to policymakers. He further argued that human capital theory attributes issues like job creation, poverty elimination, and inequality reduction merely to lack of individual skills. Additionally, he also poses three alternative approaches to human capital, namely, human rights, human capabilities, and human agencies.

Correlations between income, education, and ability measures provide evidence that education may be used for screening ability differences (Riley, 1979). This leads us to the question: What is meant by the screening hypothesis? This becomes crucial here because human capital theory came under a whole new light of criticisms with this theory.

Cohn et al.. (1987) concluded that the screening hypothesis did not provide any evidence from United States data, irrespective of the approach used. Earlier, using microdata, Psacharopoulos (1979) tested the screening hypothesis and concluded that the data from the United Kingdom does not provide evidence for a strong screening version. These arguments present various criticisms of human capital theory.

However, decades of research on human capital theory have addressed various aspects of this criticism, such as methodology, specification, social returns and externalities, socioeconomic impacts of education, and the impact of family background on education, among others, into evolving as an integral part of economics in both the micro and macro frameworks (Blundell et al., 2001; Card, 1993, 1999; Card & Krueger, 1992; Griliches, 1977; Heckman, 1979; Heckman et al., 2003; Johnson & Stafford, 1973; Lemieux, 2006; Teixeira, 2014).

2.7. ESTIMATES OF RETURNS TO EDUCATION ACROSS THE WORLD

2.7.1. Cross-Country Estimates

The impact of human capital on growth has been broadly estimated using either cross-country regressions or growth accounting methods. The methodological differences in estimation make it challenging to compare results. Since this thesis focuses on the role of human capital, we limit ourselves to discussing findings from cross-country regression. Cross-country estimation uses a measure of human capital to estimate differences in economic performances across nations. The difference in choice of the dependent variables (economic performance) and explanatory variables (human capital indicators) vary across studies, making comparisons seemingly troublesome (Sianesi & Reenen, 2003).

This section highlights some select results from cross-country regressions. Psacharopoulos (1985) highlights the returns to an additional year of schooling for a Mincer-type function for 61 countries. The estimates were 13 percent (Africa, including Ethiopia, Kenya, Morocco, Tanzania;), 11 percent (Asia; including Hong Kong, Malaysia, Pakistan, Singapore, South Korea, South Vietnam, Sri Lanka, Taiwan, Thailand), 14 percent (Latin America; including Brazil, Chile, Colombia, Costa Rica, El Salvador, Guatemala, Mexico, Venezuela), 8 percent (Intermediate; including Cyprus, Greece, Iran, Portugal) and 9 percent (Advanced; including Australia, Canada, France, Germany, Japan, Sweden, United kingdom, United states).

Using average years of school attainment for males (25 years and above) at the secondary and higher levels, Barro (1999) estimated that the coefficient for additional years of schooling raises the growth rate by 7 percent (Overall). Based on 139 countries from 1950 to 2014, Psacharopoulos & Patrinos (2018) estimates that the private average to a year of schooling is 9 percent globally. For a period from 1970 to 2014, using 853 harmonized household surveys, Montenegro & Patrinos (2021) estimate returns to schooling covering 142 economies. They found that the Mincerian equation is stable in estimation, shows higher returns to schooling for women, and returns exhibit decreasing trends over time. They observed the highest returns to tertiary education and the lowest to secondary.

In addition, Montenegro & Patrinos (2021) also found that average returns to schooling were highest for Latin America, the Caribbean region, and the Sub-Saharan Africa region, whereas for Asia was close to the world average, lower in high-income OECD economies and lowest for the non-OECD European & Middle East/North Africa group. These findings make it attractive to study the impact of education and schooling on earnings in the micro-framework, especially in India, where the literature is relatively scarce.

2.7.2. Private Returns To Education Across The World

Private returns to an additional year of education in the form of earnings have been researched extensively across nations. Leigh (2008) estimates the economic return to various levels of education using data from Australia. He finds high returns for high school attainment, which yields 30 percent returns per year depending on the adjustment for ability bias. In addition, vocational training boosts earnings, as does a bachelor’s degree (15 percent increase in annual earnings) and postgraduation. In Singapore, Sakellariou's (2003) analysis of Labor Force Survey data revealed that investing in education across various levels and types of schooling results in remarkably high private returns. Kikuchi's (2017) research in Japan suggests that pursuing an additional year of university education leads to a significant 9 percent increase in hourly wages. Those enrolled in university education experience an average effect of 17 percent per year compared to less than 2 percent for those not enrolled. Moreover, policy changes have a positive impact on university education.

In developed countries, average annual returns to education were estimated by Gunderson & Oreopolous (2020) to be about 10 percent, with a range of 5 to 15 percent across different contexts. They find that (1) returns rise over time, (2) returns were higher for females, degree or credential holders, general academic streams (compared to technical vocational streams), and professional fields (compared to social sciences and humanities), and (3) social returns were close to 10 percent as well. A panel study (Bartik & Hershbein, 2018) found the career earnings premium for individuals with a 4-year degree for persons was considerably lower for individuals from low-income backgrounds as opposed to individuals from higher-income backgrounds, career earnings for graduates with a bachelor's degree were 136 percent higher than those who stopped at high school for individuals with family income 1.85 times higher than the poverty level when in high school, whereas if the individuals family income in high school was 1.85 times lower than the poverty level then their career earnings were only 71 percent higher than high school graduates—suggesting that education contributed more to wage inequality.

Another study (Phan & Coxhead, 2013) highlighted that lower incentives to invest in education and higher levels of inequality in Vietnam resulted from continued interventions in capital and labor markets that favored state firms and workers, causing widening gaps in wages and skill premiums between them. In comparison, another study from Vietnam (Tran, 2023) finds that individuals with lower secondary schooling diplomas earn a higher wage income. Questioning lower increases in earnings dispersion and returns to education among some countries as compared to the USA and the UK; (Barrett et al., 1999) compare the distribution of earnings between 1987 and 1994 for Ireland to find larger growth in earnings dispersion contrary to their expectation and account for this by increasing returns.

Jones et al. (2023) studied returns to education in Mozambique from 1996–2015 and found private rates of return to education fell for lower schooling levels and were stable and rose for highest levels; returns were increasingly convex in non-agricultural as opposed to agricultural employment.

In the realm of global education, research on estimating private returns has addressed various demographic factors, such as gender, economic distribution, race, and ethnicity. These factors have been discussed in the relevant literature to identify and analyze trends in private returns across different groups. In China, Fan et al. (2015) found returns between 10 to 14 percent using longitudinal data and found a significant positive impact for only rural, male, and poorly educated workers. Additionally, they reported negligible returns for female workers. This suggests the presence of gender differences, implying the need to enhance women's capacity for “networking” and competitiveness. Contrarily, (2015) used three rounds of the China Urban Labor Survey (CULS; 2001, 2005, and 2010) to find that premium to education increased by about 2 to 3 percent, and education premium is higher for non-migrants and males. In Vietnam, men only require a lower secondary schooling diploma, while women need an upper secondary schooling diploma to earn a better income (Tran, 2023). Contrarily, in South Africa, women's returns were found to be higher than men's (Salisbury, 2016). Using the Household Labour Force Survey for 2017 from Turkey and applying education reform as an instrument, Patrinos et al. (2021) found that private rate of return (16 percent) for higher education and social return (10 percent) and that the returns to education for females were higher than males. Additionally, the study found that, for both men and women, returns to education rise with education level.

Tansel & Bodur (2012) discussed male wage inequality in Turkey between 1994 and 2002 using Mincer's earnings function with OLS and quantile regression. They find evidence for high male wage inequality in Turkey, declining at the lower end of the wage distribution and increasing at the top. A meta-analysis study by Abdullah et al. (2015) found that education reduces the income share among high earners and increases the share among lower earners. In Vietnam, the relationship between a lower secondary schooling diploma and wage income is absent for ethnic minorities, unlike the majority (Tran, 2023). This suggests that vulnerable groups, like women and ethnic minorities, could benefit from state support, and the benefits of education are crucial among the labor force. We next examine studies focused on India.

2.8. PRIVATE RETURNS TO EDUCATION IN INDIA

This section highlights research that has focused on India. Most studies in the Indian context have used cross-sectional datasets to estimate returns. Using national-level data from NSS for 1983 and 1993/94, Duraisamy (2002) estimated the returns to education by gender, age, and location. He found that returns to education increased up to the secondary level and then declined. He also provides evidence of gender and locational differences in the returns to education. (Vasudeva Dutta, 2006) estimated the return to education for adult male workers using national-level survey data for three cross-sectional periods. He uses a Mincerian wage function while addressing selection bias. He finds that the returns to education are significantly different: flat returns to education for casual workers and a positive and U-shaped relation for education levels for regular workers.

While these studies looked at returns across demographics, (Kingdon & Theopold, 2008) questions the theoretical ambiguity in the sign of economic returns to education when households face liquidity constraints. In acquiring education, they suggest that better economic incentives positively impact the education of girls and non-poor boys and have a negative impact on poor boys. (Schündeln & Playforth, 2014) studied government sector employees in an attempt to explain the "micro-macro paradox" in the literature linking education and growth. They attribute high returns to education at the micro level and, on the contrary, the smaller or negative coefficients on education at the macro level (growth regressions) to highly educated individuals preferring privately rewarding jobs in sectors with lower social returns. Gender differences in returns to education with consumption data are explored in (Fulford, 2014) for India. Aggregating across age cohorts and states, he finds that an extra year of education results in a 4 percent increase in consumption for males in comparison to no additional consumption for females.

In a recent study by Serneels and Dercon (2021), the significance of a mother's aspirations in determining a child's educational outcomes is explored. The authors suggest that aspirations serve as a means of intergenerational mobility and that accounting for them can lead to faster improvements in educational outcomes. Similarly, Kundu's (2023) analysis of NSS data from 2004 to 2012 in India reveals that while the inequality of educational opportunity for Indian children is low if the beginning of schooling is considered, it sharply rises if timely finishing of elementary school is not achieved. The study highlights the importance of parental education, particularly that of mothers, in contributing to the inequality of educational opportunity.

The authors recommend that school education policies be revamped to reduce the dropout rate among vulnerable children.

Higher educational attainment can lead to higher earnings and pave the way for social mobility, particularly for individuals who face historical barriers based on gender, economic distribution, caste, or religion (Sen, 2000, 2001). In the Indian context, we highlight some significant studies that have estimated private returns across socioeconomic groups.

(Chamarbagwala, 2008) found evidence that for children in the age group 10–14 years, higher regional returns to primary education raise the probability of children studying, reducing the probability they will engage in labor for the top three quintiles of the income distribution. She, however, concludes that financial limitations hinder poorer households' opportunity to benefit from education. . Using a quantile regression approach, Agrawal (2011) reported positive returns to education for all quantiles across the wage distribution.

Gender has been an important area of research in wage determination. Gender differences in returns to education with consumption data provide evidence that an extra year of education brings only male cohorts roughly 4 percent more consumption with no higher consumption to females (Fulford, 2014).

High levels of wage discrimination for women in the labor market in Urban India do exist. However, education contributes little (Kingdon & Unni, 2010). Moreover, (Vatta et al., 2016) suggested that the returns to education have not changed much in rural and urban areas over time for female workers despite higher educational levels. The likelihood of women's labor force participation increases beyond compulsory secondary schooling; it is then crucial to provide higher education opportunities for women, including technical education and vocational training (since these are positive and significant) (Kanjilal-Bhaduri & Pastore, 2018). This is in addition to studies such as (Kingdon & Unni 2010), which suggest that the lower years of education for women are offset by the higher returns to education for women than men, which could enhance women's labor force participation and better returns to education leading to reducing gender wage gap in India. Similar findings are reported by (Hussain & Mukhopadhyay, 2023). They found that private returns to education were higher for women. Significant findings are reported (Mohapatra & Luckert, 2014) that have policy implications concerning reducing gender-wage inequalities.

Throughout history, the interplay between caste and religion has been a significant factor contributing to the unequal distribution of labor market opportunities and earnings. Even today, these dynamics continue to shape the economic landscape, perpetuating disparities and hindering progress toward greater equity and inclusion. Individuals from vulnerable sections (including lower social groups) have significantly lower incomes and job opportunities (Bhandari & Bordoloi, 2006). A study by (Agrawal, 2011) reported that disadvantaged social groups earn lower wages. Moreover, (Mitra, 2019) reported the lowest returns for Scheduled Tribes and Scheduled Castes as opposed to an unreserved category, with larger differences in rates of return for the private sector.

A caste-wise disaggregation for women is provided by (Datta et al., 2020) using the Periodic Labour Force Survey (2017-18) to find lower work participation rates for women from upper castes and increases in work participation rates moving to lower castes. However, among women who study beyond graduation, these gaps in participation narrow down, mainly in urban areas. They illustrate that caste and education interaction lend a crucial understanding of women's work participation.

Additionally, they argue that mere access to higher education does not reduce gaps between social groups concerning women's participation rates.

There are mixed results from India on the role of Caste in return to education. Bhattacharya and Sato (2017) found that caste, tribe, and religion do not significantly affect real wage rates in India. However, using data from a household-level survey by the National Sample Survey (NSS), Hnatkovska et al. (2012) found convergence in education, occupation distribution, wages, and consumption levels of SC/STs toward non-SC/ST levels. They attribute this wage and consumption convergence mainly to improvements in education.

The effectiveness of major educational policy interventions in India was studied by Varughese & Bairagya (2020) with a focus on education inequalities. Their research brought to light the persistent nature of educational inequalities across gender, place of residence, and social groups, with socio-economic factors having a strong influence on educational attainment. The study revealed that affirmative policy is still necessary for marginalized groups to realize equal educational opportunities due to their more vulnerable position. A left-skewed earnings distribution is reported for women and Muslims in comparison to males and upper-caste Hindus, respectively (Balasubramanian, 2020). Additionally, Varughese & Bairagya (2020) found that social groups such as ST Hindus, SC Hindus, and Muslims report lower educational attainment when compared to mainstream Hindus (inclusive of OBC).

2.9. EDUCATION, EARNINGS, AND NATURAL DISASTERS

Natural disasters have varied direct and indirect effects on human well-being (Benson & Clay, 2004; Field et al., 2012; Khan et al., 2020), some of which have immediate consequences as well as long-term effects (Baez et al., 2013). These include impacts on mental health (Makwana, 2019), mortality and morbidity (Bourque et al., 2007), economic growth (Cavallo et al., 2013), economic status and consumption (Bui et al., 2014), among others. Some have found that only catastrophic natural disasters negatively affect output growth in the short or long run (Cavallo et al., 2013).

The constantly evolving international literature has examined the causal relationship between natural disasters and educational attainment (Cuaresma, 2010; Drzewiecki et al., 2020). One of the long-term effects is the impact on human capital, specifically education (Baez, De la Fuente, and Santos, 2010; Kousky, 2016). Earlier studies have examined the impact of natural disasters on school enrolment (Ezaki, 2018) and years of schooling as an indicator of human capital (González et al., 2021). Natural disasters could have inter-generational effects, leading to lower human capital accumulation of such families across generations (Caruso, 2017). Given the mixed evidence, Cuaresma (2010) has argued that the evidence of natural disasters and Education is mixed and inconclusive.

Regions that experience more frequent occurrences of natural disasters could see higher investment for adaptation and mitigation, including relief (Ferreira et al., 2013; Solecki et al., 2011). This could result in better school infrastructure (which is used during disasters as a relief center) (Mutch, 2018; Paik, 2018; Pascapurnama et al., 2018) and better educational attainment in the long run (Skidmore & Toya, 2002) as well as lower migration (Dasgupta et al., 2022). On the other hand, areas with high human capital could result in lower vulnerability among households exposed to natural disasters (Drzewiecki et al., 2020). This could be due to better awareness and access to information, leading to minimizing losses from natural disasters (Parida et al., 2021). More developed countries will likely deal with disasters better and experience fewer losses (Toya & Skidmore, 2007).

The research in this area has either focussed on the role of educational attainment in reducing vulnerability to natural disasters (Fuchs & Thaler, 2018; Muttarak & Lutz, 2014) or the impact of natural disasters on educational attainment (Baez et al., 2010; De Vreyer et al., 2015). We follow the latter by estimating the impact of natural disasters on educational attainment in India.

Most studies are region or disaster-specific, like floods (Thamtanajit, 2020), droughts (Garbero & Muttarak, 2013), and earthquakes (Di Pietro, 2018; Paudel & Ryu, 2018), among others. Some studies based on cross-section data (Joshi, 2019) have the limitation that they do not observe the same households over time. Other studies have been region-specific (Suar & Kar, 2005) and lack the coverage of a national-level study. All these studies fail to account for the impact on Education based on the frequency of natural disasters at a country level using household longitudinal data, especially in large countries like India. Our study bridges this research gap by incorporating the severity of natural disasters in a longitudinal dataset covering all states and union territories in India.

India is highly vulnerable to climate change-induced natural disasters (Anees et al., 2020; BMTPC, 2006). According to some estimates, almost 85 percent of its geographical area is vulnerable to natural disasters (Paik, 2018). Natural disasters have been reported to negatively impact educational attainment in India (Parida et al., 2021).

Given that (a) India is considered to be a highly vulnerable country to extreme events in the context of climate change (BMTPC, 2006) and (b) large segments of the Indian population lack basic educational access (Jana, 2020), it is essential to understand how natural disasters have impacted education attainment. This would be important for policymakers and communities to mitigate and adapt to climate change.

2.10. CONCLUSION

It is clear from the literature review that human capital has evolved through decades of research on the microeconomic as well as the macroeconomic front. Firstly, though critics of human capital theory have time and again questioned the idea, there is continued use of the framework because it yields important policy insights. Issues pertaining to methodology, specification, social returns, and externalities, among others, have been addressed by the framework with growing sophistication.

The link between education and earnings was described in sections 2.1 to 2.9, and section 2.10 highlighted the impact of natural disasters on education and earnings. Since natural disasters are likely to increase in frequency and intensity with climate change, it is an important area of research. It has implications for quality education, gender equality, reduced Inequality, and climate change (SDG 4, 5, 10, 13, respectively).

In the next chapter (chapter 3), we discuss the methods (methodology, framework, and modeling) and materials (data and variables) used in the thesis. It discusses the details of the empirical estimation methods used. Additionally, we analyze preliminary results from cross-sectional and panel data sets. It provides a glimpse of the underlying issues we try to address in the subsequent chapters. Issues such as the gender wage gap, inequality in labor markets, differentials in private returns to investment in education across socio-economic groups, and the impact of natural disasters.

CHAPTER 3: DATA AND METHODOLOGY

In the previous chapter, we highlighted various contributions from researchers worldwide in the evolution of human capital theory. We discussed literature from the microeconomic and macroeconomic perspectives and methodological improvements in the estimation of private returns to education, differences in these private returns across gender, economic distributions, and social groups (caste and religion). Additionally, we also discussed the literature on natural disasters and educational attainment. We summarized findings from studies globally and in India.

In this chapter, we discuss dataset, variables, and the methodology, used in this thesis. This chapter is divided into 12 sections. Section 3.1 provides a brief introduction, 3.2 discusses the data and variables, 3.3 discusses the methods (Methodology, framework, and econometric specifications) for chapters 4 – 8 (returns to education), 3.9 discusses methods for Chapter 9 (natural disasters and educational outcomes), Section 3.10 discusses the classification of disaggregated variables (economic distributions, genders, caste and religion, 3.11 provides graphical summaries and comparison of means for key variables and 3.12 concludes the chapter.

3.1. INTRODUCTION

This chapter provides a detailed account of the data set, variables and methods used in each main chapter. It provides a background of available datasets in India, and describes in detail the Indian Human Development Survey (IHDS) data. We discuss the unique features of IHDS data that make it suitable for this thesis. The chapter also specifies the nature and empirical basis of the estimated models and econometric tools utilized in those models for each type of data used. We discuss our use of econometric techniques such as the Ordinary Least Squares Method (OLS), Heckman Selection Method, Generalized Least Squares Method (GLS), and other model specification tests like the Hausman test that we use to estimate and then verify the efficiency of the estimates derived from these models in the main chapters. The methods used in this thesis have a strong base in the literature; however, some techniques that earlier required manual step-wise evaluation are now computable in econometric software like Stata. One such estimation model used in this thesis is the Heckman model to account for selection bias, which we will discuss in section (3.3.2 b). We also provide classification for demographic factors such as economic distribution, gender, caste and religion. We use statistical tools to graphically visualize and t-test to compare means of key variables across sub-samples. Lastly, we discuss the use 3-dimensional graphical analysis to plot projections of derived estimates from our models that we use in chapters 5 – 8. Section 3.7 concludes this chapter.

3.2. DATA AND VARIABLES

3.2.1. Background

Currently, three nationwide panel data sets are available at the household level: IHDS, the PLFS, and CPHS. Over two waves at the household level, the IHDS panel data set has been used to study the impact of natural disasters (Tamuly & Mukhopadhyay, 2022) and rural insurance (Azam, 2018). It has also been used at the individual level as a repeated cross-section to study domestic violence (Akter & Chindarkar, 2020) and an individual panel to study women's health (Chatterjee & Sennott, 2021). Sarkar et al. (2019) also used a panel similar to our study but limited themselves to the determinants of women's participation in the labor force in India.

The government of India also publishes a new data set on employment and unemployment status, PLFS, which is collected every quarter (NSO, 2021). Although the data is collected every quarter, one household is interviewed only once a year, making it a rotational panel. This data has been used by researchers such as Sengupta (2023) to estimate the Mincer equation with a Heckman selection model to predict wages in India. The panel characteristic of the PLFS database has been used to study urban employment in India (Roy, Saroj, & Pradhan, 2022). However, it has not been exploited to respond to the question of wage determination.

The other panel data set currently available at the household level is the CPHS. However, this database focuses on tracking consumption (Abraham & Shrivastava, 2022), and, similar to the PLFS, it is too short to study medium or long-term impacts on wages attributable to education or experience.

3.2.2. Indian Human Development Survey (IHDS)

We use two waves of the Indian Human Development Survey (IHDS I and IHDS II, hereafter) that followed households over seven years (Desai & Vanneman, 2008, 2015). The first round (IHDS I) was conducted in 2004-05, and the second round (IHDS II) was conducted in 2011-12. IHDS I consist of 215,754 individuals, 41,554 households in 1,504 villages and 970 urban neighborhoods across India). The IHDS II comprises 204,565 individuals and 42,152 households from 1,503 villages and 971 urban neighborhoods throughout India). Of the nine data modules, namely individual, household, village, district, education(school), and women available in IHDS I and II, the thesis uses the individual level data to estimate and report results on average private returns to additional years of investment in human capital.

This data set emerged as a collaborative research program between the National Council of Applied Economic Research, New Delhi, India, and the University of Maryland, United States (Desai, Vanneman & National Council of Applied Economic Research 2005, 2015). Certain features of the IHDS data set make it unique compared to the commonly used data sets in Indian literature. First, topics in IHDS extend across indicators of caste, community, consumption, the standard of living, energy use, income, agriculture, employment, government subsidies, education, social and cultural capital, household and family structure, marriage, gender relations, fertility, health, village, infrastructure, among others. Second, the documented human development indicators are extensive, and the data set contains various contextual measures. Third, the panel components allow for richer inferences between periods.

Lastly, in context to Chapter 9, IHDS is the only national survey that provides longitudinal information on natural disasters and a wide range of socioeconomic characteristics with individual, household, and village-level identifiers. This survey provides information on natural disasters in India at the village level (rural) from 2006 to 2012. It also provides data on children's educational attainment (among other socioeconomic variables) at the household level, making this kind of study amenable.

Next, we provided a detailed description of merging IHDS I and II in constructing the panel data set.

3.2.3. Construction of Panel Data

We first downloaded the Stata-formatted files from the IHDS repository, which are free for researchers after registration (Desai, Vanneman & National Council of Applied Economic Research 2005, 2015). We merged the individual records from the two rounds of IHDS and arrived at 150,998 individuals. Only individuals who had been interviewed in both rounds were retained for observation.

The method to identify and merge these files is available from the IHDS repository. Upon completing the merger, we cross-checked the number of observations with IHDS documents to ensure that we had the same numbers derived from the master notes. The merged data set provided a balanced individual-level panel. A detailed account of the individual-level merging is given below.

The cross-sectional data from IHDS I and IHDS II are used to form longitudinal data (panel data) to analyze the effects of human capital investment on individual earnings within individuals and between two time periods. Data merging follows the guidelines from the IHDS instruction presentation to merge both rounds. The following steps were observed in merging the two rounds from IHDS: We use the IHDS II data file with a link containing identification markers for state, district, PSU, household, household split, and persons. We then sort and merge using STATEID DISTID PSUID HHID HHSPLITID PERSONID in the link file to the IHDS II data file with the merge 1:1 command on state. The resulting output is summarised in Table 3.1, which follows.

Table 3. 1: Result of preliminary merging of data with linking files

Result	Number of Observations
not matched	1
from master	0 (_merge==1)
from using	1 (_merge==2)
matched	204,568 (_merge==3)

Source: Authors calculation from IHDS I and II

Following this, we rename the estimate generated for a merging variable as 'merge_.' The next step was to rename all the variables in this data file for 2012 with the suffix '_2012'. Once this was achieved, we saved the file with a suitable name. The next order of business was to use the IHDS I data file and rename the identification markers to match the ones we had in the 2012 linked file. Once the renaming is achieved, sort the dataset using all the identification makers simultaneously.

The next and final step in the merging of the two rounds of the IHDS data for 2004-05 and 2011-12 is using Stata's merge 1: many command using STATEID DISTID PSUID HHID2005 HHSPLITID2005 PERSONID2005 within the 2004-05 file with the linked 2011-12 file for which the results are presented in table 3.2. that follows.

Table 3. 2: Result of final merging of IHDS I & II

Result	**Number of Observations**
not matched	118,347
from master	64,766 (_merge==1)
from using	53,581 (_merge==2)
matched	150,988 (_merge==3)

Source: Authors calculation from IHDS I and II

The result reported above (table 3.2) indicates that 150,988 individuals are common to both rounds; in other words, these individuals were interviewed in both rounds of the IHDS, namely IHDS I and IHDS II. The analysis covered in chapters 4, 5, 6, 7, and 8 of this thesis is restricted to these 150,988 individuals represented in both rounds, hence allowing us to stay focused on the interpretations we derive from our analysis. After that, we merged the household-level dataset from the two rounds to obtain a balanced household-level panel. Next, we linked the individual and household characteristics by merging the individual panel with the household panel, which provided us with an extensive panel data set for a comprehensive analysis.

3.2.4. Variables

Estimation of private returns to education in India

The dependent variable we use is the natural logarithm of hourly earnings (lnHourly_Earnings) reported by each individual in the outcome equations, in keeping with the literature on this subject (Eisnecker & Adriaans, 2023; Heckman & Polachek, 1974; Hoffmann & Kassouf, 2005). The dependent variable in the selection equations is a binary variable that indicates whether a person is in the labor force (Labour Force Participation). We constructed this variable based on whether a person reported an earning. Among the covariates, we have used variables that are commonly used in the literature for predicting earnings: the individual's education in completed years (Education), potential experience (Experience), and the square of potential experience (Experience_square).

We have used family characteristics like the number of children (Nchildren) (Mroz, 1987), marital status (Married) (Comola & de Mello, 2013), and the number of adults in a household (Nadults) (Ashraf, Weil, & Wilde, 2013; Kugler & Kumar, 2017), among others. In India, women are anticipated to exit the labor market when they become mothers (Jaumotte, 2003). This can be partly attributed to the patriarchal structure of Indian society, as the main burden of child-rearing falls on the mother (Bhambhani & Inbanathan, 2018). Therefore, variable Nchildern adversely influences women's presence in the labor market and the nature of their employment (Chun & Oh, 2002).

The family's economic status is often reflected in their consumption expenditures. We used household consumption expenditure to classify households below or above poverty. To create consumption quintiles, the entire sample was divided into two broad groups by level of consumption: those below the poverty line (BPL) and those above it

(APL). We further disaggregated the APL group into five quintiles (labeled APL1 (low) to APL5 (high)).

Gender identities have been widely acknowledged as a determinant of employment and earnings worldwide (Blau, Gielen, & Zimmermann, 2012), and studies show that a gender wage gap exists in the labor market (Sengupta & Das, 2014). We have included the variable *Female* (binary values with Female = 1, Male = 0) to examine any gap in returns to education.

Social groups in India are broadly classified along two identities—caste and religion. The role of caste and religion in determining earnings has been examined before (Arabsheibani, Gupta, Mishra, & Parhi, 2018; Bhaumik & Chakrabarty, 2010; Rani, 2013). India also has a unique social hierarchical system based on caste (Deshpande, 2011), which has well-known implications for economic outcomes (Bhaumik & Chakrabarty, 2008; Lolayekar & Mukhopadhyay, 2020). The hierarchy places the general castes at the top, followed by the other backward castes (OBC) and scheduled castes (SC). The scheduled tribes (ST) are strictly not part of the caste system. Rather, they are indigenous groups and are often ranked at the bottom of the social hierarchy (GoI, 2016; Mosse, 2018). There have been many social movements to abolish social hierarchies based on caste (Ray & Katzenstein, 2005). The state policy has also undertaken affirmative action by providing educational and employment reservations for lower castes (Deshpande, 2013). However, there is evidence that caste hierarchies still exist (Arabsheibani, Gupta, Mishra, & Parhi, 2018).

In IHDS I, there were five caste categories – namely Brahmin, OBC, SC, ST and others. In IHDS II, there were six caste categories – namely Brahmin, Forward /General (except Brahmin), OBC, SC, ST, and others. To make the categorization uniform across the two rounds and to match the government classification for employment reservations, we have classified the caste data into four categories, namely General (Forward, Brahmin, and Others together), OBC, SC, and ST. We created categorical groups as a variable *Caste* (General = 1, OBC= 2, SC=3, ST = 4).

Religion has been identified as a determinant of economic outcomes (Barro & McCleary, 2003). It is recognized as a site for determining social and economic hierarchy. Studies have found significant differences in development indicators based on religious groupings (Rani, 2013). A government report found that Muslims as a group have the lowest development indicators among all religious minority groups in India (GoI, 2006). In IHDS I and IHDS II, there were nine religion categories – namely Hindu, Muslim, Christian, Sikh, Buddhist, Jain, Tribal, Others, and None. To make the categorization tractable, we created three groups, namely Hindu, Muslim, and Other minorities (including all the remaining groups, Christian, Sikh, Buddhist, Jain, Tribal, Others, and None). The new variable, Religion, was categorical (Hindu = 1, Other minorities= 2, Muslims = 3).

Social capital can be understood as a form of investment in social relations with an expected return. Literature on social capital provides various definitions of social capital, such as the Trust definition, Excess cooperation, and Network definition, among others (Paldam, 2000). In this context, we adhere to the Network definition of social capital in this thesis. The network definition of social capital defines social capital as the social density of people's informal network. The expected returns from investment in social capital, as given by the network definition, can be summarised as a) facilitation of the flow of information, b) forms of influence, c) perception of social credentials, and d) reinforcement of identity and recognition (Lin, 2017). Social networking has been found to influence individual earnings in China (Liu, 2017), the

United States (Smith, 2000), Sweden (Behtoui & Neergaard, 2010) as well as India (Deshpande & Khanna, 2021). We added an indicator of social networking to the outcome equation (Networking Intensity). We constructed this variable by utilizing information regarding a household's membership in one or more social or economic networking groups (membership in co-operatives, caste or religious groups, and self-help groups, among others) (Dasgupta, 2005; Deshpande & Khanna, 2021).

In keeping with the literature, we constructed a variable to measure potential experience since the IHDS surveyed the experience of individuals directly (Lemieux, 2006). This variable is equal to the respondent's age minus the years of education minus five years (to account for preschool years as relevant in the Indian context). We have capped the maximum experience at 45 years (the gap between the entry and exit age group for labor force participation). It is debatable whether such an assumption is valid. We acknowledge that this assumption may not be universally valid, but it could reflect the labor market situation under specific circumstances.

We contend that during the years the two rounds of the IHDS surveys were conducted, the Indian economy experienced high growth, and the labor market underwent informalisation. The data from the NSSO's quinquennial rounds (2004–2005 and 2011–2012) suggests that the prevailing unemployment rates in the overall population were 2.2% and 2.3% (Mehrotra, 2019). Therefore, even if there was underemployment in the economy, the actual percentage of unemployment recorded was relatively low.

Impact of natural disasters on education in India

The dependent variable (*Education*) is the number of years of completed education of an individual. A continuous treatment of Natural Disaster Intensity (NDI) is expected to influence this. The variable *Natural Disaster Intensity (NDI)* is a constructed variable aggregating various natural disasters and the frequency of their occurrences between 2006 and 2012. We include other controls such as the number of children in the household (*Number of children in the household*), school fees paid by the household (*School Fees)*, and the type of educational institutions: *Government* schools, *Aided* schools, and *Private* schools. The reference category (base level) included all other types of educational institutions (madrasa, convent, junior college, college, post-graduate, among others).

We also added constructed variables such as *Membership Intensity* and *Confidence Intensity. Membership Intensity* is a constructed variable depicting the household's intensity in networking (a form of social capital). We constructed this variable by utilizing information regarding household membership in one or more social or economic membership groups like membership in co-operatives, caste or religious groups, self-help groups, etc.

The variable *Confidence Intensity* is a constructed variable depicting the household's confidence level in public institutions such as law enforcement, hospitals, banks, etc. Individuals' perceptions regarding their level of trust in public institutions were aggregated to form the variable 'Confidence Intensity.' We also use binary indicators for *Life insurance* and *Health insurance* as controls for the financial resilience of the household. We include a control *Repeated* to account for any individual repeating the same class.

3.3. METHODOLOGY AND FRAMEWORK

3.3.1. Cross-Sectional Analysis

The foundation of the methodology used in this thesis lies with Mincer's earnings function (Mincer, 1974). The earning function essentially consists of a semi-log function of earnings that depends on education (years of education) and experience (labor market experience). The model also accounts for the nonlinear relationship of experience with earnings by including a quadratic term for experience.

The methodology used in this thesis includes descriptive statistics, comparison of mean across groups, regression model estimation, and specification tests. These are discussed for two rounds of cross-sectional data, Indian Human Development Survey: IHDS I (2004-05) and IHDS II (2011-12).

3.3.2. Statistical Tools For Data Description

Data description is an essential part of preliminary analysis. Using large datasets requires researchers to summarize and aggregate to understand central tendencies, dispersion, and skewness within the used variables. A wide range of econometric models exist in establishing causal relations and estimation. However, each model makes crucial assumptions on the underlying distribution from which the data was obtained. It is then imperative that we first study and describe the data set that would be used in further analysis. We use descriptive statistics at the beginning of each cross-sectional analysis. The descriptive statistics include the number of observations for each variable, the mean (as a measure of central tendency), the standard deviation (a measure of dispersion), and the range (minimum and maximum values). Next, we discuss the t-test used to hypothesize and test differences in variables of interest.

3.3.3. Tests For Difference Of Means Across Groups

Hypothesis testing tools form an integral part of inferential statistics. Broadly, hypothesis testing methods can be classified as parametric and non-parametric (Harwell, 1988; Keller, 2018; Sheskin, 2003). The basis of this lies in the knowledge of the underlying distribution. These tools provide a gateway to analyzing differences across central tendencies and dispersions and proportions, among others; in their simplest use, they help hypothesize the significance of estimates in more complex model frameworks and determine the importance of overall models.

We use a simple two-sample t-test to hypothesize and test differences in the mean value of earnings across groups. In addition, these tests are also used in our estimated models like OLS. Next, we discuss another preliminary form of analysis we used: correlation analysis.

3.3.4. Use Of Software

Throughout the process of completing and compiling this thesis, we utilized a variety of software tools. Our preliminary and main data analyses were conducted using Stata 16.1, while we employed R for 3-dimensional graphical projections. We also utilized spreadsheets for compiling and formatting our results, and Grammarly for language editing purposes.

3.3.5. Regression Analysis

A correlation analysis does not suffice to establish causation between variables. This task is most commonly performed using regression methods. Regression methods help study the impact of the independent variable(s) on the dependent variable. In the sections to follow (3.3.2), we discuss the econometric models of regression estimation using ordinary least squares and Heckman's selection corrected method.

3.4. FRAMEWORK FOR SPECIFYING MINCER'S EARNINGS FUNCTION

The Mincer earnings function explains changes in an individual's hourly earnings as a function of schooling and experience (Mincer, 1974). It consists of a single regression equation in semi-log form. The model uses schooling, experience, and the square of experience as explanatory variables to explain changes in the natural logarithm of individuals' hourly earnings. Over the years, researchers have tried many variations in Mincer's original model specification to suit different data types.

Mincer's Earnings Function can be conceived of as a relational model:

$$Ln\ Y = f\,(S, Exp, Exp^2) \qquad (3.1)$$

Where:

LnY = Natural Log of Hourly Earnings

S = Years of Schooling

Exp = Potential Experience

Exp^2 = Square of Potential Experience

The framework of the Mincer function has been extensively used in estimating returns to education for over half a century. The ordinary least squares is the most widely used econometric method for estimating the Mincer's earnings function. As such, in the next section, we specify the econometric model for Mincer's earnings function and discuss its estimation using ordinary least squares.

3.4.1. Estimating The Econometric Model For Mincer's Earnings Function Using Ordinary Least Squares

The econometric model specification used for Mincer's earnings function can then be specified as follows:

$$LnY = a_0 + a_1 S + a_2 Exp + a_3 Exp^2 + u \qquad (3.2)$$

Where:

LnY = Natural Log of Hourly Earnings

S = Years of Schooling

Exp = Potential Experience

Exp^2 = Square of Potential Experience

u = Error term

Researchers have developed extended versions of this model by including polynomial terms in education (quadratic, cubic) and experience (cubic and quartic) to test its validity for various national and international data. We then specify our extended model for estimating private returns to education, experience, and social capital.

We estimate an extended Mincer Earnings function by adding an explanatory variable, namely social capital (measured by Networking Intensity). Networking Intensity is a constructed variable depicting the household's membership in various community social groups and organizations, like NGOs, religious groups, caste groups, etc.

The extended Mincer's earnings function can then be stated as follows:

$$\boldsymbol{Ln\,Y_i = \beta_0 + \beta_1 Educ_i + \beta_2 Exp_i + \beta_3 Exp_i^2 +}$$
$$\boldsymbol{\beta_4 Networking\ Intensity_i + u_i} \qquad \textit{\textbf{(3.3)}}$$

Where:

LnY = Natural Log of Hourly Earnings

Educ = Years of Schooling

Exp = Potential Experience

Exp^2 = Square of Potential Experience

Networking Intensity = Proxy for social capital

u = Error term

The returns for investment in additional years of education and experience would then be evaluated as the coefficients estimated in the OLS model. The coefficient $\boldsymbol{\beta_1}$ would be an individual's private returns to an additional year of education. The coefficient $\boldsymbol{\beta_2}$ would be an individual's private returns to an additional year of potential labor market experience. The coefficient $\boldsymbol{\beta_3}$ on the square of potential experience would help establish whether the relationship is linear or nonlinear. The coefficient of Networking Intensity $\boldsymbol{\beta_4}$ will show if any significance of social capital exists in determining earnings.

However, literature on the estimation of private returns to education in the form of earnings suggests that such estimation using OLS may suffer from at least two sources of endogeneity. Namely, selection bias and Ability bias. These issues and the solutions for these issues were discussed in the literature review section (section 2.4). While the methodology to address selection bias in the earnings model is well established, the methods of addressing Ability bias have limitations. This thesis covers our analysis by addressing selection bias using Heckman's selection model. We discuss our modified equation and the estimation models in the next section.

3.4.2. Estimating Econometric Model For Mincer's Earnings Function Using Heckman Selection Model

Though the Mincerian earnings function (J. A. Mincer, 1974) is still perhaps the used model for such estimation, the estimation method using ordinary least squares is criticized. The 1970s and 80s saw testing of the validity of the econometric specification of Mincer's earnings function. Traditional models using Mincer's earnings function with ordinary least squares methods have been reported to suffer unobserved individual heterogeneity (Flabbi & Gatti, 2018) in the form of both selection and endogeneity bias (ability bias) (Heckman, Lochner, & Todd, 2006; Wooldridge, 2018). Due to the nature of censored data (J. J. Heckman, 1976), researchers have established that using OLS to estimate such an earnings function would overestimate the coefficients of the regression model. Hence, it would give a misguided representation of the returns to education.

The suggested remedy for selection bias is using a Heckman selection model. This model consists of two equations, namely a main equation (outcome equation) of the regression of earnings and its explanatory variables and a selection equation that runs a probit estimation using a dummy variable to establish whether a particular individual's earnings are observed or not (participation in the labor force) against a set of instrumental variables, that determine the likelihood of labor force participation.

The Heckman selection model (Heckman, 1976, 1979), can be estimated using two distinct estimators.

The first specification of Heckman selection model uses a maximum likelihood estimator (henceforth MLE) (Heckman,1976). This estimator assumes an underlying regression relationships between a dependent 'y' and an independent 'x' variable, however the dependent variable is not always observed. Whether the dependent variable is observed or not for a unit (i = 1, 2, ….N) may be determined by a set of factors (instruments 'z' in selection equation) that influence it. This deterministic relationship can then be used as a first stage equation (selection equation). 'y' for unit 'i' is observed if the selection equation condition is fulfilled.

$$\boldsymbol{y_i = \alpha x_i + u_{1i}} \qquad \textbf{(outcome equation) (3.4)}$$

$$\boldsymbol{\gamma z_i + u_{2i} > 0} \qquad \textbf{(selection equation) (3.5)}$$

$$\boldsymbol{u_{1i} \sim N(0, \sigma)}$$

$$\boldsymbol{u_{2i} \sim N(0, 1)}$$

$$\boldsymbol{corr\ (u_1 u_2) = \rho}$$

Where,

y = Dependent variable

x = Independent variable

z = Set of instrument variables determining if 'y' is observed or not

u_1 and u_2 = Error term in outcome and selection equations respectively

The second, option is the estimation of Heckman selection model using Heckman's two-step consistent estimator provided by Heckman (1979) (henceforth Heckman two-step consistent estimator). This method uses Heckman's (1979) procedure. It involves the estimation of the Probit estimates for the selection equation (equation 3.6).

$$\mathbf{Pr}(\boldsymbol{y_i\ observed}\mid \boldsymbol{z_i}) = \boldsymbol{\phi}(\boldsymbol{z_i}, \boldsymbol{\lambda}) \qquad \textbf{(Probit estimates) (3.6)}$$

These estimates are then used to compute 'Inverse of the mills ratio' (the non-selection hazard , henceforth IMR). The significance of IMR, indicates the presence of selection bias and the justification for the use of the selection model. The Heckman selection model (MLE) also generates $'\lambda'$ (Lambda) as an indicator of the presence of selection bias. The IMR integrated to into the outcome equation to form an augmented outcome regression equation (equation 3.7).

$$\boldsymbol{y_i = a\, x_i + a_m\, IMR_i + e_{1i}} \qquad \textbf{(augmented outcome equation) (3.7)}$$

where;

a_m = additional parameter estimate - coefficient of the variable 'IMR' (non-selection hazard)

We estimate both these options of Heckman selection model.

Our econometric model can then be specified as follows in equation (3.8) and (3.9).

Main Equation specification:

$$\boldsymbol{Ln\, Y_i = \beta_0 + \beta_1 Educ_i + \beta_2 Exp_i + \beta_3 Exp_i^2 +}$$
$$\boldsymbol{\beta_4 Networking\ Intensity_i + \beta_5 IMR_i + \beta_6 Z_i + u_i} \qquad \textit{\textbf{(3.8)}}$$

Where:

LnY = Natural Log of Hourly Earnings

Educ = Years of Schooling

Exp = Potential Experience

Exp^2 = Square of Potential Experience

Networking Intensity = Proxy for social capital

IMR = Inverse Mills ratio (significance means the presence of selection bias)

Z = list of instruments from the selection equation

u = Error term

Selection Equation specification:

$$\boldsymbol{Labourforce\ Participation_i = a_0 + a_1 Married_i + a_2 NChildren_i + a_3 Nadults_i^2 + a_4 X_i + e_i} \qquad \textbf{\textit{(3.9)}}$$

Where:

Labourforce Participation = Dummy variable for Labourforce participation

Married = Dummy variable for marital status

Nchildren = Number of children in the Household

Nadults = Number of adults in the Household

X = list of covariates from the main equation

e = Error term

Equations (3.8) and (3.9) are the main and selection equations for our econometric specification, respectively, that are used in the Heckman selection models to be estimated in the following chapters. The Heckman selection model with two-step consistent estimates, additionally estimates a value for the Inverse Mills Ratio (IMR) to allow for testing the hypothesis of whether selection bias is an issue in the estimation. A significant value for IMR would imply the existence of selection Bias and, as such, justify the use of Heckman's selection model for estimation.

3.5. PANEL DATA ANALYSIS

Though suitable for many forms of analysis, the cross-sectional analysis discussed in the previous section suffers from limitations. Prior studies using pooled unit-level data will have the limitations that cross-section models encounter, such as omitted variable bias and dynamic relationships, among other issues (Hsiao, 2007). Using panel data is very helpful for policy analysis because it removes the limitations of cross-sectional and pooled data sets (Wooldridge, 2012). Since panel data refers to data for the same unit of observation over time, we can see differences between entities and time. The following sections discuss the various panel analysis tools used in this thesis.

3.5.1. Statistical Tool For Data Description

The methodology used to illustrate descriptive statistics at the beginning of panel analysis is similar to that used in the cross-sectional section. The descriptive statistics comprise the number of observations for each variable, the mean (which is a measure of central tendency), the standard deviation (which is a measure of dispersion), and the range (minimum and maximum values); however, the descriptive statistics for the panel data provide these measures for overall, in between and within. This allows us to see differences in these measures between and within the periods considered.

3.5.2. Panel Regression Analysis

We now proceed from the cross-sectional regression analysis to a panel regression analysis. We must establish the model specification for reporting estimates from the panel data. At this conjecture, the most crucial decision a researcher must make is to choose between two distinct model specifications: the random effects model of estimating panel data and the fixed effects model. The difference between the two lies in the assumptions regarding the unobserved effect. The random effects model assumes that these unobserved effects are independent of all explanatory variables in all periods (Wooldridge, 2012).

The usual empirical test to choose between these models is the Hausman test. A Chi-squared test is used to test significant differences in the estimated coefficients using the random and fixed effect model. Consequently, a significant p-value for the Chi-squared test indicates the suitability of a fixed effects model in estimation (Wooldridge, 2012).

Further, the literature review also indicates that the models represented above may suffer from selection bias due to unobservable or missing values for the earnings of individuals resulting from non-participation in the labor force (J. Heckman, 1990). Selecting only those individuals who reported earnings to estimate the returns to an additional year of education could overestimate the effect of education. This estimation error, using only the sub-sample of individuals who reported earnings, is called selection bias (Heckman, 1974; 1976; 1979). It has been well established that estimation from merely a least squares estimator suffers from selection and ability bias (Angrist & Krueger, 1991). (J. J. Heckman, 1979b). However, another problem researchers discuss in estimating returns to education for additional years of education encompasses the issue of endogeneity (J. D. Angrist & Krueger, 1999; Card, 1999; Griliches, 1977b). Individuals may embody innate abilities or genetic traits that aid productivity over and above additional years of education. The average private return to additional years of education may be an overestimation due to these unobservable traits. Children with greater ability are more likely to continue in school than others. The returns to education may also be overestimated in such cases (Patrinos, 2016).

These issues of specification and endogeneity bias have been the focus of more recent research in human capital investment and, specifically, on returns to education. Different methods have been proposed to tackle the problem of endogeneity. The most popular is the instrumental variable (IV) approach (Carneiro, Heckman, & Vytlacil, 2011). Instruments are variables that are strongly correlated to the endogenous covariate but not correlated to the outcome variable (Briggs, 2004). The challenge in empirical studies has been to find effective instruments (Angrist & Krueger, 2001; Baltagi & Khanti-Akom, 1990). The current recommendation is to use an instrument or treatment assigned randomly rather than an individual-related characteristic (Wooldridge, 2018).

In non-experimental data, the challenge of finding the right instrument has led researchers to abandon the IV method (Rossi, 2014; Truong, Ogawa, & Sanfo, 2021), as weak instruments lead to greater bias than other techniques (Bound, Jaeger, & Baker, 1995; Cruz & Moreira, 2005). For this reason, we have chosen not to use the IV method. Instead, we use the workhorse of such studies—the fixed-effects model with Heckman correction—to control for selection bias (Wooldridge 2018), with longitudinal data to estimate fixed-effects models (Kamhöfer & Schmitz, 2016; Leigh, 2008). This model controls for average individual differences in any observable or unobservable predictors. The within-group estimate explains the effect of covariates on an individual's earnings over time.

These issues, though seriously challenging the initial models presented in the data, are an issue of model specification. These issues are addressed in literature in different ways. The most common methods to address these biases involve manual evaluation and inclusion of a selection correction variable, 'Inverse Mills Ratio,' or an alternative estimation method widely known as the Heckman Two-Step model for selection Correction. The two differ only because the former involves the manual estimation of the 'Inverse Mills ratio' from a probit model, and the latter generates this inverse mills ratio as a two-step procedure. Hence, departing from the previously used OLS method, the Heckman-type model follows a selection of samples based on the probit estimates of the selection model specified.

3.6. FRAMEWORK OF PANEL DATA ANALYSIS

3.6.1. Econometric Model Specifications For Estimating The Earnings Function In The Panel Data Analysis

This section highlights the econometric model specifications for estimating the earnings function in the panel data analysis. The panel data analysis involves estimating an extended Mincer's earnings function for individuals, first using the random effects and fixed effects model in Stata 16.1 to determine the suitability of the data. The Hausman specification test compares the estimates of each of these models using its random effects and fixed effects. Once we establish the suitability of the random or fixed effects, we estimate the Heckman selection model with fixed effects for this panel data.

Fixed Effects model can be specified as:

$$\boldsymbol{Y_{it}} = \boldsymbol{\beta_1 X_{it}} + \boldsymbol{a_i} + \boldsymbol{e_{it}} \text{ ; i = 1, 2, …N and t = 1, 2, …T.} \qquad \textbf{\textit{(3.10)}}$$

Where $\boldsymbol{a_i}$ are the individual specific intercepts that capture heterogeneity, i is the observed units (individual level, in our data), and t is the time period (two-periods, in our data).

The econometric equation in this section is stated as:

$$\boldsymbol{Ln\ Hourly\ Earnings_{it}} = \boldsymbol{B_0} + \boldsymbol{\beta_1 Education_{it}} + \boldsymbol{\beta_2 Experience_{it}} + \boldsymbol{\beta_3 Experience^2_{it}} + \boldsymbol{\beta_4 Network\ Intensity_{it}} + \boldsymbol{u_{it}} \qquad \textbf{\textit{(3.11)}}$$

Where, $'\boldsymbol{B_0}'$ the intercept is inclusive of the individual specific intercepts. The overall intercept is actually the average of the individual-specific intercepts (Wooldridge, 2012).

In **equation (3.11)**, the dependent variable is *Ln_Hourly_Earnings* (Natural logarithm of Hourly earnings) of an individual. The explanatory variables are Education (Years of completed education), *Experience* (Potential Experience), *Experiencesqr* (Square of Potential Experience) and *Networking Intensity* (a proxy for social capital).

The model is similar to the one used in the cross-sectional analysis but incorporates time-specific and individual-specific effects. The subscripts 't' and 'i' represent the period and the individual, respectively.

3.6.2. Econometric Model Specification For Heckman Selection Model

The Heckman model for selection bias is panel data, which has been explored in the Indian context. Internationally, the method is used by researchers to model returns to additional years of education in the presence of selection bias (Heckman & Li, 2004). However, these studies have been limited in their ability to employ fixed effects in the model. The model presented in this section overcomes this limitation in previous studies while addressing selection bias to derive better estimates for returns to additional years of education.

Main Equation:

$$\begin{aligned}\boldsymbol{Ln\ Hourly\ Earnings_{it}} &= \boldsymbol{B_0 + \beta_1 Education_{it} + \beta_2 Experience_{it} + \beta_3 Experience_{it}^2} \\ &+ \boldsymbol{\beta_4 Networking\ Intensity_{it}} + \\ &\boldsymbol{\beta_5 IMR_{it} + \beta_6 Z_{it} + u_{it}} \qquad \textit{(3.12)}\end{aligned}$$

Selection Equation:

$$\begin{aligned}\boldsymbol{Labourforce\ Participation_{it}} &= \boldsymbol{a_0 + a_1 Married_{it} + a_2 Nchildren_{it} + a_3 Nadults_{it}^2} \\ &+ \boldsymbol{a_4 X_{it} + e_{it}} \qquad \textit{(3.13)}\end{aligned}$$

In the **main equation (3.12)**, the dependent variable in all the models is *Ln Hourly Earnings* (Natural logarithm of Hourly earnings of an individual). The explanatory variables are *Education* (Years of completed education), *Experience* (Potential Experience), *Experiencesqr* (Square of Potential Experience) and *Networking Intensity* (a proxy for social capital). The model also generates estimates for the year-specific inverse mills *ratio (IMR)* and instruments used in the selection equation.

In the **selection Equation (3.9)**, the dependent variable is *laborforce participation* (Dummy for participation in the labor force); the instruments used are Married (Dummy for marital status), Nchildren (Number of children in the household), and Nadults (Number of adults in the household). The selection equation also uses all the explanatory variables from the main equation and the interactions of variables from the explanatory variables and the selection instruments.

3.7. HAUSMAN TEST

The usual empirical test to choose between these models is the Hausman test. A Chi-squared test is used to test significant differences in the estimated coefficients using the random and fixed effect model. Consequently, a significant p-value for the Chi-squared test indicates the suitability of a fixed effects model in estimation. The Hausman test is follows a Chi-squared distribution and can be estimated as follows (equation 3.14):

$$\boldsymbol{H} = (\boldsymbol{\beta_c} - \boldsymbol{\beta_e})' \, (\boldsymbol{V_c} - \boldsymbol{V_e})^{-\mathbf{1}} \, (\boldsymbol{\beta_c} - \boldsymbol{\beta_e}) \qquad \textbf{(3.14)}$$

Where;

$$\beta_c = coefficient\ vector\ from\ consistent\ estimator$$

$$\beta_e = coefficient\ vector\ from\ efficient\ estimator$$

$$V_c = covariance\ matrix\ of\ the\ consistent\ estimator$$

$$V_e = covariance\ matrix\ of\ the\ efficient\ estimator$$

The models and test described above form the basis of our panel analysis, which we conduct and examine in Chapter 4 (overview of returns to education). The entire chapter is dedicated to understanding the framework and the derived estimates.

Using this framework, we establish our final model, which we then replicate in chapters 5 (across economic distributions), 6 (across gender), 7 (across caste), and 8 (across religion).

3.8. GRAPHICAL PROJECTIONS

In this section, we use estimates derived from the Heckman Selection model from panel data for education and experience to generate projections for earnings patterns of individuals by class (consumption quintile), gender, caste, and religion. In this attempt, we try to find evidence of convergence or divergence in return for an additional year of education and experience. We generate four graphs of projections of earnings: by class (chapter 5), by gender (chapter 6), by caste (chapter 7), and by religion (chapter 8). We believe these projections provide visual insight into the wide range of estimates derived from models.

3.9. METHODOLOGY FOR IMPACT OF NATURAL DISASTERS ON EDUCATION

3.9.1. Introduction

Studies have found that natural disasters disrupt welfare through widening inequalities, increased disparities, and long-term educational impacts (Baez et al., 2010; Butz et al., 2014). This impact on education can lead to human capital loss and adversely affect future growth and development. We use data from two of the latest rounds of the Indian Human Development Survey (IHDS) and employ a panel Difference in Differences (DiD) regression model with continuous treatment fixed effects at the individual level. This allows us to examine the impact of natural disasters on education outcomes between 2004-05 and 2011-12 at the individual level.

3.9.2. Methodology and Framework

We use a natural experimental study approach (Quasi-experimental design), using Natural Disasters as a continuous treatment to estimate the average treatment effect of natural disasters on Educational Attainment in India. A natural experiment is said to be a form of quasi-experimental design if an intervention where the researcher does not control the circumstances of implementation of the intervention itself (Leatherdale, 2019). Such approaches are most preferred in the absence of randomized controls in studies where researchers must evaluate the effectiveness of programs or interventions (Chen et al., 2020; González et al., 2021). (Tamuly & Mukhopadhyay, 2022) used this approach to estimate and examine the impact of natural disasters on the well-being (measured by consumption expenditure) of households in India. We follow (Tamuly & Mukhopadhyay, 2022) in constructing and using a continuous measure of the frequency of occurrence of natural disasters as *Natural Disaster Intensity (NDI)*, trust in public institutions as *Confidence Intensity*, membership in various socioeconomic groups as *Membership Intensity* (a measure of social capital) and other covariates used by them to estimate their impact on Educational attainment. In addition, we include *School fees*, *School Type (Government, Aided, Private)*, and *Repeated* (if an individual has ever repeated a class) as covariates to estimate their impact on Educational Attainment.

Multiple household and institutional characteristics determine educational attainment. These include household characteristics like household structure (proportion of children in the household) (Blake, 1989; Perkins, 2019), nature of the household indicated by the ownership of household assets (Cooper & Stewart, 2021; Filmer & Pritchett, 1999) and school type (Pianta & Ansari, 2018). However, a substantial variation in educational attainment remains unexplained after accounting for these. We believe these unexplained variations may partly be explained by observing the exposure of the villages to natural disasters. In our model, we incorporate an explanatory variable measuring the cumulative effect of natural disasters on households between 2006 and 2012 (*Natural Disaster Intensity, NDI*). These natural disasters are listed as floods, droughts, earthquakes, hailstorms, tsunamis, and droughts, among others. Past studies have recorded the role of gender (Arnold, 2015), economic status (Blanden, 2004), social groups (Borooah, 2012), religion, and ethnicity in determining educational attainment (Sander, 1992).

The structure of the difference-in-differences regression models can be illustrated as follows (equation 3.10):

$$\boldsymbol{Y_i = \beta_0 + \beta_1 Time_{Period_i} + \beta_2 Treated_i + \beta_3 Time_{Period_i} * Treated_i + \beta_4 Covariates_i + \varepsilon_i} \quad \textbf{\textit{(3.10)}}$$

In equation 3.10, Y_i is the observed value of the response variable, $Time\_Period_i$ is the binary time variable indicating pre or post-treatment, $Treated_i$ is the binary variable indicating the treatment, $Time\_Period_i * Trperiodeated_i$ is the interaction term between the time period and the treatment, $Covariates_i$ are a set of independent variables that possible affect the response variable, and ε_i is the error term.

The functional relationship that we wish to explore is as follows (equation 3.11):

Education = f (Natural Disasters Intensity (treatment), set of controls (Number of Children in Household, Number of Household Assets, Economic cost of education, Institutional type, Membership Intensity, Confidence Intensity, Financial resilience, Repetition of a class) (3.11)

In equation 3.11, Education is the dependent variable measured in an individual's years of completed education, which depends on Natural disaster intensity, Household characteristics (*Number of Children in Household, Number of Household Assets)*, Economic cost of Education (School Fees) and Institutional type (Government, Aided and Private), attainment *Membership Intensity* (construct to measure social capital), Confidence in public institutions (construct to measure the level of trust in public institutions), Financial resilience (binary indicators for Life insurance and Health Insurance), and Repetition of class (a binary indicator of having repeated a class)

Our objective is to estimate the impact of natural disasters (with differential intensity) on educational attainment in households in rural India. We provide estimates disaggregated by *Gender* (Male and Female), *Consumption quintiles* (Below the poverty line (BPL) and quintiles above the poverty line, from APL1 to APL5), Caste (General, Other backward castes, Scheduled castes, and Scheduled tribes) and *Religion* (Hindu, Other minorities and Muslim). This disaggregation, we believe, would provide more robust estimates to analyze the impact of natural disasters on educational attainment across sections of the sub-population.

3.9.3. Model Specification

The basic intuition behind the use of a difference-in-differences is to evaluate the differences in the causal effect of an event between a treatment group (those affected by the event) and a control group (those unaffected by the event) (Donald & Lang, 2007). In such a case, one could infer that if the event had never occurred, then differences between the treatment group and control group would have stayed the same. *Natural Disaster Intensity* serves as a treatment that is similar to 'natural program intervention.' In contrast to conventional DiD models that use a binary treatment variable, we use a continuous treatment effect. This enables us to capture the slope effect (incremental change due to unit increase in the intensity of the treatment, Natural Disasters) of the treatment in addition to the level effect. In other words, we could capture the marginal change in educational attainment for unit changes in the intensity of natural disasters.

The specific equation is stated as (equation 3.12):

$$\textbf{\textit{Education}}_{it} = \alpha_0 + \alpha_1 \textbf{\textit{Natural Disaster Intensity}}_{it} + \alpha_2 \textbf{\textit{Number of children in Household}}_{it} + \alpha_3 \textbf{\textit{Number of Household Assets}}_{it} + \alpha_4 \textbf{\textit{School Fees}}_{it} + \alpha_5 \textbf{\textit{Government School}}_{it} + \alpha_6 \textbf{\textit{Aided School}}_{it} + \alpha_7 \textbf{\textit{Private School}}_{it} + \alpha_8 \textbf{\textit{Membership Intensity}}_{it} + \alpha_9 \textbf{\textit{Confidence Intensity}}_{it} + \alpha_{10} \textbf{\textit{Life Insurance}}_{it} + \alpha_{11} \textbf{\textit{Health Insurance}}_{it} + \alpha_{12} \textbf{\textit{Repeated}}_{it} + \alpha_{13} \textbf{\textit{Year}}_{it} + \varepsilon_{it} \qquad \textbf{\textit{(3.12)}}$$

Where,

Education = Years of Education

Natural Disaster Intensity = Ordered rank variable (range lies between 0 to 18)

Number of children in household = Number of children in the household

Number of Household Assets = Number of household assets owned by the household

School Fees = School fees paid by households (in terms of 2011-12 values)

School Type = Type of educational institution (Others (1), Government (2), Aided (3), Private (4))

Membership Intensity = Ordered rank variable for household membership in one or more social or economic membership groups (range lies between 0 to 9)

Confidence Intensity = Ordered rank variable for a household's level of Confidence in public institutions (range lies between 10 to 30)

Life Insurance = Dummy variable for financial resilience (Yes (1), No (0)) of household

Health Insurance = Dummy variable for financial resilience (Yes (1), No (0)) of household

Repeated = Dummy variable for an individual repeating the same class (Yes (1), No (0))

Year = Year indicator

εit = Residual

We then provide the disaggregated estimates obtained from the above model by *Gender* (Male and Female), *Consumption quintiles* (Below the poverty line and quintiles above the poverty line), *Caste* (General, Other backward castes, Scheduled castes, and Scheduled tribes), and *Religion* (Hindu, Other minorities and Muslim).

3.10. CLASSIFICATION OF VARIABLES USED FOR DISAGGREGATION

3.10.1. Consumption Quintiles

Table 3. 3: Composition of Consumption Quintiles in IHDS I

Consumption Categories 2005	Frequency	Percent	Cumulative
Below Poverty	50,486	30.67	30.67
Above Poverty Q1 (APL1)	22,834	13.87	44.54
Above Poverty Q2 (APL2)	22,827	13.87	58.41
Above Poverty Q3 (APL3)	22,823	13.86	72.27
Above Poverty Q4 (APL4)	22,824	13.86	86.13
Above Poverty Q5 (APL5)	22,825	13.87	100.00
Total	164,619	100.00	

Source: Author's calculation using IHDS I Data

Table 3.3 provides the composition of frequency and percentage of the individuals belonging to different consumption quintiles in the data. In IHDS I, we find that of the total sample used for analysis, BPL consisted of 30.67 percent, APL1 consisted of 13.87

percent, APL2 consisted of 13.87 percent, APL3 consisted of 13.86 percent, APL4 consisted of 13.86 percent, and APL5 consisted of 13.87 percent. The APL1 to APL5 (Above poverty line quintiles) comprised 69.33 percent of observations.

Table 3. 4: Composition of Consumption Quintiles in IHDS II

Consumption Categories 2012	Frequency	Percent	Cumulative
Below Poverty	29,479	18.35	18.35
Above Poverty Q1 (APL1)	26,242	16.33	34.68
Above Poverty Q2 (APL2)	26,239	16.33	51.01
Above Poverty Q3 (APL3)	26,240	16.33	67.34
Above Poverty Q4 (APL4)	26,240	16.33	83.67
Above Poverty Q5 (APL5)	26,239	16.33	100.00
Total	160,679	100.00	

Source: Author's calculation using IHDS II Data

Table 3.3 provides the composition of frequency and percentage of the individuals belonging to different consumption quintiles in the data. In IHDS II, we find that of the total sample used for analysis, BPL consisted of 18.35 percent, APL1 consisted of 16.33 percent, APL2 consisted of 16.33 percent, APL3 consisted of 16.33 percent, APL4 consisted of 16.33 percent, and APL5 consisted of 16.33 percent. The APL1 to APL5 (Above poverty line quintiles) comprised 81.65 percent of observations.

3.10.2. Gender

Table 3. 5: Composition of Gender in IHDS I

Gender Categories 2005	Frequency	Percent	Cumulative
Male	109,805	50.89	50.89
Female	105,949	49.11	100.00
Total	**215,754**	**100.00**	

Source: Author's calculation using IHDS I Data

Table 3.5 provides the composition of Males and Females in the data. Males comprise 50.89 percent of the observations in the data, while Females comprise 49.11 percent.

Table 3. 6: Composition of Gender in IHDS II

Gender Categories 2012	Frequency	Percent	Cumulative
Male	102,062	49.89	49.89
Female	102,506	50.11	100.00
Total	**204,568**	**100.00**	

Source: Author's calculation using IHDS II Data

Table 3.6 provides the composition of Males and Females in the data. Males comprise 49.89 percent of the observations in the data, while Females comprise 50.11 percent.

3.10.3. Caste

Table 3. 7: Composition of Caste in IHDS I

Category	Frequency	Percent	Cumulative
Brahmin	9,997	5.55	5.55
Forward/General (except Brahmin)	71,880	39.87	45.41
Other Backward Castes (OBC)	36,094	20.02	65.44
Scheduled Castes (SC)	14,220	7.89	73.32
Scheduled Tribes (ST)	48,096	26.68	100
Total	1,80,287	100	

Source: Author's calculation using IHDS I Data

Table 3.7 provides the composition of frequency and percentage of the individuals belonging to different Religions in the Data. In IHDS I, we find that of the total sample used for analysis, Brahmin caste consisted of 5.55 percent, Forward/ General (except Brahmin) consisted of 39.87 percent, Other Backward castes consisted of 20.02 percent, Scheduled castes consisted of 7.89 percent and Scheduled tribes consisted of 26.68 percent.

Table 3. 8: Composition of Caste in IHDS II

Category	Frequency	Percent	Cumulative
Brahmin	10,253	5.02	5.02
Forward/General (except Brahmin)	46,862	22.96	27.98
Other Backward Castes (OBC)	83,929	41.12	69.1
Scheduled Castes (SC)	43,137	21.13	90.23
Scheduled Tribes (ST)	17,380	8.51	98.75
Others	2,553	1.25	100
Total	2,04,114	100	

Source: Author's calculation using IHDS II Data

Table 3.8 provides the composition of frequency and percentage of the individuals belonging to different Religions in the Data. In IHDS II, we find that of the total sample used for analysis, Brahmin caste consisted of 5.02 percent, Forward/ General (except Brahmin) consisted of 22.96 percent, Other Backward castes consisted of 41.12 percent, Scheduled castes consisted of 21.13 percent and Scheduled tribes consisted of 8.51 percent and other 1.25 percent.

In our analysis, we constructed four broad categories of caste categories for analysis since the two rounds of IHDS provided different classification norms. General/ Forward castes (including Brahmins and others) (1), Other Backward castes (2), Scheduled castes (3), and Scheduled tribes.

3.10.4. Religion

Table 3. 9: Composition of Religion in IHDS I

Religion	Frequency	Percent	Cumulative
Hindu	1,42,997	79.32	79.32
Muslim	23,914	13.26	92.58
Christian	4,984	2.76	95.35
Sikh	4,738	2.63	97.97
Buddhist	1,103	0.61	98.59
Jain	591	0.33	98.91
Tribal	1,835	1.02	99.93
Others	110	0.06	99.99
None	15	0.01	100
Total	180,287	100.00	

Source: Author's calculation using IHDS I Data

Table 3.9 provides the composition of frequency and percentage of the individuals belonging to different Religions in the Data. In IHDS I, we find that of the total sample used for analysis, Hindu consisted of 79.32 percent, Muslim consisted of 13.26 percent, Christian consisted of 2.76 percent, and all other groups combined consisted of 4.66 percent.

Table 3. 10: Composition of Religion in IHDS II

Religion	Frequency	Percent	Cumulative
Hindu	1,63,876	80.11	80.11
Muslim	27,792	13.59	93.69
Christian	5,272	2.58	96.27
Sikh	4,786	2.34	98.61
Buddhist	1,205	0.59	99.2
Jain	486	0.24	99.44
Tribal	914	0.45	99.88
Others	185	0.09	99.98
None	51	0.02	100
Total	2,04,567	100	

Source: Author's calculation using IHDS II Data

Table 3.10 provides the composition of frequency and percentage of the individuals belonging to different Religions in the Data. In IHDS II, we find that of the total sample used for analysis, Hindu consisted of 80.11 percent, Muslim consisted of 13.59 percent, Christian consisted of 2.58 percent, and all other groups combined consisted of 3.72 percent.

In our analysis, we constructed three broad categories of religious groups for analysis since the proportion of most minority groups was very small in comparison. Hindu (1), Other Minorities (2) including (Christian, Jain, Sikh, Jain, Buddhist, Tribals and others and None) and Muslims (3).

3.11. GRAPHICAL SUMMARIES AND COMPARISON OF THE MEAN OF KEY INDICATORS ACROSS GENDER AND SOCIO-ECONOMIC GROUPS

3.11.1. Indian Human Development Survey I

In this particular section of our study, we aim to provide a comprehensive analysis of the graphical illustrations and statistical tests depicting the differences in averages of the two critical variables - Hourly Earnings (Natural logarithm) and Educational Attainment (measured in years) - based on gender and socio-economic groups by utilizing IHDS I data. Our goal is to gain a better understanding of variations in these key variables based on socio-economic factors, which will allow us to provide disaggregated estimates for further analysis. By providing detailed insights into these differences, we hope to shed light on the key factors that contribute to such variations across different socio-economic groups and gender.

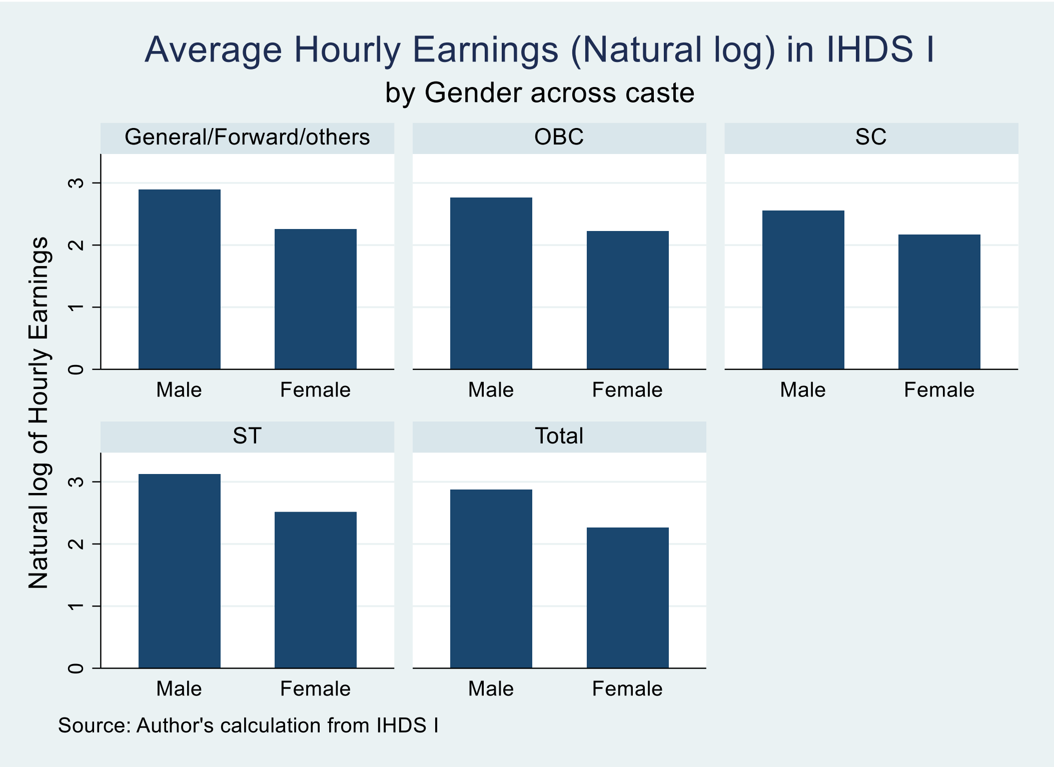

Figure 3. 1: Graphical illustration of Average Hourly Earnings in IHDS I by Gender across Castes

Figure 3.1 depicts the difference in average hourly earnings of men and women across caste categories. Gender differences in average hourly earnings are persistent across all caste categories. In other words, regardless of social status, women from across all castes tend to earn less than their male counterparts.

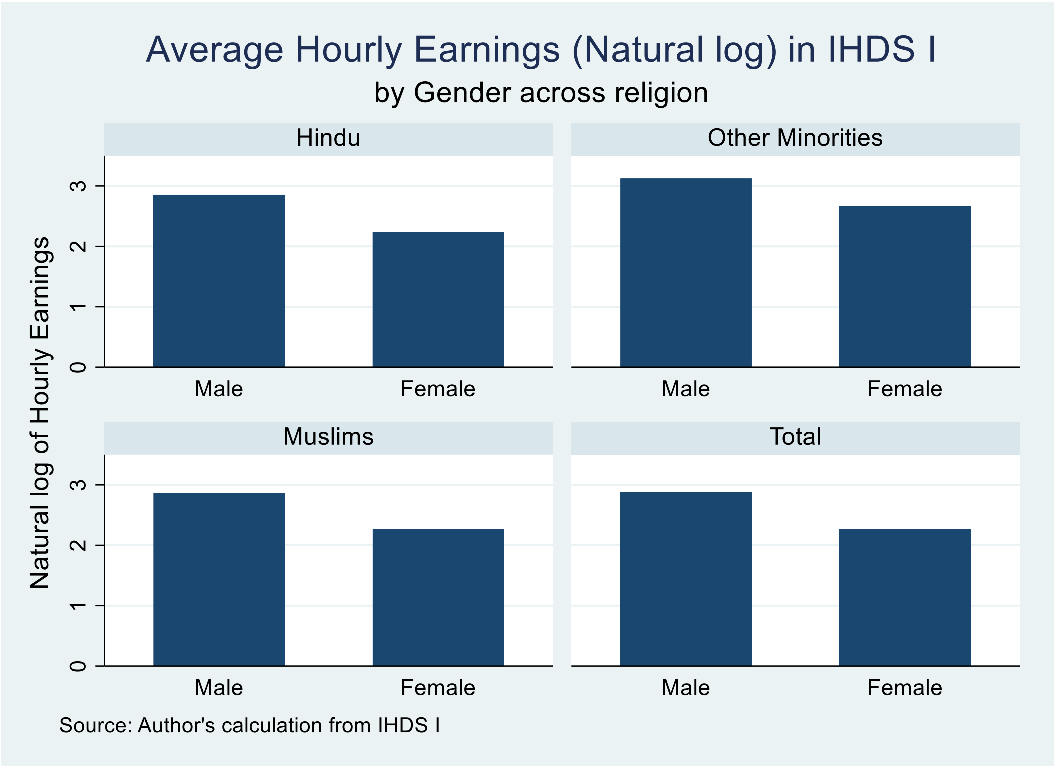

Figure 3. 2: Graphical illustration of Average Hourly Earnings in IHDS I by Gender across Religion

Figure 3.2 depicts the difference in average hourly earnings of men and women across religious groups. Gender differences in average hourly earnings are persistent across all religious groups.

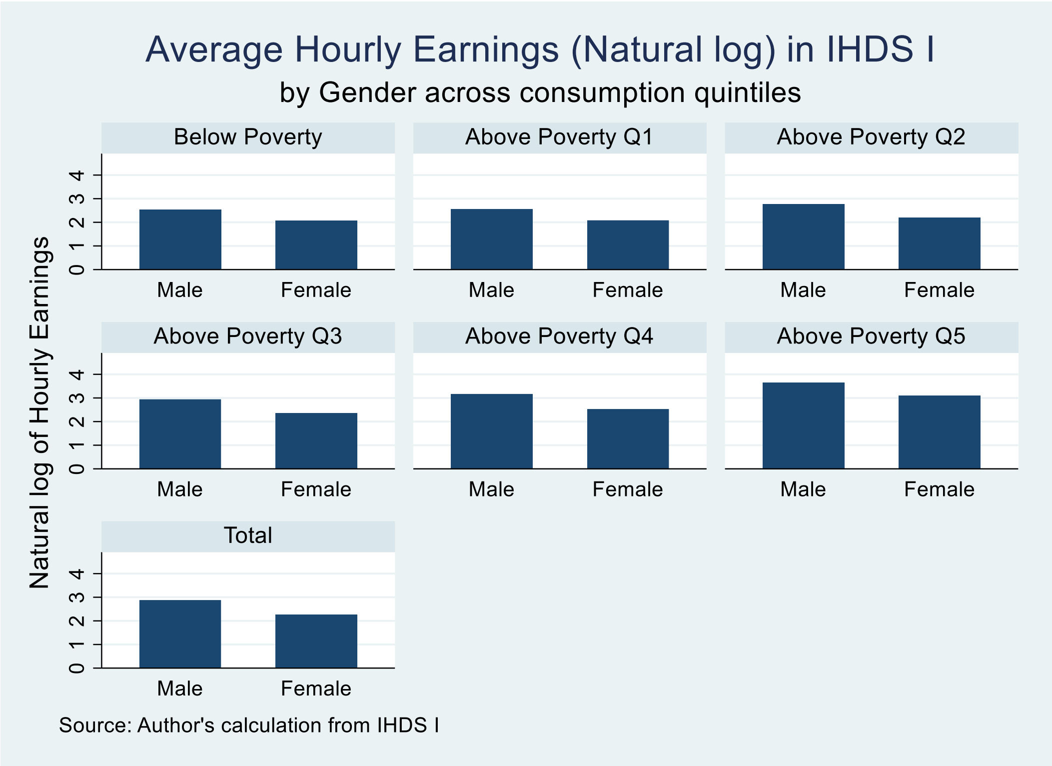

Figure 3. 3: Graphical illustration of Average Hourly Earnings in IHDS I by Gender across Consumption quintiles

Figure 3.3 depicts the difference in average hourly earnings of men and women across consumption quintiles. Gender differences in average hourly earnings are persistent across all consumption quintiles.

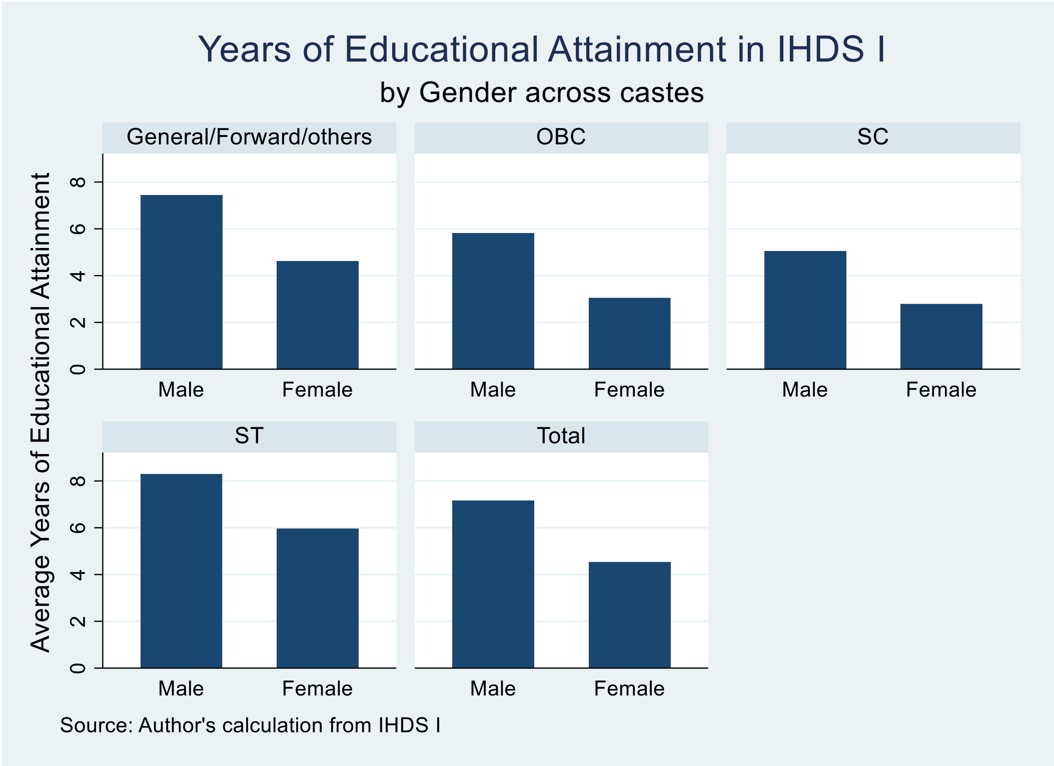

Figure 3. 4: Graphical illustration of Average Educational Attainment in IHDS I by Gender across Caste

Figure 3.4 depicts the difference in average educational attainment of men and women across caste categories. Gender differences in average educational attainment are persistent across all caste categories.

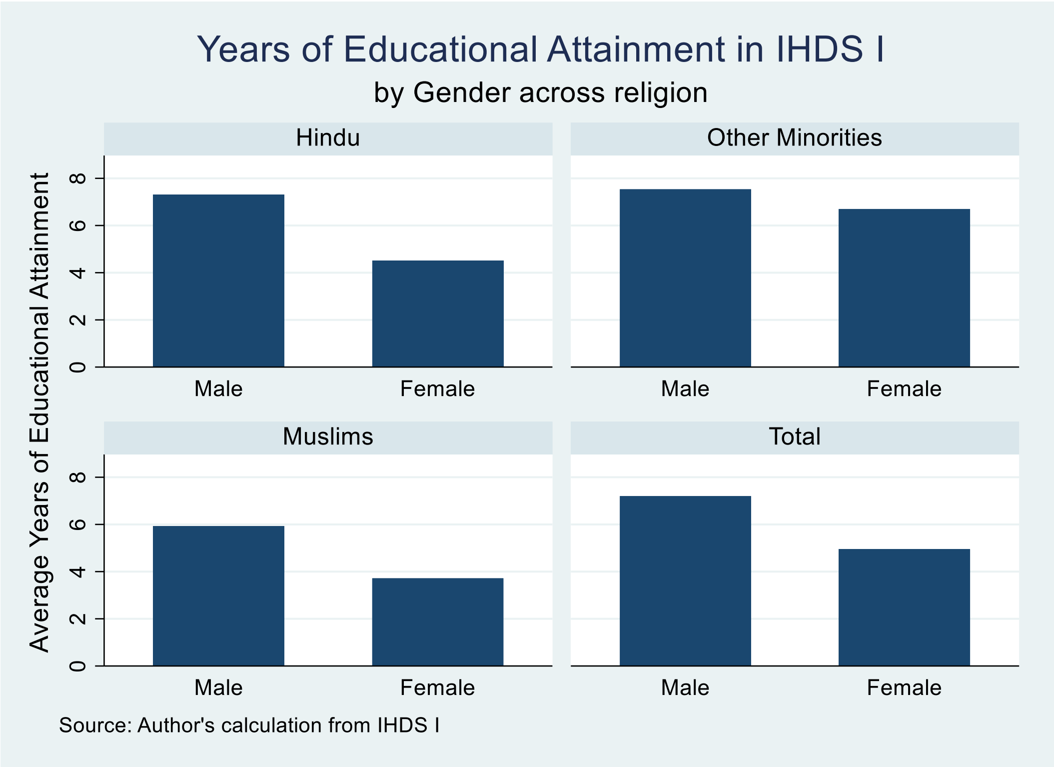

Figure 3. 5: Graphical illustration of Average Educational Attainment in IHDS I by Gender across Religion

Figure 3.5 depicts the difference in average educational attainment of men and women across religious groups. Gender differences in average educational attainment are persistent across all religious groups.

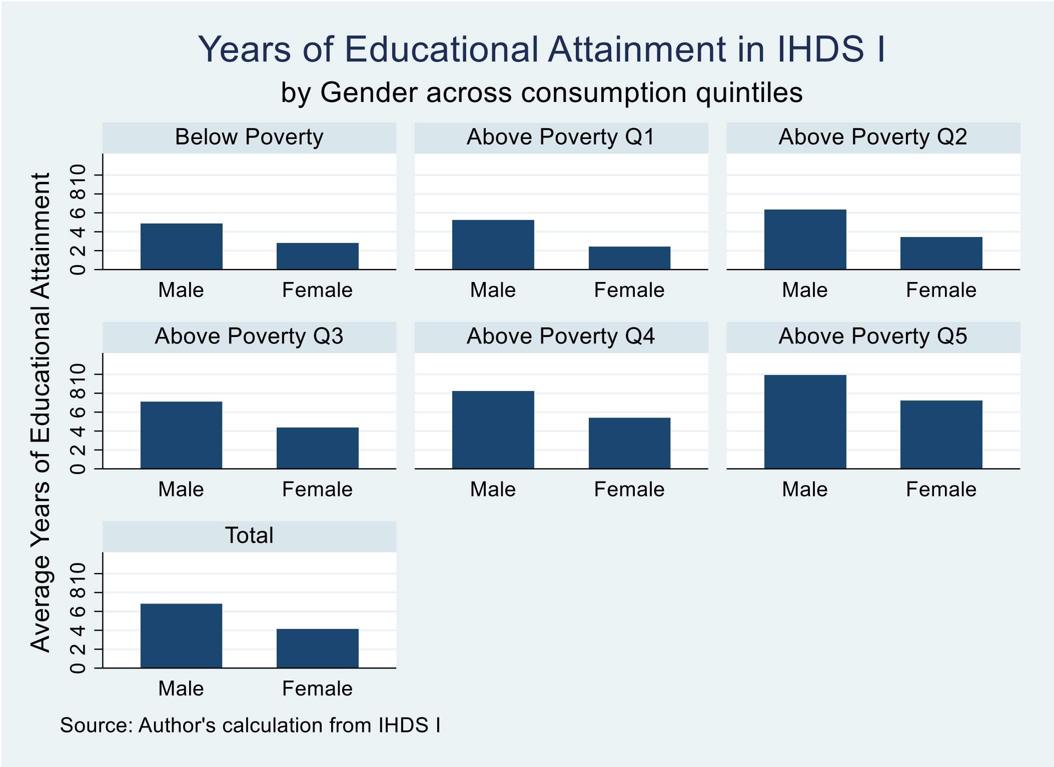

Figure 3. 6: Graphical illustration of Average Educational Attainment in IHDS I by Gender across Consumption quintiles

Figure 3.6 depicts the difference in the average educational attainment of men and women across consumption quintiles. Gender differences in average educational attainment are persistent across all consumption quintiles.

Table 3. 11: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (overall sample)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	24,856	2.875111	.0047697	.7519753	2.865762	2.88446
Female	9,685	2.26521	.0068789	.6769704	2.251726	2.278694
combined	34,541	2.7041	.004204	.7813183	2.69586	2.71234
diff		.6099015	.0087649		.5927219	.627081

```
    diff = mean(Male) - mean(Female)                              t =  69.5844
Ho: diff = 0                                     degrees of freedom =    34539

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS I

Results of two sample t-tests (table 3.11) across genders for the overall sample used in the analysis revealed a significant difference in hourly earnings between men (2.88) and women (2.26). We can also conclude that men have significantly higher hourly earnings than women in the overall sample.

Table 3. 12: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (General/ Forward and other Castes)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	10,476	2.894902	.0074243	.7598977	2.880349	2.909455
Female	3,983	2.252926	.0106377	.6713564	2.23207	2.273782
combined	14,459	2.718058	.0065734	.7904198	2.705173	2.730942
diff		.6419755	.0137114		.6150994	.6688516

```
    diff = mean(Male) - mean(Female)                              t =  46.8205
Ho: diff = 0                                     degrees of freedom =    14457

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS I

Results of two sample t-tests (table 3.12) across gender for General/Forward Castes reveal a significant difference in hourly earnings between men (2.89) and women

(2.25). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 13: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Other Backward Castes)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	6,539	2.764969	.0078479	.6346135	2.749584	2.780353
Female	2,761	2.221468	.0106763	.5609884	2.200534	2.242402
combined	9,300	2.603613	.0068645	.661991	2.590157	2.617069
diff		.5435007	.0139282		.5161983	.570803

```
    diff = mean(Male) - mean(Female)                              t =  39.0216
Ho: diff = 0                                     degrees of freedom =     9298

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS I

Results of two sample t-tests (table 3.113) across gender for Other Backward Castes reveal a significant difference in hourly earnings between men (2.76) and women (2.22). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 14: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Scheduled Caste)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	2,531	2.556056	.0143242	.7206388	2.527967	2.584144
Female	1,622	2.165019	.0159758	.6434122	2.133684	2.196355
combined	4,153	2.403332	.0111301	.717268	2.381511	2.425153
diff		.3910365	.0219941		.3479162	.4341568

```
    diff = mean(Male) - mean(Female)                              t =  17.7791
Ho: diff = 0                                     degrees of freedom =     4151

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author’s calculation from IHDS I

Results of two sample t-tests (table 3.14) across gender for Scheduled Castes reveal a significant difference in hourly earnings between men (2.56) and women (2.17). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 15: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Scheduled Tribes)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	5,310	3.123779	.0109653	.7990351	3.102282	3.145275
Female	1,319	2.517071	.0239577	.8700965	2.470071	2.56407
combined	6,629	3.003059	.0104263	.8488957	2.982621	3.023498
diff		.6067081	.0250322		.557637	.6557792

```
    diff = mean(Male) - mean(Female)                              t =  24.2371
Ho: diff = 0                                     degrees of freedom =     6627

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author’s calculation from IHDS I

Results of two sample t-tests (table 3.15) across gender for Scheduled Tribes reveal a significant difference in hourly earnings between men (3.12) and women (2.51). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 16: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Hindus)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	20,279	2.85511	.0052801	.751912	2.844761	2.86546
Female	8,526	2.23832	.0069416	.6409602	2.224712	2.251927
combined	28,805	2.672546	.0045597	.7738797	2.663609	2.681483
diff		.6167905	.0093043		.5985536	.6350275

```
    diff = mean(Male) - mean(Female)                              t =  66.2906
Ho: diff = 0                                     degrees of freedom =    28803

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS I

Results of two sample t-tests (table 3.16) across gender for Hindus reveal a significant difference in hourly earnings between men (2.86) and women (2.23). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 17: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Other Minorities)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	1,726	3.125333	.0199258	.8278206	3.086251	3.164414
Female	571	2.661692	.0408331	.9757318	2.58149	2.741894
combined	2,297	3.010078	.0185615	.8895986	2.973679	3.046478
diff		.4636408	.0418523		.3815686	.545713

```
    diff = mean(Male) - mean(Female)                              t =  11.0780
Ho: diff = 0                                     degrees of freedom =     2295

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS I

Results of two sample t-tests (table 3.17) across genders for Other Minorities reveal a significant difference in hourly earnings between men (3.12) and women (2.66). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 18: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Muslims)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	2,851	2.865893	.0126076	.6731786	2.841172	2.890614
Female	588	2.270096	.0293406	.7114726	2.212471	2.327722
combined	3,439	2.764023	.0122067	.7158362	2.74009	2.787957
diff		.5957962	.0307933		.5354212	.6561711

```
    diff = mean(Male) - mean(Female)                              t =  19.3483
Ho: diff = 0                                     degrees of freedom =     3437

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS I

Results of two sample t-tests (table 3.18) across gender for Muslims reveal a significant difference in hourly earnings between men (2.86) and women (2.27). We can also conclude that men have significantly higher hourly earnings than women.

3.11.2. Indian Human Development Survey II

In this section, we present graphical illustrations and test differences in averages of the critical variables, namely Hourly Earnings (Natural logarithm) and Educational Attainment (measured in years) by gender and across socio-economic groups using IHDS II data. We believe this will help us understand the differences in these key variables across socio-economic factors, for which we later provide disaggregated estimates. By analyzing these variables in detail, we can identify patterns and trends that can help us better understand the underlying factors that contribute to socio-economic disparities.

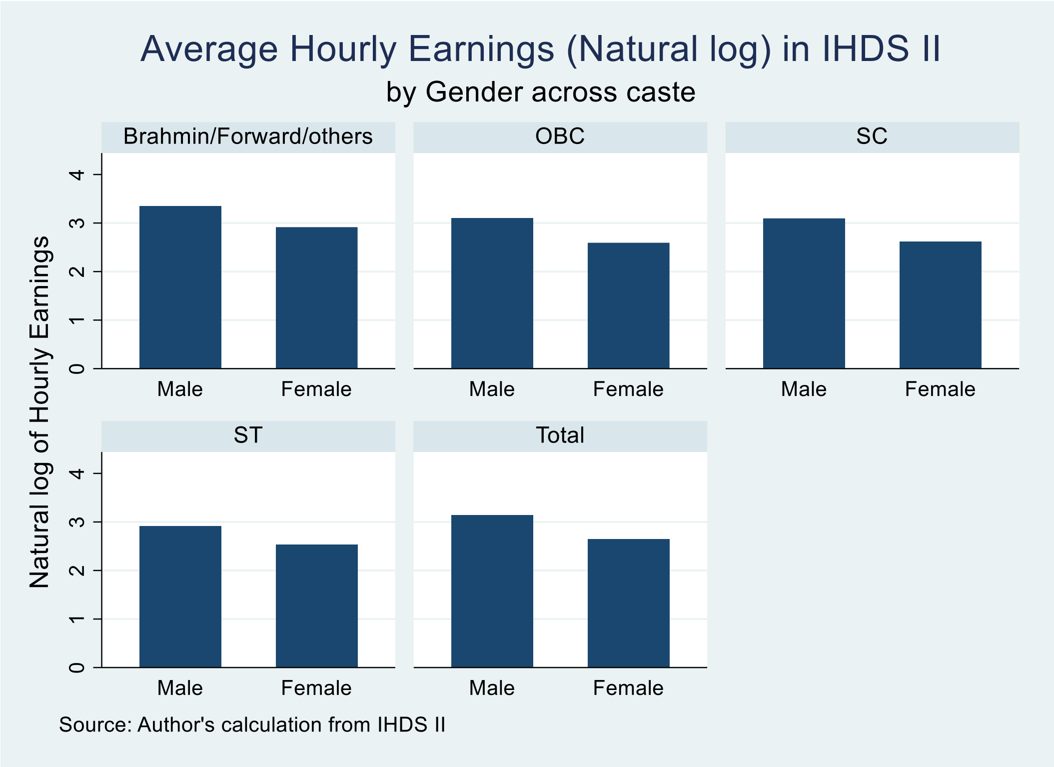

Figure 3. 7: Graphical illustration of Average Hourly Earnings in IHDS II by Gender Across Castes

Figure 3.7 depicts the difference in average hourly earnings of men and women across caste categories. Gender differences in average hourly earnings are persistent across all caste categories.

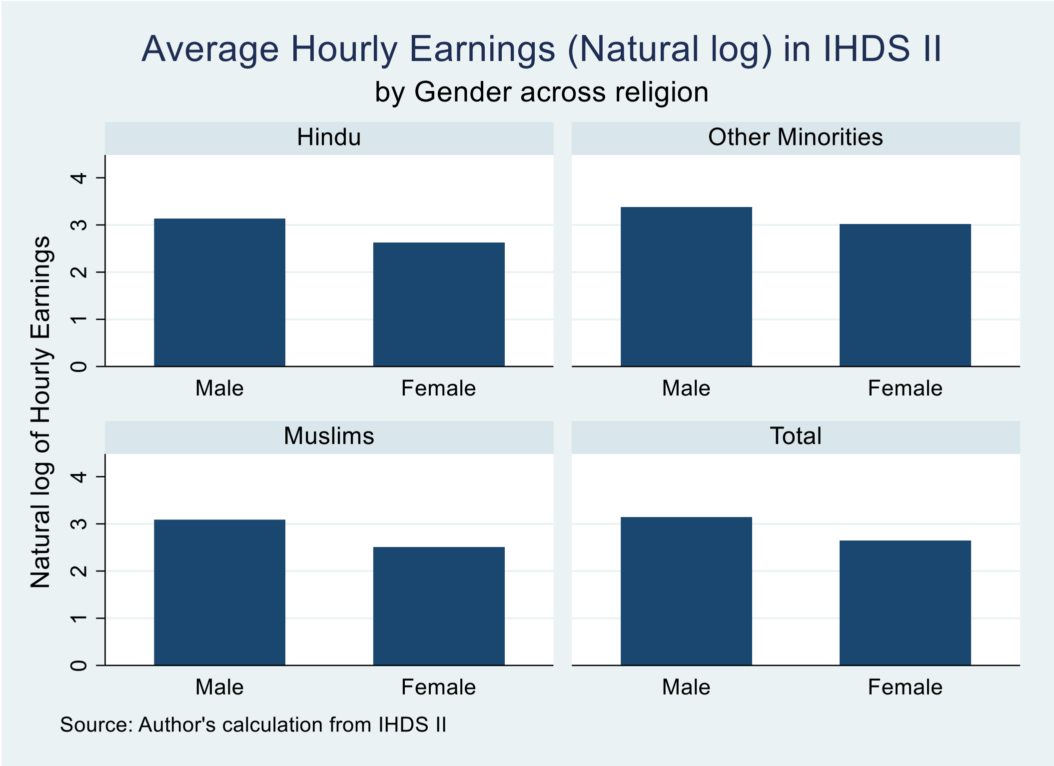

Figure 3. 8: Graphical illustration of Average Hourly Earnings in IHDS II by Gender across Religion

Figure 3.8 depicts the difference in average hourly earnings of men and women across religious groups. Gender differences in average hourly earnings are persistent across all religious groups.

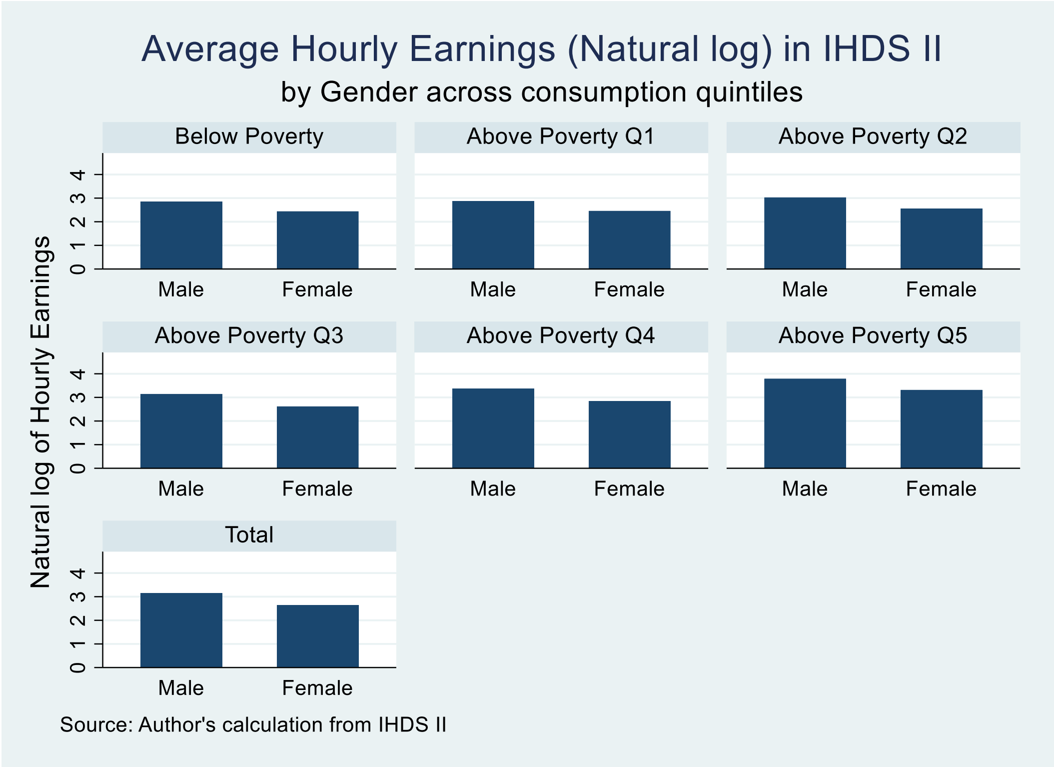

Figure 3. 9: Graphical illustration of Average Hourly Earnings in IHDS II by Gender across Consumption quintiles

Figure 3.9 depicts the difference in average hourly earnings of men and women across consumption quintiles. Gender differences in average hourly earnings are persistent across all consumption quintiles.

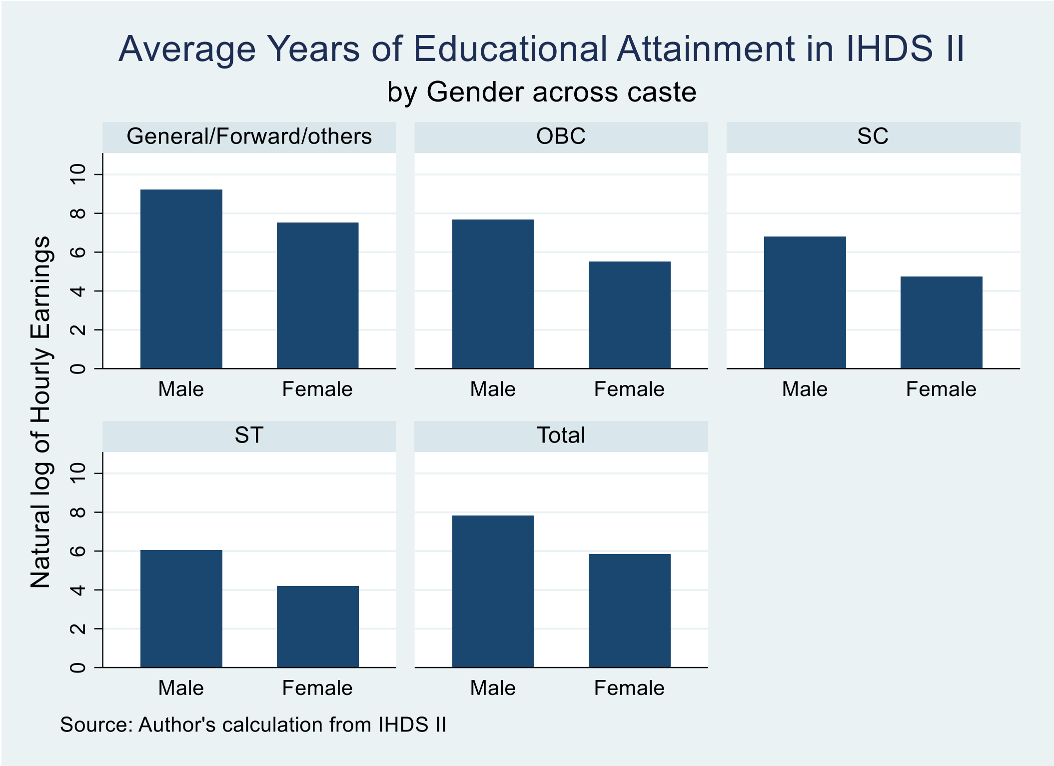

Figure 3. 10: Graphical illustration of Average Educational Attainment in IHDS II by Gender across Caste

Figure 3.10 depicts the difference in average educational attainment of men and women across castes. Gender differences in average hourly earnings are persistent across all castes.

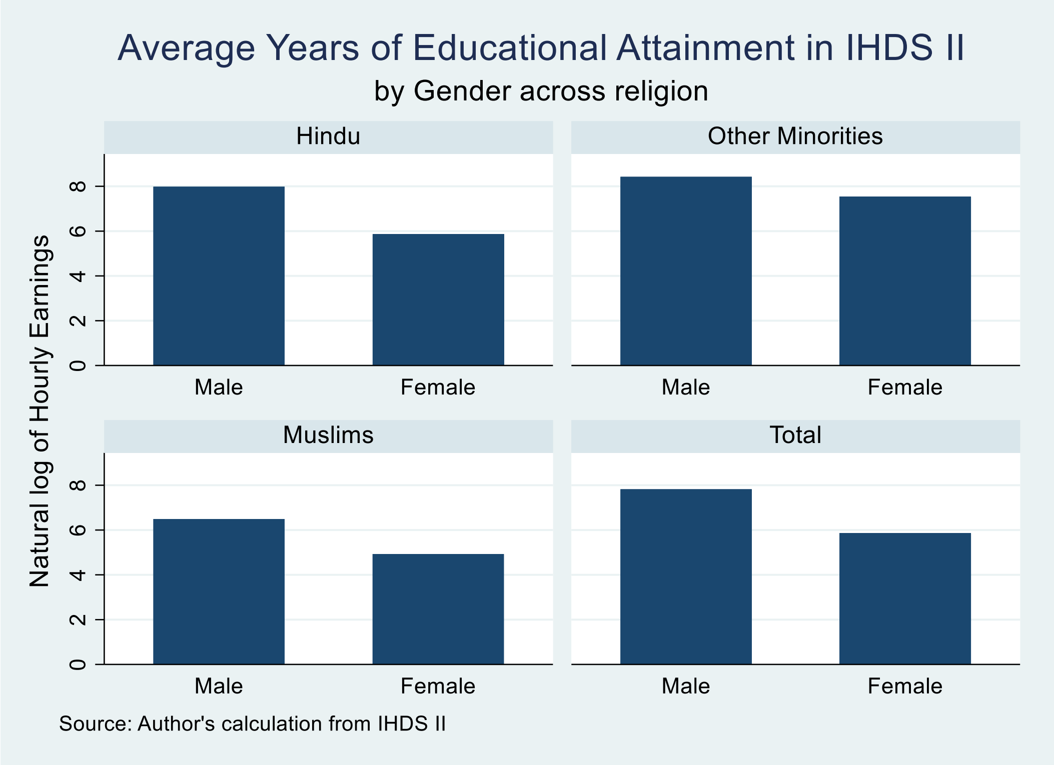

Figure 3. 11: Graphical illustration of Average Educational Attainment in IHDS II by Gender across Religion

Figure 3.11 depicts the difference in average educational attainment of men and women across religions. Gender differences in average hourly earnings are persistent across all religions.

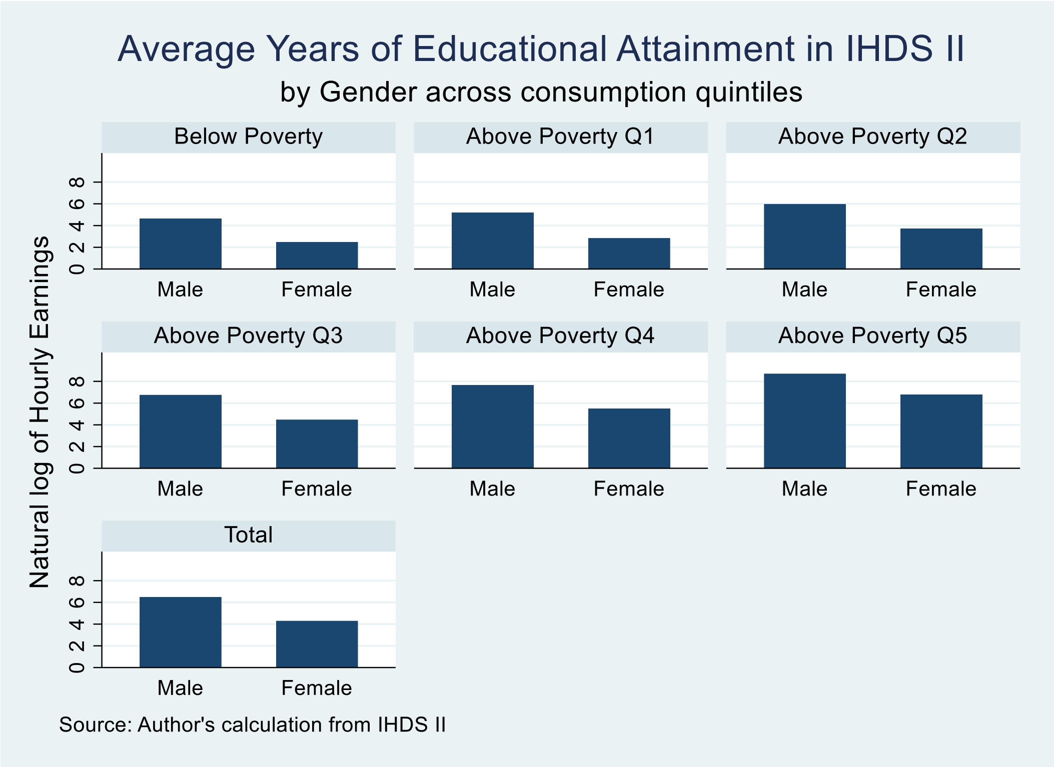

Figure 3. 12: Graphical illustration of Average Educational Attainment in IHDS II by Gender across Consumption quintiles

Figure 3.12 depicts the difference in the average educational attainment of men and women across all consumption quintiles. Gender differences in average hourly earnings are persistent across all consumption quintiles.

Table 3. 19: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (overall sample)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	35,620	3.143215	.0038265	.7221846	3.135715	3.150715
Female	15,609	2.641399	.0055428	.6924961	2.630535	2.652264
combined	51,229	2.990316	.0033124	.7497293	2.983824	2.996809
diff		.5018155	.0068466		.488396	.5152349

```
    diff = mean(Male) - mean(Female)                          t =  73.2937
Ho: diff = 0                                     degrees of freedom =    51227

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author’s calculation from IHDS II

Results of two sample t-tests (table 3.19) across genders for the overall sample used in the analysis revealed a significant difference in hourly earnings between men (3.14) and women (2.64). We can also conclude that men have significantly higher hourly earnings than women in the overall sample.

Table 3. 20: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (General/Forward and Other Castes)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	8,792	3.350903	.0086728	.813212	3.333903	3.367904
Female	2,528	2.912377	.0182774	.9189753	2.876537	2.948218
combined	11,320	3.252971	.0080607	.857621	3.237171	3.268772
diff		.438526	.0189115		.4014561	.4755959

```
    diff = mean(Male) - mean(Female)                          t =  23.1883
Ho: diff = 0                                     degrees of freedom =    11318

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author’s calculation from IHDS II

Results of two sample t-tests (table 3.20) across gender for General/Forward and other Castes reveal a significant difference in hourly earnings between men (3.35) and women (2.91). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 21: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Other Backward Castes)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	13,741	3.100152	.0059626	.6989454	3.088465	3.11184
Female	6,093	2.591368	.0080233	.6262812	2.57564	2.607097
combined	19,834	2.943854	.0050908	.7169496	2.933876	2.953832
diff		.5087838	.010427		.488346	.5292217

diff = mean(Male) - mean(Female)   t = 48.7947
Ho: diff = 0   degrees of freedom = 19832

Ha: diff < 0   Ha: diff != 0   Ha: diff > 0
Pr(T < t) = 1.0000   Pr(|T| > |t|) = 0.0000   Pr(T > t) = 0.0000

Source: Author's calculation from IHDS II

Results of two sample t-tests (table 3.21) across gender for Other Backward Castes reveal a significant difference in hourly earnings between men (3.10) and women (2.59). We can also conclude that men have significantly higher hourly earnings than women

Table 3. 22: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Scheduled Castes)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	9,361	3.095611	.0064298	.6220992	3.083008	3.108215
Female	4,527	2.614491	.0087553	.5890799	2.597326	2.631655
combined	13,888	2.938783	.0055306	.6517725	2.927942	2.949624
diff		.4811208	.0110707		.4594208	.5028207

diff = mean(Male) - mean(Female)   t = 43.4591
Ho: diff = 0   degrees of freedom = 13886

Ha: diff < 0   Ha: diff != 0   Ha: diff > 0
Pr(T < t) = 1.0000   Pr(|T| > |t|) = 0.0000   Pr(T > t) = 0.0000

Source: Author's calculation from IHDS II

Results of two sample t-tests (table 3.22) across gender for Scheduled Castes reveal a significant difference in hourly earnings between men (3.09) and women (2.61). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 23: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Scheduled Tribes)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	3,633	2.916654	.0114992	.6931042	2.894109	2.9392
Female	2,445	2.530703	.0136553	.675212	2.503926	2.55748
combined	6,078	2.761398	.0091268	.7115414	2.743506	2.779289
diff		.3859514	.0179436		.3507757	.4211272

```
    diff = mean(Male) - mean(Female)                              t =  21.5092
Ho: diff = 0                                     degrees of freedom =     6076

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author’s calculation from IHDS II

Results of two sample t-tests (table 3.23) across gender for Scheduled Tribes reveal a significant difference in hourly earnings between men (2.91) and women (2.53). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 24: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Hindus)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	28,941	3.133709	.0042143	.7169428	3.125449	3.14197
Female	13,512	2.623969	.0056349	.6550129	2.612924	2.635015
combined	42,453	2.971469	.0035775	.7371115	2.964457	2.978481
diff		.5097401	.0072709		.495489	.5239911

```
    diff = mean(Male) - mean(Female)                              t =  70.1072
Ho: diff = 0                                     degrees of freedom =    42451

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author’s calculation from IHDS II

Results of two sample t-tests (table 3.24) across gender for Hindus reveal a significant difference in hourly earnings between men (3.13) and women (2.62). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 25: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Other Minorities)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	2,232	3.37883	.017112	.8084411	3.345273	3.412388
Female	1,008	3.019066	.0286344	.9091155	2.962876	3.075256
combined	3,240	3.266904	.0150604	.8572526	3.237375	3.296433
diff		.359764	.0319163		.2971857	.4223422

```
    diff = mean(Male) - mean(Female)                              t =  11.2721
Ho: diff = 0                                     degrees of freedom =     3238

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS II

Results of two sample t-tests (table 3.25) across genders for Other Minorities reveal a significant difference in hourly earnings between men (3.37) and women (3.02). We can also conclude that men have significantly higher hourly earnings than women.

Table 3. 26: Two sample mean comparison of Hourly Earnings (Natural Log) by Gender (Muslims)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	4,447	3.086817	.0103133	.687748	3.066597	3.107036
Female	1,089	2.508088	.0241066	.7955165	2.460787	2.555389
combined	5,536	2.972973	.0100331	.7465029	2.953305	2.992642
diff		.5787286	.0240131		.5316535	.6258038

```
    diff = mean(Male) - mean(Female)                              t =  24.1005
Ho: diff = 0                                     degrees of freedom =     5534

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Source: Author's calculation from IHDS II

Results of two sample t-tests (table 3.26) across gender for Muslims reveal a significant difference in hourly earnings between men (3.08) and women (2.50). We can also conclude that men have significantly higher hourly earnings than women.

3.11.3. Indian Human Development Survey I and II - Balanced Panel Data

In this section, we provided a graphical disaggregation of Average Hourly Earnings (Natural logarithm) and Average Educational attainment (measured by years) by consumption quintiles, gender, caste, and religion for the panel data constructed. The result reported in the previous section highlights that 150,988 individuals are common to both rounds, namely IHDS I and IHDS II.

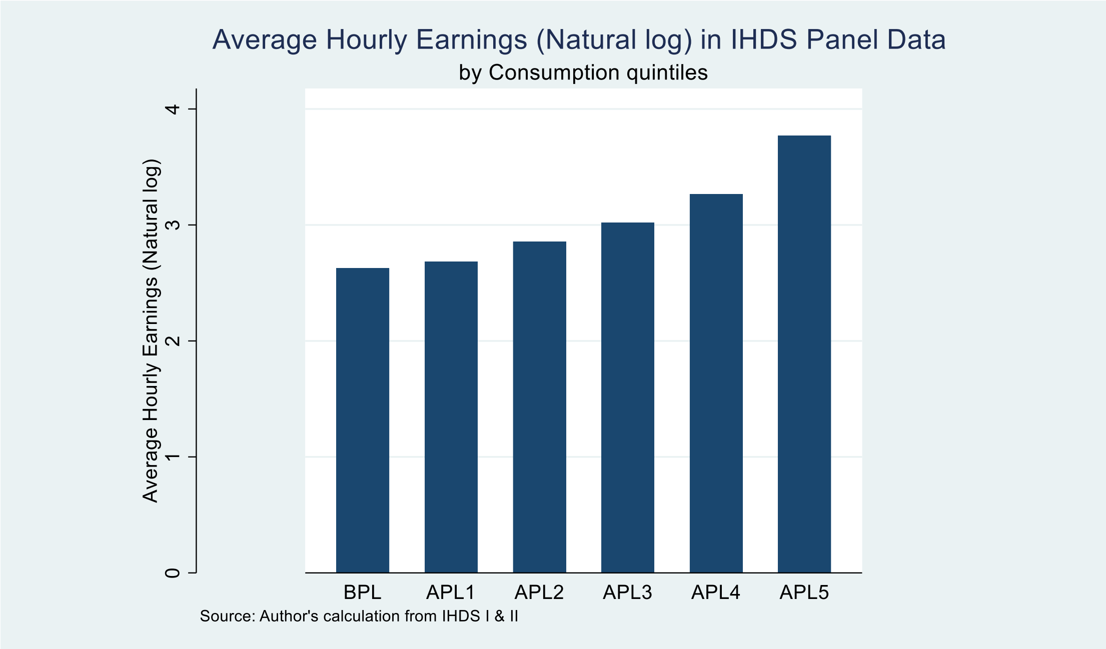

Figure 3. 13: Graphical illustration of Average Hourly Earnings (natural log) in IHDS I and II Panel Data by Consumption quintiles

Figure 3.13 depicts the difference in average hourly earnings across consumption quintiles. We see average hourly earnings rise as we move to higher consumption quintiles.

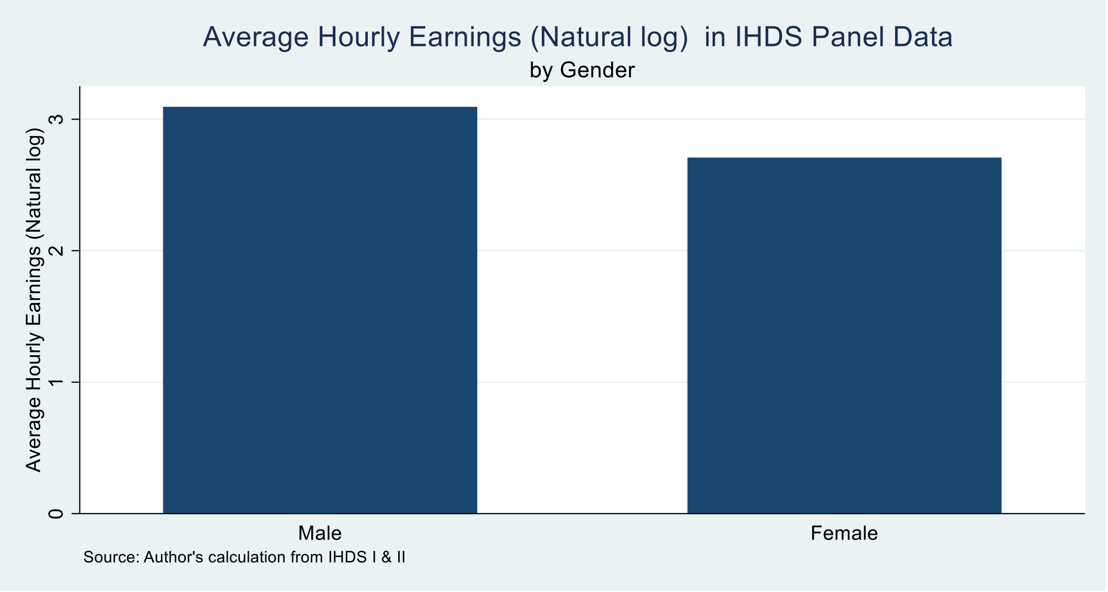

Figure 3. 14: Graphical illustration of Average Hourly Earnings (natural log) in IHDS I and II Panel Data by Gender

Figure 3.14 depicts the difference in average hourly earnings of men and women. We see that average hourly earnings are higher for men as opposed to women.

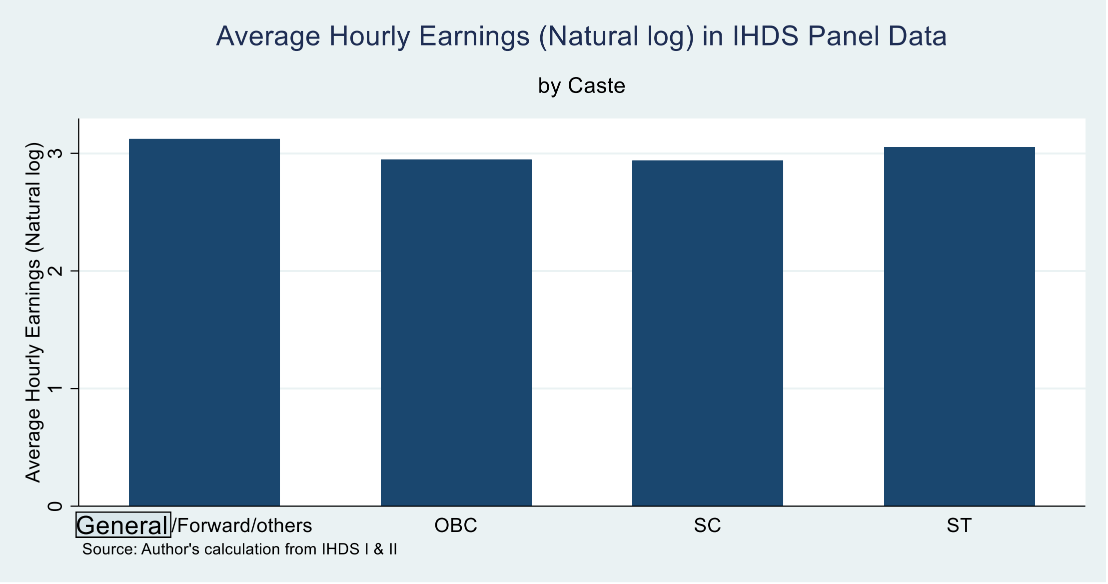

Figure 3. 15: Graphical illustration of Average Hourly Earnings (natural log) in IHDS I and II Panel Data by Caste

Figure 3.15 depicts the difference in average hourly earnings across castes. The average hourly earnings are highest for General/Forward and other Caste categories, followed by Scheduled Tribes.

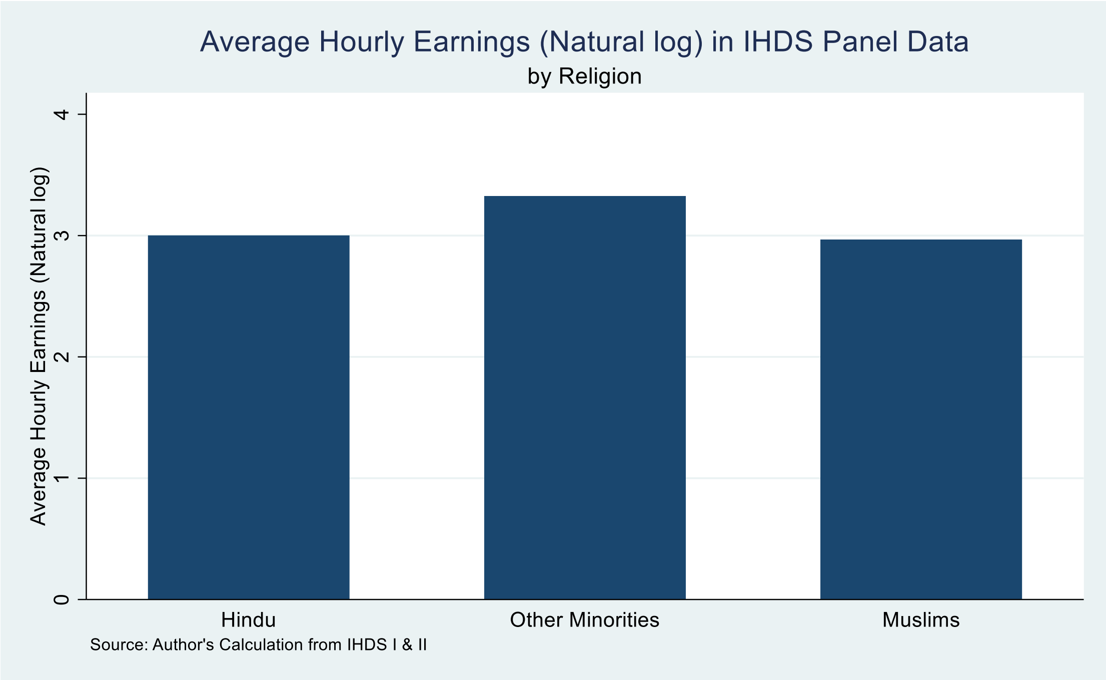

Figure 3. 16: Graphical illustration of Average Hourly Earnings (natural log) in IHDS I and II Panel Data by Religion

Figure 3.13 depicts the difference in average hourly earnings across religions. We see that the average hourly earnings are highest for other Minorities, followed by Hindus.

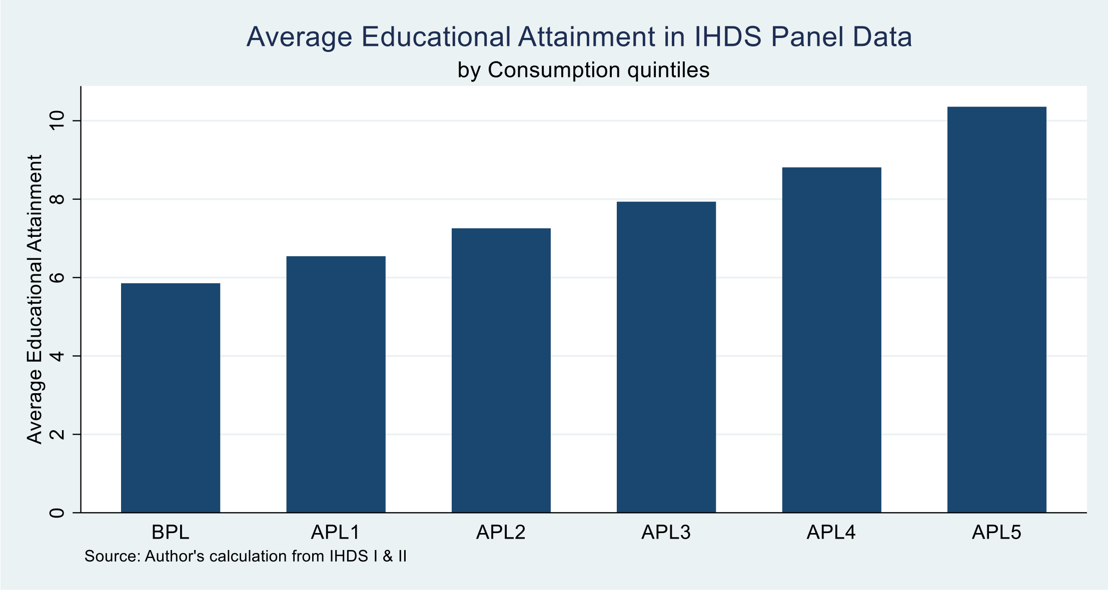

Figure 3. 17: Graphical illustration of Average Educational Attainment in IHDS I and II Panel Data by Consumption quintiles

Figure 3.17 depicts the difference in average educational attainment across consumption quintiles. The average educational attainment is highest for APL5 (5th quintile above the poverty line) and lowest for BPL (below the poverty line). The average educational attainment is higher for higher consumption quintiles.

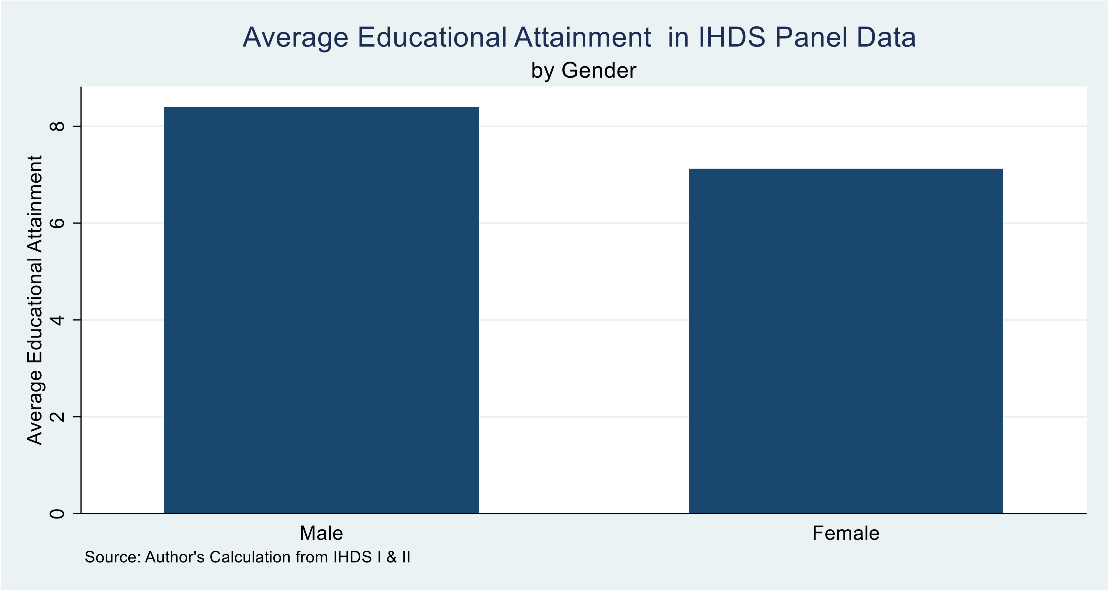

Figure 3. 18: Graphical illustration of Average Educational Attainment in IHDS I and II Panel Data by Gender

Figure 3.18 depicts the difference in average educational attainment across genders. We see that the average educational attainment is higher for men than women.

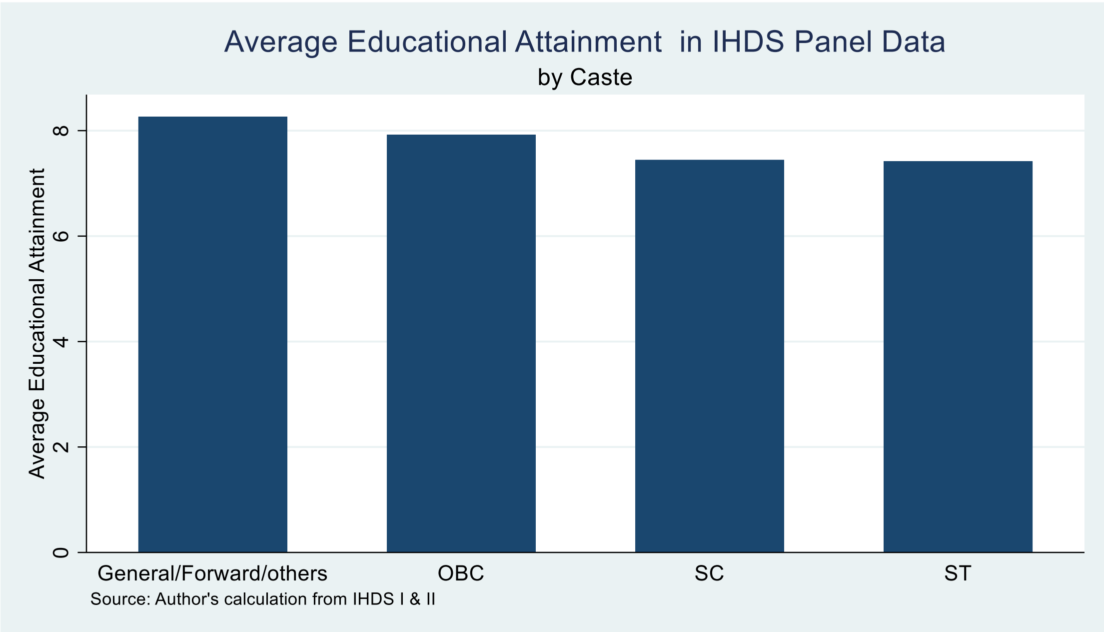

Figure 3. 19: Graphical illustration of Average Educational Attainment in IHDS I and II Panel Data by Caste

Figure 3.19 depicts the difference in average educational attainment across castes. The average educational attainment is highest for General/Forward/other castes and lowest for Scheduled Tribes.

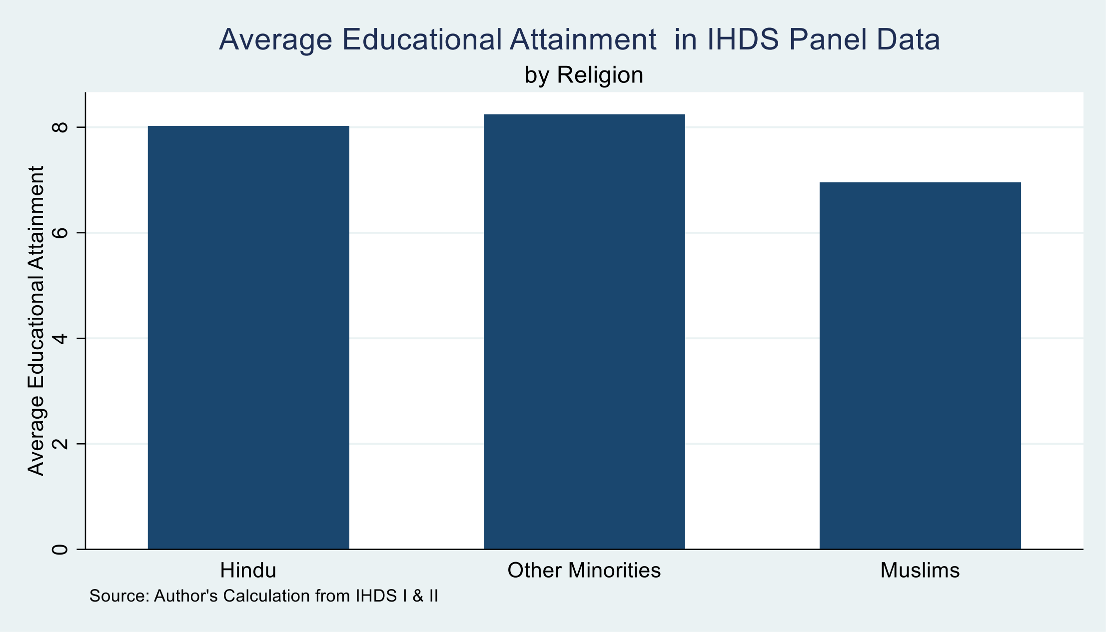

Figure 3. 20: Graphical illustration of Average Educational Attainment in IHDS I and II Panel Data by Religion

Figure 3.20 depicts the difference in average educational attainment across religions. We see that the average educational attainment is highest for other minorities, followed by Hindus, and lowest for Muslims.

3.12. CONCLUSION

This chapter describes the data, variables, and the methodology used in the thesis. The methods sections provided elaborate details of the framework utilized to study a) private returns to education and b) the impact of natural disasters on educational attainment. We provided a background to data sets available in the Indian context. We described the IHDS data and the variables to be used in the following chapters have been discussed, including those that were constructed. The section on Data and variables (3.2) also accounts for the detailed process of the merging used in the construction of the panel data (sub-section3.2.3). Additionally , we discussed the statistical methods employed and software used. The methodology and framework section also discussed in detail the econometric specifications. We provided a graphical illustration of key variables in each round of the IHDS (IHDS I and IHDS II), and conduct statistical tests to hypothesize differences in average values of these key variables across gender and socio-economic groups.

In the next chapter (chapter 4), we provide an overview of the estimation of private returns to education using the methodologies discussed in this chapter. We estimate these returns to education using cross-sectional and panel data sets while testing the suitability of the estimators and assessing the presence of selection bias using specification tests.

CHAPTER 4: EARNINGS, EDUCATION, AND EXPERIENCE IN THE INDIAN CONTEXT: AN OVERVIEW

The previous chapter provided a detailed overview of the methodology, data, and variables used in the thesis. It included the methods used to study private returns to education and the impact of natural disasters on educational attainment, as well as statistical tests to compare key variables across different groups. The chapter also discussed the merging process used in constructing the panel data and the variables to be used in the following chapters.

This chapter presents an analysis of an individual's human capital investment in years of education and experience and its returns in the form of hourly earnings. It addresses issues and differences in estimating these average private returns using an extended Mincer's earnings function from the Ordinary Least Squares method and Heckman's correction for selection bias.

In this chapter, we analyze and present results for Individual level data from the IHDS. Our analysis comprises of Cross-sectional analysis for IHDS I , Cross-sectional analysis for IHDS II, and Panel data analysis (IHDS I and II). The chapter focuses on estimating returns to human capital investment (Years of education), Potential job market experience (Years) at the individual level using Mincer's earnings function. In addition, we also estimate the impact of social capital (as measured by Networking Intensity) on earnings.

4.1. INTRODUCTION

Returns to human capital investment are commonly estimated using model specifications with the help of Ordinary Least Square. These estimates are known to suffer from ability and selection bias. The analysis in this chapter attempts to ascertain that the traditional OLS estimates suffer from selection bias in cross-sectional data as well as panel data, as found in the literature. Alternatively, researchers have used the instrumental variables approach (Card, 1993) and Heckman selection approach (Heckman, 1976, 1979,1990) to address this issue. We use the traditional OLS approach and Heckman's two-step correction approach to compare estimates from these models. We compare the estimates from an extended Mincer function using two methods, namely, OLS and Heckman's two-step selection bias correction method, and provide the range of these comparative estimates. Additionally, we also test for the persistence of selection bias.

4.2. THEORETICAL FRAMEWORK

4.2.1. Mincer Earnings Function

Mincer's earliest form of the earnings function ran a single regression equation in the semi-log form (Mincer 1974). The model used education, experience, and the square experience as co-variates. Over the years, researchers have used many variations of Mincer's original model for use in different contexts. One of the reasons for using a quadratic form for experience was to account for any non-linear effect that it may have on earnings, which could emerge from a loss in productivity due to age.

Mincer's earnings function can be framed as:

lnHourly_Earnings = f (Education, Experience, Experiencesqr) (4.1)

The early econometric method for estimating the Mincer earnings function used the ordinary least squares (OLS) equation (Hartog & Gerritsen, 2016; Polachek, 2008).

The econometric equation would then be stated as:

$$lnHourly_{Earnings} = \beta 0 + \beta 1\, Education + \beta 2\, Experience + \beta 3\, Experiencesqr + e \quad (4.2)$$

where e = stochastic error term

Researchers have developed extended versions of the model by including polynomial terms in education (quadratic, cubic) and experience (cubic and quartic) in an attempt to test its validity for various national and international data. Hence, firstly, we estimate an extended version of Mincer's equation using data from IHDS.

However, it has been recognized that OLS estimates suffer from at least two sources of bias— selection and ability bias (Angrist & Krueger, 1991). Selecting only those individuals who reported earnings to estimate the returns to an additional year of education could overestimate the effect of education. This error in estimation, using only the sub-sample of individuals who reported earnings, is called selection bias (Heckman 1974; 1976; 1979).

These other form of bias reported in studies on returns to education is that of ability bias. Different methods have been proposed to tackle the problem of endogeneity. The most popular one is the instrumental variable (IV) approach (Carneiro, Heckman, & Vytlacil, 2011). The challenge in empirical studies has been to find effective instruments (Angrist & Krueger, 2001; Baltagi & Khanti-Akom, 1990). In non-experimental data, the challenge of finding the right instrument has led researchers to abandon the IV method (Rossi, 2014; Truong et al., 2021), as weak instruments lead to greater bias than other techniques (Bound et al., 1995; Cruz & Moreira, 2005). For this reason, we have chosen not to use the IV method.

However, we use the Heckman selection model to control for selection bias (Wooldridge, 2018), in cross-sectional data and the Heckman two-step correction model with fixed effects in longitudinal data (Kamhöfer & Schmitz, 2016; Leigh, 2008). This model controls for average differences across individuals in any observable or unobservable predictors. The within-group estimate explains the effect of covariates on an individual's earnings over time.

4.2.2. Heckman Selection Model

In the first stage (selection equation), we predicted participation in the workforce with instruments such as Married, Nchildren, and Nadults apart from the explanatory variables of the outcome equation (Semykina & Wooldridge, 2010). The selection equation is stated as:

$$\text{Labour Force Participation} = \alpha 0 + \alpha 1 Married + \alpha 2 Nchildren + \alpha 3 Nadults + \alpha 4(X) + \varepsilon \quad (4.3)$$

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 4 and ε = stochastic error term.

In the second stage of the Heckman procedure (outcome equation), we used a semi-log polynomial model to allow second-order covariates for experience (equation 4.4). The outcome equation is stated as follows:

$$lnHourly\_Earnings = \beta 0 + \beta 1\, Education + \beta 2\, Experience + \beta 3\, Experiencesqr + \beta 4\, Networking\ Intensity + \beta 5\, IMR + \beta 6\, Zi + e \quad (4.4)$$

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (eq 4.3) (Married, NChildren, and Nadults), e = error term.

4.2.3. Sub- Sample Used For Analysis

The analysis is restricted to individuals who were interviewed in both rounds of IHDS. The sub-sample selected for analysis in this chapter is restricted to individuals in the age group of 15 to 64 years as per the guidelines of Labour force participation followed by NSSO. Experience (Potential Experience) has been restricted to an upper bound of 45 years.

4.3. CROSS-SECTIONAL ANALYSIS

4.3.1. Descriptive Statistics IHDS I

Table 4. 1:Descriptive Statistics: IHDS I

Variables	Description	Obs	Mean	Std. Dev.	Min	Max
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	35,519	2.70	0.78	-2.22	7.25
Education	Completed years of education	214,399	4.67	4.68	0.00	15.00
Experience	Potential experience: {(Age - Education - 5) Years}	269,335	22.34	18.53	0.00	45.00
Experiencesqr	Square of potential experience	269,335	842.53	878.68	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups household's were members of or associated with.	179,682	0.61	1.02	0.00	7.00
Age	Age in years	215,754	27.35	19.35	0.00	116.00
Labor force participation	Labour force participation dummy variable ('1' if hourly earnings are observed, '0' otherwise)	269,335	0.18	0.38	0.00	1.00
Married	Married dummy variable ('1' if individual is married, '0' otherwise)	269,335	0.36	0.48	0.00	1.00
Nchildren	Number of children in the household	215,754	2.20	1.86	0.00	17.00
Nadults	Number of adults in household	215,754	3.27	1.72	0.00	18.00

Source: Author's calculation using IHDS I Data

In table (4.1), we find that average hourly earnngs (natural log) is 2.70 (std deviation, 0.78). Average education is 4.67 years (std deviation, 4.68 years). Average experience in job market is 22.34 years (std deviation, 18.53 years). And 36 percent of the individuals were married. Average age in our sub-sample was 27.35 years (std deviation, 19.35).

4.3.2. Comparative Models - IHDS I (Cross-sectional)

Here, we estimate the impact of Education and Experience on the *Hourly Earnings* (natural log)of individuals using Heckman's two-step correction procedure. This model addresses the issue of selection bias in the OLS model estimated in previous sub-

Table 4. 2: Comparative table of estimates from OLS model and Heckman models

Outcome Equation Dependent Variable: lnHourly_Earnings	**OLS**	**HECKMAN (MLE)**	**HECKMAN (Two-step)**
Education	0.106***	0.104***	0.105***
	(0.00)	(0.00)	(0.00)
Experience	0.0484***	0.0553***	0.0531***
	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.00059***	-0.000714***	-0.000675***
	(0.00)	(0.00)	(0.00)
Networking Intensity	-0.00035	-0.00303	-0.00217
	(0.00)	(0.00)	(0.00)
Constant	1.408***	0.967***	1.113***
	(0.02)	(0.05)	(0.03)
Selection Equation Dependent Variable: Labourforce Participation			
Married	-	0.284***	0.290***
	-	(0.01)	(0.01)
Number of Children	-	0.0353***	0.0330***
	-	(0.00)	(0.00)
Number of Adults	-	-0.132***	-0.135***
	-	(0.00)	(0.00)
Constant	-	-0.342***	-0.332***
	-	(0.01)	(0.01)
athrho	-	0.530***	-
	-	(0.05)	-
insigma	-	-0.371***	-
	-	(0.02)	-
IMR	-	-	0.221***

	-	-	(0.02)
Total obs	34013	109385	109385
Selected obs	-	34013	34013
Non selected obs	-	75372	75372
R Square	0.350	-	-
Adjusted Rsquare	0.350	-	-
IMR		0.335	0.221
F	3650.1	-	-
chi2		16494.7	18042.2
P value	**0.0000**	**0.0000**	**0.0000**

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

The table (4.2) summarizes the estimates for the extended Mincer's earnings function using OLS model, Heckman selection model (MLE), and Heckman two-step selection model (TSE) from IHDS I.

In the OLS model, the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Square of Experience,* and *Networking Intensity*. The coefficient for *Education* is 0.106. The variable is positive and significant. The implication is that a change of one unit in years of education causes a positive change in earnings by 10.6 percent. The coefficient of *Experience* is 0.045. The negative sign on the quadratic term for experience indicates that the rate of change in earnings as a result of experience is increasing at a decreasing rate , showing an inverted-U kind of relationship. A change of one unit in years of experience causes a positive change in earnings of 4.5 percent. The coefficient of the intercept term is 1.41 and is positive and significant. We find that the coefficients for *Education* (private returns to education) and *Experience* are positive and significant. Hence, we can conclude that the variables affect an individual's hourly earnings. However, *Networking Intensity* is not significant.

The overall model is significant, and the R square for the model is 0.35, indicating that the model explains 35 percent of the change in the natural log of hourly earnings of individuals.

In the Heckman selection model (MLE), the main equation, comprises of the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Experiencesqr*, and *Networking* Intensity. The coefficient for *Education* is 0.104. The variable is significant at the 1 percent level of significance. A change of one year in years of education causes a positive change in earnings of 10.4 percent. The coefficient of *Experience* is 0.055. The negative sign on the quadratic term for experience indicates non-linearities that present an inverted U-shaped relationship between earnings and experience. The variable is significant at the 1 percent level of significance. A change of one year in years of potential experience causes a positive change in earnings of 5.5 percent. The coefficient of the intercept term is 0.967, which is positive and significant. However, we find that *Networking Intensity* is not significant in the Heckman selection model (MLE). The Inverse Mills ratio (IMR) lets us decide if the model would otherwise suffer from selection bias. The coefficient for IMR is 0.335, however the model does not test for its significance. However, it does estimate 'athrho' (inverse hyperbolic tangent of correlation between the errors of selection and main equation) and 'insigma' (natural log of the standard error of the residual in the main equation) which are significant. The IMR here is the product of the standardized rho and sigma.

In the Heckman Two-step selection model (TSE), the main equation, comprises of the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Experiencesqr*, and *Networking* Intensity. The coefficient for *Education* is 0.105. The variable is significant at the 1 percent level of significance. A change of one year in years of education causes a positive change in earnings of 10.5 percent. The coefficient of *Experience* is 0.055. The negative sign on the quadratic term for experience indicates non-linearities that present an inverted U-shaped relationship between earnings and experience. The variable is significant at the 1 percent level of significance. A change of one year in years of potential experience causes a positive change in earnings of 5.3 percent. The coefficient of the intercept term is 1.11, which is positive and significant. *Networking Intensity* is also not significant in the Heckman Two-step selection model (TSE). The coefficient for IMR is 0.22, and is statistically significant indicating presence of selection bias.

In the Selection equation of the Heckman selection models (MLE and TSE), the dependent variable is *labour force participation*, and the instrumental variables in the selection equation are *Married, Number of Children*, and *Number of Adults* in the household. Each of these instrumental variables is significant.

The significance of the overall Heckman selection model, is determined with a Wald test. It estimates a Chi-Square of 16494.7 (MLE) and 18042.2 (Two step consistent), both are significant. The total number of observations OLS (34,013), Heckman (MLE) (1,09,385), and Heckman Twostep (1,09,385). The OLS model did not account for more than 68 percent of the observations.

4.3.3. Descriptive Statistics IHDS II

Table 4. 3:Descriptive statistics: IHDS II

Variable	**Description**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	53,351	2.98	0.75	0.00	7.82
Education	Completed years of education	2,04,336	5.32	4.88	0.00	16.00
Experience	Potential experience: {(Age - Education - 5) Years}	2,69,335	24.48	18.81	0.00	45.00
Experiencesqr	Square of potential experience	2,69,335	952.71	900.05	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups household's were members of or associated with.	1,60,427	0.68	1.05	0.00	7.00
Age	Age in years	2,04,565	29.82	20.36	0.00	99.00
Labourforce Participation	Labour force participation dummy variable ('1' if hourly earnings are observed, '0' otherwise)	2,69,335	0.20	0.40	0.00	1.00
Married	Married dummy variable ('1' if individual is married, '0' otherwise)	2,69,335	0.35	0.48	0.00	1.00
Number of Children	Number of children in the household	2,04,568	1.84	1.66	0.00	18.00
Number of Adults	Number of adults in household	2,04,568	3.41	1.70	0.00	18.00

Source: Author's calculation using IHDS II Data

In table (4.3), we find that average hourly earnngs (natural log) is 2.98 (std deviation, 0.75). Average education is 5.32 years (std deviation, 4.88 years). Average experience in job market is 24.48 years (std deviation, 18.81 years). And 35 percent of the individuals were married. Average age in our sub -sample was 29.82 years (std deviation, 20.36).

4.3.4. Comparative Models - IHDS II (Cross-sectional)

Table 4. 4: Comparative table of estimates from OLS model and Heckman models

Outcome Equation Dependent Variable: lnHourly_Earnings	OLS	HECKMAN (MLE)	HECKMAN (Two-step)
Education	0.0835***	0.0826***	0.0826***
	(0.00)	(0.00)	(0.00)
Experience	0.0334***	0.0345***	0.0347***
	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.00039***	-0.00037***	-0.000380***
	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0307***	0.0294***	0.0292***
	(0.00)	(0.00)	(0.00)
Constant	1.914***	1.744***	1.706***
	(0.02)	(0.02)	(0.03)
Selection Equation Dependent Variable: Labourforce Participation			
Married	-	0.0100	0.0127
	-	(0.01)	(0.01)
Number of Children	-	0.00991***	0.00554**
	-	(0.00)	(0.00)
Number of Adults	-	-0.129***	-0.128***
	-	(0.00)	(0.00)
Constant	-	0.213***	0.213***
	-	(0.01)	(0.01)
athrho	-	0.275***	-
	-	(0.03)	-

insigma	-	-0.400***	-
	-	(0.01)	-
IMR	-	-	0.218***
	-	-	(0.02)
Total obs	46488	111669	111669
Selected obs	-	46488	46488
Non selected obs	-	65181	65181
R Square	0.239	-	-
Adjusted Rsquare	0.239	-	-
IMR	-	0.180	0.218
F	2676.4	-	-
chi2	-	13783.7	14019.6
P value	**0.0000**	**0.0000**	**0.0000**

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

The table (4.4) summarizes the estimates for the extended Mincer's earnings function using OLS model, Heckman selection model (MLE), and Heckman two-step selection model (TSE) from IHDS II.

In the OLS model, the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Square of Experience,* and *Networking Intensity*. The coefficient for *Education* is 0.084. The variable is positive and significant. The implication is that a change of one unit in years of education causes a positive change in earnings by 8.4 percent. The coefficient of *Experience* is 0.033. The negative sign on the quadratic term for experience and its significane indicates that the rate of change in earnings as a result of experience is increasing at a decreasing rate , showing an inverted-U kind of relationship. A change of one unit in years of experience causes a positive change in earnings of 3.3 percent. The coefficient of the intercept term is 1.9 and is positive and significant. We find that the coefficients for

Education (private returns to education) and *Experience* are positive and significant. Hence, we can conclude that the variables affect an individual's hourly earnings. The coefficient of *Networking Intensity* (0.031) is positive and significant. Indicating a positive impact on natural log earnings of 3.1 percent.

The overall model is significant, and the R square for the model is 0.24, indicating that the model explains 24 percent of the change in the natural log of hourly earnings of individuals.

In the Heckman selection model (MLE), the main equation, comprises of the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Experiencesqr*, and *Networking* Intensity. The coefficient for *Education* is 0.0826. The variable is significant at the 1 percent level of significance. A change of one year in years of education causes a positive change in earnings of 8.3 percent. The coefficient of *Experience* is 0.0345. The negative sign on the quadratic term for experience indicates non-linearities that present an inverted U-shaped relationship between earnings and experience. The variable is significant . A change of one year in years of potential experience causes a positive change in earnings of 3.5 percent. The coefficient of the intercept term is 1.74, which is positive and significant. Additionaly, we find that *Networking Intensity* (0.029) is positive and significant in the Heckman selection model (MLE). Indicating a positive impact on natural log earnings of 2.9 percent. The coefficient for IMR is 0.18, however the model does not test for its significance.

In the Heckman Two-step selection model (TSE), the main equation, comprises of the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Experiencesqr*, and *Networking* Intensity. The coefficient for

Education is 0.0826. The variable is positive and significant. A change of one year in years of education causes a positive change in earnings of 8.3 percent. The coefficient of *Experience* is 0.0347. The negative sign on the quadratic term for experience indicates non-linearities that present an inverted U-shaped relationship between earnings and experience. A change of one year in years of potential experience causes a positive change in earnings of 3.5 percent. The coefficient of the intercept term is 1.7, which is positive and significant. *Networking Intensity* (0.029) is positive and significant in the Heckman Two-step selection model (TSE). The coefficient for IMR is 0.218, and is statistically significant indicating presence of selection bias.

In the Selection equation of the Heckman selection models (MLE and TSE), the dependent variable is *labour force participation*, and the instrumental variables in the selection equation are *Married, Number of Children*, and *Number of Adults* in the household. *Number of Children*, and *Number of Adults are* significant.

The significance of the overall Heckman selection model, is determined with a Wald test. It estimates a Chi-Square of 13783.7 (MLE) and 14019.6 (Two step consistent), both are significant. The total number of observations OLS (46,488), Heckman (MLE), and Heckman Twostep (1,11,669). The OLS model did not account for more than 58 percent of the observations.

4.4. PANEL DATA ANALYSIS (IHDS I AND IHDS II)

In this section, we estimate returns to an additional year of education (private returns) using individual panel data from IHDS I & II. Our estimation consists of 3 steps. Firstly, we provide a detailed descriptive statistic for the variables used in the analysis. Next, we estimate and compare the random and fixed effects model using the Hausman test to determine the suitability of the appropriate model. Lastly, we use the suitable model to estimate Heckman's corrected estimates addressing selection bias.

4.4.1. Descriptive Statistics

Table 4. 5: Descriptive statistics

Variable	Description		Mean	S.D	Min	Max	Obs
lnHourly_ Earnings	Natural logarithm of hourly earnings observed	overall	3.02	0.82	-0.84	6.92	21413
		bet		0.77	-0.84	6.71	15601
		within		0.28	1.10	4.94	1.37
Education	Completed years of education	overall	7.90	4.44	0	16	52134
		bet		4.03	0	16	29878
		within		1.63	-0.10	15.90	1.74
Experience	Potential experience: {(age - education - 5) years}	overall	19.51	13.53	0	45	52378
		bet		13.62	0	45	29889
		within		2.71	-1.99	41.01	1.75
Experiencesqr	Square of potential experience	overall	563.77	629.60	0	2025	52378
		bet		627.80	0	2025	29889
		within		138.98	-446.7	1574.27	1.75

Networking intensity	Number of social or economic committees that household members were associated with	overall bet within	0.71	1.08 0.90 0.62	0 0 -2.79	7 7 4.21	52225 29870 1.75
Labor force participation	Binary variable ('1' if hourly earnings are observed, '0' otherwise)	overall bet within	0.41	0.49 0.43 0.25	0 0 -0.09	1 1 0.91	52378 29889 1.75
Married	Binary variable ('1' if the individual is married, '0' otherwise)	overall bet within	0.62	0.49 0.46 0.18	0 0 0.12	1 1 1.12	52378 29889 1.75
Children	Number of children in the household	overall bet within	1.54	1.62 1.41 0.84	0 0 -4.96	18 15 8.04	52378 29889 1.75
Adults	Number of adults in the household	overall bet within	3.47	1.74 1.50 0.90	0 0 -4.53	18 18 11.47	52378 29889 1.75
Consumption quintiles	'0' if below the poverty line, '1' if the lowest quintile above the poverty line, and	overall bet within	2.66	1.75 1.54 0.89	0 0 0.16	5 5 5.16	52365 29889 1.75

	'5' if the highest quintile above the poverty line						
Female	Binary variable ('1' if the individual is female, '0' otherwise)	overall bet within	0.39	0.49 0.49 0.03	0 0 -0.11	1 1 0.89	52378 29889 1.75
Caste	Categories: '1' if general /others, '2' if other backward classes, '3' if scheduled castes, and '4' if scheduled tribes	overall bet within	2.11	1.10 0.81 0.78	1 1 0.61	4 4 3.61	52310 29888 1.75
Religion	Categories: '1' if Hindu, '2' if Other Minorities, and '3' if Muslims	overall bet within	1.33	0.69 0.70 0.10	1 1 0.33	3 3 2.33	52378 29889 1.75

Source: Author's calculations based on IHDS I and II

We found that the average of hourly earnings (natural log) of an individual (overall) was 3.02 with a standard deviation of 0.8 (see Table 4.7). Average education was 7.9 years, with a standard deviation of 4.4 years (overall). The average experience was 19.5 years, with a standard deviation of 13.5 years (overall). Of the total sample, 62% were married, and 39% were female. The average number of children was 1.54, and the average number of adults was 3.47.

4.4.2. Panel Estimates For Fixed And Random Effects Models

Table 4. 6: Panel estimates for fixed and random effects models (dependent variable: lnHourly_Earnings).

Variables	Random	Fixed
Education	0.101***	0.0666***
	(0.0013)	(0.0033)
Experience	0.0374***	0.0608***
	(0.0016)	(0.0046)
Experiencesqr	-0.000381***	-0.000575***
	(0.0000)	(0.0001)
Networking Intensity	0.0191***	0.0174**
	(0.0044)	(0.0074)
Constant	1.656***	1.570***
	(0.0198)	(0.0508)
Observations	21291	21291
R-Square	0.191	0.207

Source: Author's calculations based on IHDS I and II

Standard errors in parentheses

*** p<0.1, ** p<0.05, *** p<0.01**

First, we estimated the Mincer earnings function using fixed effect (FE) and random effect (RE) models. Then, we used the Hausman test (Hausman command) to confirm that the FE model was more suitable for our analysis. Table 4.6, illustrates the panel estimates for the fixed and random effect models and results for Hausman test (table 4.7). Using the dependent variable (natural log) *Hourly Earnings*, we estimate the coefficient for *Education, Experience, Experiencesqr,* and *Networking Intensity*. We estimate the coefficient using a random effects model and a fixed effects model. The coefficient for *Education* was 0.066 in the fixed effects model and 0.101 in the random effects model. The coefficient for *Experience* is 0.0374 in the random effect model and 0.06 in the fixed effects model. The square of experience was negative and significant in both models. *Networking Intensity* was positive and significant in both models as well. The random effects model provided an R square (Between) of 0.19, and the fixed model provided an R square (Within) of 0.21. Results from the Hausman test are illustrated below in Table 4.7.

4.4.3. Hausman Test To Choose Between Fixed And Random Effects Model

Table 4. 7: Hausman test to choose between fixed and random effects models (dependent variable: lnHourly_Earnings).

Variables	Fixed	Random	Difference	S.E.
Education	0.0666	0.1015	-0.0349	0.00
Experience	0.0608	0.0374	0.0234	0.00
Experiencesqr	-0.0006	-0.0004	-0.0002	0.00
Networking Intensity	0.0174	0.0191	-0.0018	0.00

Source: Author's calculations based on IHDS I and II

Test: Ho: Differences in coefficients are not systematic

Chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 186.29

Prob>Chi2 = 0.0000

The Chi-squared value of 186.29 is significant, indicating the use of the fixed effects model. We use this result to establish that the fixed effects model is appropriate and proceed along this line to estimate the Heckman selection model with fixed effects using STATA 16.1 newly defined command 'xtheckmanfe'.

4.4.4. Panel Estimates From The Fixed-Effects Heckman Model (full sample)

Unlike earlier panel commands for the Heckman model, *xtheckmanfe* allows us to incorporate FE (fixed effects) while addressing selection bias. This, we believe, is a methodological improvement over other previous studies (Briggs, 2004; Schwiebert, 2012), and our estimates update their results.

Table 4. 8: **Panel Estimates from the Fixed-effects Heckman Model (full sample) (dependent variable: lnHourly_Earnings).**

Main Equation	**Estimate**	***SE***	**95 % CI**		***p***
			LL	***UL***	
Education	0.0348***	0.004	0.026	0.0436	0.0000
Experience	0.012**	0.005	0.0013	0.0226	0.0290
Experiencesqr	-0.0001	0.000	-0.0003	0.0001	0.3100
Networking Intensity	0.0164**	0.007	0.0031	0.0298	0.0160
2012	0.2981***	0.057	0.1871	0.4092	0.0000
IMR					
2005	-0.2213***	0.038	-0.2956	-0.1471	0.0000
2012	-0.3799***	0.049	-0.4765	-0.2832	0.0000
Constant	1.5735***	0.056	1.4628	1.6843	0.0000

Note: Total N = 51,989, CI = confidence interval, LL = lower limit, and UL= upper limit.

Source: Author's calculations based on IHDS I and II

** p<0.1, ** p<0.05, *** p<0.01*

To decide which of the FE and RE models was more suitable for this study, we used the Hausman test (results can be found in Table 4.7). The test statistic chi-square value was 186.29 (with a p-value 0.0000), which indicates rejection of the null hypothesis (of RE). Therefore, we accepted the alternate hypothesis that the FE model was better suited to our data.

The FE Heckman model (for the full sample) suggests that earnings are positively and significantly affected by Education (0.035). The coefficient of Experience (0.012) is positive and significant, but that of the square term is negative and insignificant. Networking intensity (0.016)has a positive and significant impact on earnings. The coefficients for year specific IMR (2005, (-) 0.22 & 2012, (--) 0.38) are significant and indicate a selection bias, thus justifying the use of the selection model. The results of the selection equation are reported in table 4.9.

Selection equation and estimates

In the selection equation (4.3), we used *Married*, *Nadults*, and *Nchildren* as excluded instruments (Comola & de Mello, 2013). These coefficients have been used in the recent literature as well (Beam et al., 2020; Sarkar et al., 2019). These variables could be acceptable instruments for selection in the labor market and we tested the same for confirmation. The model also estimates a value for the IMR by year to test whether selection bias exists in the estimation. IMR is included in the outcome equation (4.4) as a covariate. A significant value of IMR would imply the existence of selection bias.

Table 4. 9: Selection equation estimates from fixed effects Heckman Model (without disaggregation) (dependent variable: lnHourly_Earnings).

Variables	Coefficient	Standard Error	Z-Statistic	P value	95 Percent Confidence Interval	
Year Specific Education						
2005	0.0417	0.009	4.41	0.0000	0.0232	0.0603
2012	-0.0271	0.010	-2.61	0.0090	-0.0474	-0.0067
Year Specific Experience						
2005	-0.0006	0.015	-0.04	0.9700	-0.0292	0.0281
2012	0.0605	0.009	6.65	0.0000	0.0427	0.0783
Year Specific Experience_square						
2005	-0.0003	0.000	-1.03	0.3020	-0.0008	0.0002
2012	-0.0011	0.000	-6.52	0.0000	-0.0014	-0.0008
Year Specific NetworkingIntensity						

2005	-0.0200	0.014	-1.39	0.1660	-0.0484	0.0083
2012	0.0162	0.013	1.26	0.2090	-0.0091	0.0416
Year Specific Married						
2005	-0.8115	0.057	-14.19	0.0000	-0.9236	-0.6994
2012	0.8594	0.046	18.50	0.0000	0.7684	0.9504
Year Specific Nchildren						
2005	-0.0573	0.012	-4.95	0.0000	-0.0801	-0.0346
2012	-0.0081	0.011	-0.75	0.4510	-0.0291	0.0129
Year Specific Nadults						
2005	-0.0386	0.012	-3.11	0.0020	-0.0629	-0.0143
2012	-0.0088	0.011	-0.81	0.4180	-0.0300	0.0125
Year-Specific Individual Average Education						
2005	-0.0254	0.010	-2.43	0.0150	-0.0458	-0.0049
2012	0.0253	0.010	2.42	0.0160	0.0048	0.0459
Year-Specific Individual Average Experience						
2005	0.1390	0.017	8.42	0.0000	0.1066	0.1713
2012	-0.0014	0.010	-0.15	0.8830	-0.0206	0.0177
Year Specific Individual Average Experience_square						
2005	-0.0020	0.000	-7.13	0.0000	-0.0026	-0.0015
2012	-0.0002	0.000	-0.86	0.3920	-0.0005	0.0002
Year-Specific Individual Average Networking Intensity						
2005	0.0180	0.016	1.13	0.2590	-0.0132	0.0492
2012	0.0064	0.017	0.37	0.7100	-0.0273	0.0401
Year Specific Individual Average Married						
2005	0.7754	0.063	12.36	0.0000	0.6524	0.8983
2012	-1.0516	0.057	-18.48	0.0000	-1.1631	-0.9400
Year Specific Individual Average Nchildren						
2005	0.0901	0.014	6.41	0.0000	0.0626	0.1177
2012	-0.0047	0.012	-0.38	0.7030	-0.0290	0.0195
Year Specific Individual Average Nadults						

2005	-0.1076	0.014	-7.69	0.0000	-0.1350	-0.0801
2012	-0.1148	0.012	-9.81	0.0000	-0.1377	-0.0919
Year						
2012	1.6346	0.070	23.36	0.0000	1.4974	1.7718
Constant	-1.6452	0.067	-24.58	0.0000	-1.7763	-1.5140

Source: Author's calculations based on IHDS I and II

* p<0.1, ** p<0.05, *** p<0.01

4.5. FINDINGS

The estimates from IHDS I do show evidence of selection bias in the model. However, the coefficient for education, which is the private returns to education for an additional year, is similar in all estimated models (around 10 percent).

A comparison of the coefficients from the two models suggests that the coefficient of experience, however, is lower in the OLS (0.48) model as compared to the Heckman models (MLE, 0.55 and TSE, 0.53). All the explanatory variables are significant in either model except *Networking Intensity* (not significant any overall model).

The estimates from IHDS II also show evidence of selection bias in the model. However, the coefficient for education, which is the private returns to education for an additional year, is similar in all estimated models (8.3 percent).

A comparison of the coefficients from the two models suggests that the coefficient of experience, is also similar in OLS and the Heckman models (between 3.3 to 3.5 percent). All the explanatory variables are significant across models. The coefficient for *Networking Intensity* marginally higher in the OLS (0.31) as compared to the Heckman models (0.29).

In the Heckman selection model with fixed effects using the xtheckmanfe command, we found that the within estimator (OLS) suffers from selection bias with the coefficient for *IMR* for 2005 (-0.2213) and 2012 (-0.3799) being highly significant. The coefficients for *Education* were positive and significant in the within estimator (OLS) (0.0666) and Heckman selection model (0.0348). This illustrates that the within estimator (OLS) overestimated the coefficients for private returns to an additional year of education when compared with the Heckman model with fixed effects. The coefficient for *Education* was nearly twice as large in the within estimator (OLS).

The coefficient for *Experience* was positive and significant in both the models as well; however, the estimates were overestimated in the within estimator (OLS) (0.0608) as compared to the Heckman selection model (0.012) with fixed effects. The coefficient for square of experience was negative and significant in the within the estimator (OLS) but not significant in the Heckman model.

The variable *Networking Intensity* was positive and highly significant in the within estimator (OLS) (0.0174) and the Heckman model (0.0164); the coefficient was marginally higher for the OLS model. The selection instruments were highly significant (except Nadults 2012) in the Heckman-type model. The number of observations in the within estimator (OLS) is reported as 21291, and in the Heckman, it is reported as 51989.

4.6. CONCLUSION

The estimate of returns to an additional year of education estimated using OLS leads to the problem of selection bias. The way forward is to estimate these returns to an additional year of schooling using the Heckman selection approach to addresses selection bias.

We also conclude that *Education* and *Experience* are highly explanatory variables in this model. We find that both education and experience positively impact earning in the overall data. We find evidence for an inverted U-shaped relationship between earnings and experience. This is true in the context of cross-sectional analysis (OLS and Heckman models) and panel data analysis (in within estimator only). This is consistent with the literature, however in the Heckman model with fixed effect we find it only jointly significant .

We also explored the possibility of social capital in the form of *Networking Intensity* in explaining variations in earnings between and within. Using cross-sectional and panel data analysis, we find that social capital positively affects earnings in cross-sectional models for IHDS II and panel data models.

In the next four chapters, we use our models to provide disaggregated estimates across class (economic distributions), gender, castes, and religion. This disaggregation, we believe, will help us throw light on some important findings within the context of the Indian labor market and the disparities in returns to education across sgender, and ocio-economic groups.

The next chapter (Chapter 5) discusses private returns to education disaggregated by consumption quintiles (economic distribution) using the Heckman models. This will help discuss economic inequalities in the Indian context.

CHAPTER 5: TRENDS IN RETURNS TO HUMAN CAPITAL INVESTMENT ACROSS CLASS

In the previous chapter, we summarized the estimates for returns and an additional year of education, experience and the impact of social capital for the overall sample. We also explored the use of three models to estimate our extended Mincer's earnings function. We further analyzed the data using the Ordinary Least squares and the Heckman type selection models (MLE and TSE), leading to the conclusion that the Heckman models are more suitable for our sample, that has shown the presence of selection bias. Lastly, in our panel data analysis, we derived a similar conclusion in addition to the suitability of a fixed effects model.

5.1. INTRODUCTION

In Chapter 5, we examine the trends in estimates of private returns to human capital investment in education, experience across class (economic distribution). We create six consumption categories using household consumption data from IHDS. The consumption categories consist of one group of BPL (Below the poverty line) and five quintiles of APL (Above the poverty line (hereafter APL1 to APL5). This chapter uses cross-sectional and panel data.

5.2. FRAMEWORK

The basic framework for the model follows from Chapter 4, where we use Mincer's earnings function as the basic equation: this is given by:

lnHourly Earnings = f (Education, Experience, Experience_square)   (5.1)

We use an extended form of the equation to incorporate Networking Intensity ***(a proxy for social capital).***

5.2.1. Models

Model 1: Extended Mincer's Earnings function using OLS

$$\boldsymbol{Ln\ Hourly\ Earnings_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \beta_3 Experience_i^2 + \beta_4 Networking\ Intensity_i + u_i} \qquad \textit{\textbf{(5.2)}}$$

Where u = stochastic error term and β1 = the average private returns to an additional year of education; however, as pointed out in Chapter 4, the estimation of this suffers from at least two sources of endogeneity biases—namely, selection Bias and ability Bias. Hence, we proceed to estimate our model using Heckman's Two-step procedure to address selection bias.

Model 2: Extended Mincer's Earnings function using Heckman selection models to address selection bias (using MLE and TSE)

First stage equation:

$$\boldsymbol{Labourforce\ Participation_i = a_0 + a_1 Married_i + a_2 Nchildren_i + a_3 Nadults_i^2 + a_4 X_i + e_i} \qquad \textit{\textbf{(5.3)}}$$

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 5.4 and e = stochastic error term.

Second stage equation:

$$\boldsymbol{Ln\ Hourly\ Earnings_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \beta_3 Experience_i^2 + \beta_4 Networking\ Intensity_i + \beta_5 IMR_i + \beta_6 Z_i + u_i} \qquad \textit{\textbf{(5.4)}}$$

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (eq 5.3) (Married, Nchildren, and Nadults), u= error term.

5.3. ANALYSIS

5.4. CROSS-SECTIONAL ANALYSIS IHDS I

In this section of the chapter, we proceed towards analyzing data from round 1 of IHDS, namely IHDS I. Our analysis consists of descriptive statistics of the selected sample, followed by estimation of OLS and Heckman selection models disaggregated by consumption categories (as described in sub-section 5.1).

5.4.1. Descriptive Statistics for IHDS I

Table 5. 1: Table of Descriptive Statistics

Variables	Description	Obs	Mean	Std. Dev.	Min	Max
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	35,519	2.70	0.78	-2.22	7.25
Education	Completed years of education	214,399	4.67	4.68	0.00	15.00
Experience	Potential experience: {(Age - Education - 5) Years}	269,335	22.34	18.53	0.00	45.00
Experiencesqr	Square of potential experience	269,335	842.53	878.68	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups households were members of or associated with.	179,682	0.61	1.02	0.00	7.00
Age	Age in years	215,754	27.35	19.35	0.00	116.00

Labourforce Participation	Labor force participation dummy variable ('1' if hourly earnings are observed, '0' otherwise)	269,335	0.18	0.38	0.00	1.00
Married	Married dummy variable ('1' if the individual is married, '0' otherwise)	269,335	0.36	0.48	0.00	1.00
Number of Children	Number of children in the household	215,754	2.20	1.86	0.00	17.00
Number of Adults	Number of adults in the household	215,754	3.27	1.72	0.00	18.00
Consumption Quintiles	Consumption quintiles: '0' Below the poverty line, '1' lowest quintile above the poverty line, and '5' Above poverty line highest quintile	164,619	2.08	1.82	0.00	5.00

Source: Author's calculation using IHDS I Data

We found that an individual's (natural log of) average hourly earnings were 2.70 with a standard deviation of 0.78. Average education was 4.67 years, with a standard deviation of 4.7 years. The average experience was 22.34 years, with a standard deviation of 18.5 years. Of the total sample, 36% were married. The average number of children was 2.2, and adults was 3.3.

5.4.2. Ordinary Least Squares Model – IHDS I

Table 5. 2: OLS model estimates disaggregated by consumption quintiles

Outcome Equation Dependent Variable: lnHourly_Earnings						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	**APL5**
Education	0.0530***	0.0527***	0.0621***	0.0806***	0.103***	0.130***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0196***	0.0230***	0.0358***	0.0402***	0.0591***	0.0619***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.00025***	-0.0003***	-0.00049***	-0.0005***	-0.0007***	-0.0007***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0400***	0.0197***	-0.00526	-0.0206***	-0.0260***	-0.0456***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.883***	1.849***	1.808***	1.770***	1.438***	1.330***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Total obs	9654	5686	5126	4915	4396	4236
R square	0.110	0.105	0.151	0.227	0.331	0.460
Adj Rsquare	0.110	0.105	0.150	0.226	0.331	0.459
F	241.3	122.5	172.1	327.7	550.9	1062.0
P value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Author's calculation using IHDS I Data

The table (5.2) summarizes the disaggregated estimates for the extended Mincer's earnings function using OLS model. In the OLS model, the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Square of Experience,* and *Networking Intensity*. The coefficient of Education was positive and significant for all consumption quintiles, BPL (0.053), APL1 (0.053), APL2 (0.062), APL3 (0.081), APL4 (0.10) and APL5 (0.13). The coefficients for Experience were positive and highly significant across all consumption quintiles, BPL (0.0196), APL1 (0.023), APL2 (0.036), APL3 (0.040), APL4 (0.059), and APL5 (0.062). The coefficients for Experience_square were negative and significant for all consumption quintiles, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and significant for BPL (0.04), APL1(0.013), whereas the coefficients were negative and significant for APL3 (-0.021), APL4 (-0.0026) and APL5 (-0.046) except APL2.The overall model is significant across quintiles.

5.4.3. Heckman Selection Model (Maximum Likelihood Estimates) – IHDS I

Table 5. 3: Heckman selection model estimates (using maximum likelihood estimator) by consumption quintiles

Outcome Equation Dependent Variable: lnHourly_Earnings						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	**APL5**
Education	0.0517***	0.0457***	0.0563***	0.0764***	0.0996***	0.129***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0313***	0.0255***	0.0351***	0.0398***	0.0587***	0.0638***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.0004***	-0.0003***	-0.0005***	-0.0005***	-0.0007***	-0.0007***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0327***	0.00659	-0.00850	-0.0247***	-0.0275***	-0.0473***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.361***	1.388***	1.337***	1.326***	1.045***	1.088***
	(0.04)	(0.04)	(0.05)	(0.06)	(0.08)	(0.09)
Selection Equation Dependent Variable: Labourforce Participation						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	**APL5**
Married	0.497***	0.178***	0.0565**	0.0171	0.00253	0.0396
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Number of Children	-0.000630	-0.0262***	-0.0516***	-0.0262***	-0.0423***	-0.00238
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of Adults	-0.123***	-0.136***	-0.118***	-0.131***	-0.118***	-0.123***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-0.172***	0.221***	0.143***	0.094***	-0.0425	-0.202***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
athrho	0.830***	1.108***	0.934***	0.774***	0.569***	0.286***
	(0.05)	(0.04)	(0.07)	(0.07)	(0.09)	(0.08)
lnsigma	-0.475***	-0.390***	-0.339***	-0.326***	-0.308***	-0.332***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
Total obs	23227	12870	13480	13727	14069	14735
Selected obs	9654	5686	5126	4915	4396	4236
Non selected obs	13573	7184	8354	8812	9673	10499
rho	0.680	0.804	0.733	0.649	0.515	0.279
Lambda (IMR)	0.423	0.544	0.522	0.469	0.379	0.200
chi2	1099.2	498.7	682.1	1225.7	1963.0	3564.4
P value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table (5.3) we report the estimates from Heckman Selection model with maximum likelihood estimator, the coefficients for *Education* were positive and highly significant across all consumption quintiles, namely, BPL (0.0517), APL1 (0.0457), APL2 (0.0563), APL3 (0.0764), APL4 (0.0996) and APL5 (0.129). The coefficients for *Experience* were positive and highly significant across all consumption quintiles, namely, BPL (0.0313), APL1 (0.0255), APL2 (0.0351), APL3 (0.0398), APL4 (0.0587), and APL5 (0.0638). The coefficients for *Experiencesqr* was negative and significant for all consumption quintiles, signifying an inverted U-shaped relationship. The coefficients for *Networking Intensity* were positive and highly significant for BPL (0.032) except APL1, whereas the coefficients were negative and highly significant for APL3 (-0.0247), APL4 (-0.0275) and APL5 (-0.0473) except APL2.

5.4.4. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) – IHDS I

Table 5. 4: Heckman two-step selection model estimates (two-step estimator) by consumption quintiles

Outcome Equation Dependent Variable: lnHourly_Earnings						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	***APL5***
Education	0.0532***	0.0503***	0.0596***	0.0775***	0.0997***	0.129***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0258***	0.0259***	0.0365***	0.0405***	0.0591***	0.0649***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.000348***	-0.000342***	-0.000499***	-0.000526***	-0.000737***	-0.000763***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0366***	0.0142**	-0.00585	-0.0239***	-0.0274***	-0.0482***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.622***	1.491***	1.468***	1.348***	1.034***	0.955***
	(0.05)	(0.06)	(0.07)	(0.07)	(0.09)	(0.12)
Selection Equation Dependent Variable: Labourforce Participation						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	***APL5***
Married	0.557***	0.207***	0.0511*	0.00785	0.00319	0.0430

	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Number of Children	0.00223	-0.0329***	-0.0438***	-0.0162**	-0.0346***	-0.00101
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of Adults	-0.132***	-0.155***	-0.137***	-0.138***	-0.122***	-0.122***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-0.194***	0.276***	0.192***	0.110***	-0.0418	-0.209***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
IMR	0.200***	0.384***	0.350***	0.432***	0.384***	0.309***
	(0.03)	(0.04)	(0.05)	(0.06)	(0.07)	(0.09)
Total obs	23227	12870	13480	13727	14069	14735
Selected obs	9654	5686	5126	4915	4396	4236
Non selected obs	13573	7184	8354	8812	9673	10499
rho	0.366	0.633	0.548	0.612	0.522	0.415
Lambda (IMR)	0.200	0.384	0.350	0.432	0.384	0.309
chi2	1176.0	593.4	819.3	1302.4	1987.3	3541.8
P value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table (5.4); we observe that the coefficient for *Education*, for BPL (0.053), APL1 (0.050), APL2 (0.060), APL3 (0.078), APL4 (0.10), and APL5 (0.13). The coefficients were positive and significant for all consumption quintiles. We also observe that the coefficient for *Experience*, for BPL (0.026), APL1 (0.026), APL2 (0.037), APL3 (0.041), APL4 (0.059), and APL5 (0.065). The coefficients were positive and significant for BPL, APL1, APL2, APL3, APL4 and APL5.

The coefficients for Square of Experience were negative and significant for BPL, APL1, APL2, APL3, APL4 and APL5. The coefficient for *Networking Intensity* for BPL (0.0366), APL1 (0.014), APL3 (-0.024), APL4 (-0.0274), and APL5 (-0.048). The coefficients were positive and significant for BPL and APL1, whereas they were negative and significant for APL3, APL4, and APL5.

The coefficient for IMR was significant across all consumption quintiles, indicating the presence of selection bias. The instruments used for the first stage selection equation; were Married, Nchildren, and Nadults.

5.5. CROSS-SECTIONAL ANALYSIS IHDS II

In this section of the chapter, we proceed toward analyzing data from round 1 of IHDS, namely IHDS II (2011-12). Our analysis consists of descriptive statistics of the selected sample, followed by the estimation of multiple regression model by Consumption categories (as described in the previous section) using the Heckman Selection model.

5.5.1. Descriptive Statistics for IHDS II

Table 5. 5: Table of Descriptive Statistics

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	53,351	2.98	0.75	0.00	7.82
Education	Completed years of education	2,04,336	5.32	4.88	0.00	16.00
Experience	Potential experience: {(Age - Education - 5) Years}	2,69,335	24.48	18.81	0.00	45.00
Experiencesqr	Square of potential experience	2,69,335	952.71	900.05	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups households were members of or associated with.	1,60,427	0.68	1.05	0.00	7.00
Age	Age in years	2,04,565	29.82	20.36	0.00	99.00
Labourforce Participation	Labor force participation dummy variable ('1' if	2,69,335	0.20	0.40	0.00	1.00

	hourly earnings are observed, '0' otherwise)					
Married	Married dummy variable ('1' if the individual is married, '0' otherwise)	2,69,335	0.35	0.48	0.00	1.00
Number of Children	Number of children in the household	2,04,568	1.84	1.66	0.00	18.00
Number of Adults	Number of adults in the household	2,04,568	3.41	1.70	0.00	18.00
Consumption Quintiles	Consumption quintiles: '0' Below the poverty line, '1' lowest quintile above the poverty line, and '5' Above poverty line highest quintile	1,60,679	2.45	1.73	0.00	5.00

Source: Author's calculation using IHDS II Data

We found that an individual's (natural log of) average hourly earnings were 2.98 with a standard deviation of 0.75. Average education was 5.32 years, with a standard deviation of 4.88 years. The average experience was 24.48 years, with a standard deviation of 18.81 years. Of the total sample, 35% were married. The average number of children was 1.84, and the adults were 3.41.

5.5.2. Ordinary Least Squares Model – IHDS II

Table 5. 6: OLS model estimates disaggregated by consumption quintiles

Outcome Equation Dependent Variable: lnHourly_Earnings						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	**APL5**
Education	0.0426***	0.0391***	0.0497***	0.0643***	0.0763***	0.120***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0185***	0.0173***	0.0239***	0.0289***	0.0329***	0.0454***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.0002***	-0.0002***	-0.0003***	-0.00032***	-0.00033***	-0.0004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0729***	0.0529***	0.0472***	0.0286***	0.0282***	-0.0216**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	2.239***	2.281***	2.240***	2.129***	2.029***	1.587***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Total obs	12458	7291	6832	6442	5732	5323
R square	0.0919	0.0726	0.109	0.144	0.175	0.356
Adj Rsquare	0.0916	0.0721	0.108	0.144	0.175	0.356
F	234.6	98.52	152.8	207.0	263.5	760.5
P value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Author's calculation using IHDS II Data

The table (5.6) summarizes the disaggregated estimates for the extended Mincer's earnings function using OLS model. In the OLS model, the dependent variable is *Hourly Earnings* (natural log), and the independent variables are *Education, Experience, Square of Experience,* and *Networking Intensity*. The coefficient of Education was positive and significant for all consumption quintiles, BPL (0.043), APL1 (0.039), APL2 (0.0497), APL3 (0.064), APL4 (0.76) and APL5 (0.12). The coefficients for Experience were positive and significant across all consumption quintiles, BPL (0.019), APL1 (0.017), APL2 (0.024), APL3 (0.029), APL4 (0.033), and APL5 (0.045). The coefficients for Experience_square were negative and significant for all consumption quintiles, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and significant for BPL (0.073), APL1(0.053), APL2 (0.047), APL3 (0.029), APL4 (0.28), and negative for APL5 (-0.0216).The overall model is significant across quintiles

5.5.3. Heckman Selection Model (Maximum Likelihood Estimates)- IHDS II

Table 5. 7: Heckman selection model estimates (using maximum likelihood estimator) by consumption quintiles

Outcome Equation Dependent Variable: lnHourly_Earnings						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	**APL5**
Education	0.0412***	0.0389***	0.0494***	0.0636***	0.0762***	0.123***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0220***	0.0189***	0.0248***	0.0297***	0.0332***	0.0457***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.0003***	-0.0002***	-0.0003***	-0.0003***	-0.0003***	-0.0004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0691***	0.0519***	0.0463***	0.0285***	0.0281***	-0.0218**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.963***	2.139***	2.117***	1.955***	1.943***	2.463***
	(0.03)	(0.05)	(0.06)	(0.06)	(0.09)	(0.06)
Selection Equation Dependent Variable: Labourforce Participation						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	**APL5**
Married	0.192***	0.203***	0.0836***	0.0522**	0.0185	-0.0178
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Number of Children	-0.0176***	-0.0250***	-0.0314***	-0.0220***	-0.0186**	-0.0161**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of Adults	-0.114***	-0.117***	-0.104***	-0.0988***	-0.0795***	-0.0418***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.432***	0.341***	0.276***	0.190***	0.00757	-0.206***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
athrho	0.652***	0.294***	0.231***	0.294***	0.117	-1.208***
	(0.05)	(0.09)	(0.09)	(0.07)	(0.11)	(0.05)
insigma	-0.491***	-0.561***	-0.499***	-0.407***	-0.325***	0.0420*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Total obs	22008	13927	14422	14554	14761	15228
Selected obs	12458	7291	6832	6442	5732	5323
Non selected obs	9550	6636	7590	8112	9029	9905
rho	0.573	0.286	0.227	0.286	0.116	-0.836
Lambda (IMR)	0.351	0.163	0.138	0.190	0.0842	-0.872
chi2	1128.3	552.8	816.0	1052.4	1210.6	3283.1
P value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table (5.7) we report the estimates from Heckman Selection model with maximum likelihood estimator, the coefficients for *Education* were positive and significant across all consumption quintiles, namely, BPL (0.0412), APL1 (0.0389), APL2 (0.0494), APL3 (0.0636), APL4 (0.0762) and APL5 (0.123). The coefficients for *Experience* were positive and highly significant across all consumption quintiles, namely, BPL (0.022), APL1 (0.0189), APL2 (0.0248), APL3 (0.0297), APL4 (0.0332), and APL5 (0.0457). The coefficients for *Experiencesqr* were negative and significant for all consumption quintiles, signifying an inverted U-shaped relationship. The coefficients for *Networking Intensity* were positive and significant for BPL (0.069), APL1 (0.0519), APL2 (0.0463), APL3 (0.0285), APL4 (0.0281), and negative for APL5 (- 0.0218).

5.5.4. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) - IHDS II

Table 5. 8: Heckman two-step selection model estimates (two-step estimator) by consumption quintiles

Outcome Equation Dependent Variable: lnHourly_Earnings						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	***APL5***
Education	0.0425***	0.0390***	0.0494***	0.0634***	0.0761***	0.120***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0217***	0.0193***	0.0251***	0.0307***	0.0334***	0.0452***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0707***	0.0520***	0.0460***	0.0286***	0.0281***	-0.0213**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.980***	2.115***	2.068***	1.795***	1.906***	1.733***
	(0.04)	(0.06)	(0.07)	(0.08)	(0.12)	(0.15)
Selection Equation Dependent Variable: Labourforce Participation						
	BPL	***APL1***	***APL2***	***APL3***	***APL4***	***APL5***
Married	0.215***	0.221***	0.0882***	0.0649***	0.0204	-0.0675***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Number of Children	-0.0287***	-0.0271***	-0.0323***	-0.0227***	-0.0202**	0.00807
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of Adults	-0.109***	-0.114***	-0.103***	-0.0961***	-0.0788***	-0.0666***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.424***	0.320***	0.270***	0.172***	0.00542	-0.105***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
IMR	0.317***	0.186***	0.191***	0.358***	0.120	-0.139
	(0.05)	(0.05)	(0.07)	(0.08)	(0.11)	(0.14)
Total obs	22008	13927	14422	14554	14761	15228
Selected obs	12458	7291	6832	6442	5732	5323

Non selected obs	9550	6636	7590	8112	9029	9905
rho	0.526	0.324	0.310	0.506	0.165	-0.184
Lambda (IMR)	0.317	0.186	0.191	0.358	0.120	-0.139
chi2	1236.1	563.6	818.7	1053.7	1210.4	2927.6
P value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table (5.8); we report the estimates from Heckman Selection model with two-step consistent estimator. We observe that the coefficient for *Education*, for BPL (0.0425), APL1 (0.039), APL2 (0.049), APL3 (0.063), APL4 (0.76), and APL5 (0.12). The coefficients were positive and significant for all consumption quintiles. We also observe that the coefficient for *Experience*, for BPL (0.022), APL1 (0.019), APL2 (0.025), APL3 (0.031), APL4 (0.033), and APL5 (0.045). The coefficients were positive and significant for BPL, APL1, APL2, APL3, APL4 and APL5.

The coefficients for Square of Experience were negative and significant for BPL, APL1, APL2, APL3, APL4 and APL5. The coefficient for *Networking Intensity* for BPL (0.071), APL1 (0.052), APL2 (0.046), APL3 (0.0286), APL4 (0.028), and APL5 (-0.0213). The coefficients were positive and significant for all quintiles except APL5. The coefficient for IMR was significant across all consumption quintiles (except APL4 and APL5), indicating the presence of selection bias. The instruments used for the first stage selection equation; were Married, Nchildren, and Nadults.

5.6. PANEL DATA ANALYSIS IHDS I AND II

In this section of the chapter, we proceed towards analyzing merged data from IHDS I & II. The sub-sample selected for analysis in this chapter is restricted to individuals in the age group of 15 to 64 years as per the guidelines of Labour force participation followed by NSSO. Experience (Potential Experience) has been restricted to an upper bound of 45 years. Our analysis consists of descriptive statistics of the selected sample, followed by estimation of Heckman selection model with fixed effects model disaggregated by consumption categories (as described in sub section 5.1)

The econometric model can then be specified as follows:

Selection Equation:

$$\textbf{Labourforce Participation}_{it} = \alpha_0 + \alpha_1\textbf{Married}_{it} + \alpha_2\textbf{NChildren}_{it} + \alpha_3\textbf{Nadults}_{it} + \alpha_4(\textbf{X})_{it} + \varepsilon_{it} \qquad \textbf{(5.5)}$$

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 5.6 and e = stochastic error term.

Main equation:

$$Ln\ Hourly\ Earnings_{it} = \beta_0 + \beta_1 Education_{it} + \beta_2 Experience_{it} + \beta_3 Experience^2_{it} + \beta_4 Networking\ Intensity_{it} + \beta_5 IMR_{it} + \beta_6 Z_{it} + u_{it} \qquad \textbf{(5.6)}$$

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (eq 5.5) (Married, Nchildren, and Nadults), u= error term.

5.6.1. Descriptive Statistics

Table 5. 9: Table of Descriptive Statistics

Variable	Description		Mean	SD	Min	Max	Obs
lnHourly_Earnings	Natural logarithm of hourly earnings observed	overall	3.02	0.82	-0.84	6.92	21413
		between		0.77	-0.84	6.71	15601
		within		0.28	1.10	4.94	1.37
Education	Completed years of education	overall	7.90	4.44	0	16	52134
		between		4.03	0	16	29878
		within		1.63	-0.10	15.90	1.74
Experience	Potential experience: {(age - education - 5) years}	overall	19.51	13.53	0	45	52378
		between		13.62	0	45	29889
		within		2.71	-1.99	41.01	1.75
Experience square	Square of potential experience	overall	563.77	629.60	0	2025	52378
		between		627.80	0	2025	29889
		within		138.98	-446.7	1574.27	1.75
Networking intensity	Number of social or economic committees that household members were associated with	overall	0.71	1.08	0	7	52225
		between		0.90	0	7	29870
		within		0.62	-2.79	4.21	1.75
Labor force participation	Binary variable ('1' if hourly earnings are observed, '0' otherwise)	overall	0.41	0.49	0	1	52378
		between		0.43	0	1	29889
		within		0.25	-0.09	0.91	1.75

Married	Binary variable ('1' if the individual is married, '0' otherwise)	overall	0.62	0.49	0	1	52378
		between		0.46	0	1	29889
		within		0.18	0.12	1.12	1.75
Children	Number of children in the household	overall	1.54	1.62	0	18	52378
		between		1.41	0	15	29889
		within		0.84	-4.96	8.04	1.75
Adults	Number of adults in the household	overall	3.47	1.74	0	18	52378
		between		1.50	0	18	29889
		within		0.90	-4.53	11.47	1.75
Consumption quintiles	'0' if below the poverty line, '1' if the lowest quintile above the poverty line, and '5' if the highest quintile above the poverty line	overall	2.66	1.75	0	5	52365
		between		1.54	0	5	29889
		within		0.89	0.16	5.16	1.75

Source: Author's calculations based on IHDS I and II

We found that the (natural log of) average hourly earnings of an individual (overall) was 3.02 with a standard deviation of 0.8. Average education was 7.9 years, with a standard deviation of 4.4 years (overall). The average experience was 19.5 years, with a standard deviation of 13.5 years (overall). Of the total sample, 62% were married. The average number of children was 1.54, and the average number of adults was 3.47.

5.6.2. Heckman Two-Step Correction estimates

Table 5. 10: Estimates from the Fixed-effects Heckman Selection Model by Consumption Quintiles (dependent variable: lnHourly_Earnings).

Main Equation	BPL	APL1	APL2	APL3	APL4	APL5
Education	0.0241***	0.0660***	0.0489***	0.0878***	0.0786***	0.0743***
	(0.0089)	(0.0144)	(0.0129)	(0.0213)	(0.0240)	(0.0200)
Experience	0.00108	0.00763	0.0443**	0.00894	0.0206	0.132***
	(0.0163)	(0.0282)	(0.0190)	(0.0296)	(0.0276)	(0.0188)
Experience square	0.000292	0.0000994	-0.000198	0.000260	0.000595	-0.00178***
	(0.0003)	(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0004)
Networking Intensity	0.0336	0.113***	0.0359	-0.0492	-0.0282	0.0230
	(0.0217)	(0.0388)	(0.0522)	(0.0423)	(0.0389)	(0.0277)
2012	0.246***	-0.0550	-0.182	-0.0564	-0.149	-0.237*
	(0.0609)	(0.0830)	(0.1242)	(0.1598)	(0.1749)	(0.1298)
IMR						
2005	0.0319	-0.321***	-0.469***	-0.414***	-0.783***	0.0866
	(0.1047)	(0.1085)	(0.1070)	(0.0808)	(0.1033)	(0.1103)
2012	-0.0308	-0.217	-0.473***	-0.625***	-1.013***	0.199
	(0.1446)	(0.1730)	(0.1770)	(0.2053)	(0.2271)	(0.1589)
Constant	1.810***	2.219***	2.640***	2.482***	3.064***	1.451***
	(0.1319)	(0.1389)	(0.1243)	(0.1157)	(0.1779)	(0.1935)

Source: Author's calculations based on IHDS I and II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.0

In table (5.10) the coefficient for Education, for BPL (0.0241), APL1 (0.066), APL2 (0.0489), APL3 (0.0878), APL4 (0.0786), and APL5 (0.0743). The coefficients were positive and significant for BPL, APL1, APL2, APL3, APL4 and APL5. We also observe that the coefficient for Experience, for APL2 (0.0443), and APL5 (0.132). The coefficients were positive and significant only for APL2 and APL5.

The coefficients for the Square of Experience were negative and significant for APL5. The coefficient for Networking Intensity for APL1 (0.113) was positive and significant.

The coefficient for IMR (Inverse Mills Ratio) was significant (except BPL and APL5), indicating the presence of selection bias. Hence, it justifies the use of Heckman's Two Step Correction model to address selection bias. The instruments used for the first stage probit model; Married, Nchildren, and NAdults were also significant in the first stage equation.

We find that the returns to an additional year of education are lowest for BPL and highest for APL3. We find that the returns to an additional year of experience are lowest for APL2 and highest for APL5 (Not significant for BPL and other APL quintiles). We also find that an inverted U-shaped relationship exists between earnings and experience only for APL5. Last, we find that social capital, as measured by networking intensity, bears a positive impact only on APL1 and was not significant for the others.

5.6.3. Selection Equation Estimates

Table 5. 11: Selection equation estimates from Fixed Effects- Heckman Selection model by Consumption Quintiles (dependent variable: Labourforce Participation).

Selection Equation	BPL	APL1	APL2	APL3	APL4	APL5
Year Specific Education						
2005	0.0144	0.00208	0.0850**	0.112***	0.0788***	0.148***
	(0.0215)	(0.0309)	(0.0356)	(0.0379)	(0.0295)	(0.0300)
2012	-0.0110	-0.0451	-0.0965***	-0.0607	-0.0894***	-0.0811***
	(0.0253)	(0.0351)	(0.0318)	(0.0409)	(0.0256)	(0.0286)
Year Specific Experience						
2005	0.158***	0.0584	0.0115	0.0685	0.0734*	0.115***
	(0.0478)	(0.0475)	(0.0533)	(0.0539)	(0.0435)	(0.0335)
2012	0.0717	0.0340	0.0377	0.0331	0.0520	0.0463*
	(0.0509)	(0.0591)	(0.0537)	(0.0369)	(0.0417)	(0.0281)
Year Specific Experience_square						
2005	-0.003***	-0.00179*	-0.000702	-0.00200**	-0.00111	-0.00187***
	(0.0009)	(0.0010)	(0.0011)	(0.0010)	(0.0009)	(0.0006)
2012	-0.00137	-0.000130	-0.000396	-0.000126	-0.00109	-0.00178***
	(0.0009)	(0.0011)	(0.0012)	(0.0007)	(0.0008)	(0.0006)
Year Specific Networking Intensity						
2005	-0.0216	-0.0761	0.00494	0.0172	0.0248	0.0416
	(0.0707)	(0.1048)	(0.0783)	(0.0770)	(0.0655)	(0.0541)
2012	0.0679	-0.134	-0.0567	0.0607	0.0995	0.00927
	(0.0727)	(0.1179)	(0.0813)	(0.0673)	(0.0642)	(0.0403)
Year Specific Married						
2005	-0.863***	-0.673**	-0.797***	-1.397***	-0.226	-0.482***
	(0.2525)	(0.3002)	(0.2830)	(0.2843)	(0.2619)	(0.1776)
2012	1.112***	1.594***	0.525**	0.876***	0.338	0.375*
	(0.2060)	(0.2640)	(0.2443)	(0.2970)	(0.2339)	(0.1967)
Year Specific Nchildren						
2005	-0.0131	0.0554	-0.122*	-0.0387	-0.00130	-0.0132
	(0.0402)	(0.0634)	(0.0717)	(0.0720)	(0.0924)	(0.0471)
2012	-0.0520	-0.0423	-0.0176	-0.0178	0.00107	-0.0791
	(0.0424)	(0.0603)	(0.0664)	(0.0635)	(0.0658)	(0.0500)
Year Specific Nadults						
2005	0.00586	-0.120*	-0.0186	0.0726	-0.0965	-0.00557

	(0.0428)	(0.0723)	(0.0797)	(0.0608)	(0.0661)	(0.0517)
2012	-0.0579	-0.00609	-0.0124	-0.0438	0.0314	0.125***
	(0.0366)	(0.0531)	(0.0594)	(0.0529)	(0.0553)	(0.0419)
Year-Specific Individual Average Education						
2005	0.0141	0.0113	-0.0611*	-0.0707*	-0.0219	-0.0822***
	(0.0226)	(0.0305)	(0.0368)	(0.0386)	(0.0302)	(0.0304)
2012	0.00627	0.0305	0.0878***	0.0531	0.0944***	0.135***
	(0.0238)	(0.0352)	(0.0327)	(0.0407)	(0.0254)	(0.0292)
Year-Specific Individual Average Experience						
2005	-0.0434	0.0705	0.118**	0.0557	0.0540	0.0259
	(0.0479)	(0.0483)	(0.0557)	(0.0544)	(0.0444)	(0.0351)
2012	-0.0160	0.0139	0.00607	-0.00201	-0.0000159	0.00404
	(0.0502)	(0.0626)	(0.0548)	(0.0363)	(0.0427)	(0.0272)
Year Specific Individual Average Experience_square						
2005	0.00196**	-0.000456	-0.00156	-0.0000140	-0.00107	-0.000638
	(0.0009)	(0.0010)	(0.0012)	(0.0010)	(0.0009)	(0.0006)
2012	0.000247	-0.000958	-0.000657	-0.000702	-0.000114	0.000680
	(0.0009)	(0.0012)	(0.0012)	(0.0007)	(0.0008)	(0.0006)
Year-Specific Individual Average Networking Intensity						
2005	-0.0274	0.0507	0.0387	-0.0205	-0.0233	-0.0263
	(0.0744)	(0.1055)	(0.0751)	(0.0811)	(0.0682)	(0.0604)
2012	-0.0496	0.104	0.0900	0.0000405	-0.0664	0.0109
	(0.0746)	(0.1231)	(0.0877)	(0.0668)	(0.0648)	(0.0433)
Year Specific Individual Average Married						
2005	0.886***	0.623**	0.652**	1.194***	0.104	0.477***
	(0.2379)	(0.3115)	(0.2927)	(0.3017)	(0.2649)	(0.1792)
2012	-0.943***	-1.488***	-0.369	-0.774**	-0.368	-0.490**
	(0.2068)	(0.2909)	(0.2527)	(0.3023)	(0.2423)	(0.2049)
Year Specific Individual Average Nchildren						
2005	-0.00616	-0.114*	0.0852	0.0439	-0.0458	0.0136
	(0.0413)	(0.0657)	(0.0755)	(0.0751)	(0.0933)	(0.0553)

2012	0.0106	0.00555	-0.0617	-0.0650	-0.0994	0.00196
	(0.0469)	(0.0600)	(0.0651)	(0.0662)	(0.0677)	(0.0501)
Year Specific Individual Average Nadults						
2005	-0.120***	0.0108	-0.0790	-0.206***	-0.00869	-0.120**
	(0.0429)	(0.0714)	(0.0842)	(0.0634)	(0.0654)	(0.0564)
2012	-0.0458	-0.0943*	-0.0702	-0.0423	-0.0955*	-0.212***
	(0.0401)	(0.0536)	(0.0591)	(0.0557)	(0.0561)	(0.0434)
Year						
2012	1.201***	1.360***	1.565***	1.724***	1.640***	1.423***
	(0.1160)	(0.1509)	(0.1591)	(0.1368)	(0.1353)	(0.1363)
Constant	-1.141***	-1.171***	-1.472***	-1.585***	-1.874***	-2.152***
	(0.0856)	(0.0961)	(0.1133)	(0.1073)	(0.0970)	(0.1155)
N	8855	6694	8019	8501	9435	10478

Source: Author's calculations based on IHDS I and II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.0

The table (5.11) summarizes the results of the first stage selection equation of the Heckman model with fixed effects. In addition to the instruments used to estimate labor force participation, the model also estimates interactions between these instruments and other covariates by year. The instruments used for the first stage probit model; Married, Nchildren, and NAdults were also significant in the first stage equation. These estimates were jointly significant across the model.

5.7. PROJECTED EARNINGS BY CLASS WITH EDUCATION AND EXPERIENCE AS COVARIATES

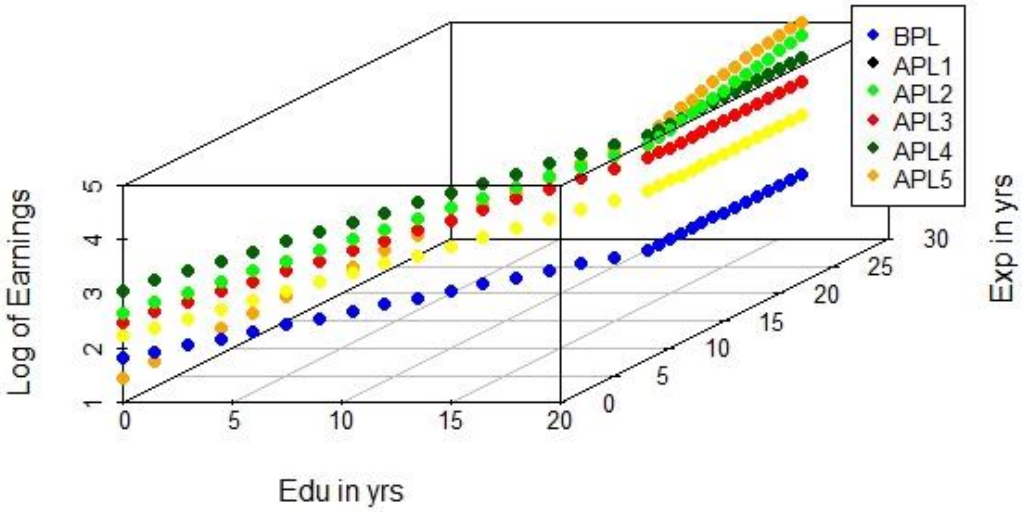

Figure 5. 1: Projected earnings by class with education and experience as covariates.

Our results demonstrate that education is significant and positive for all categories of individuals—economic strata. When we compare returns to education by consumption quintiles (Figure 5.1), the values go from 2.4% (BPL) to 8.8% (APL3). Our projections illustrate divergence in projected earnings across consumption quintiles after accounting for education and experience. We believe this is evidence of disparities between consumption groups. The low private returns for BPL suggest that there is a possibility that certain thresholds must be met to realize higher returns to investment in education. Additionally, the highest private returns for APL3 are suggestive of an inverted u pattern. This kind of analysis to explore trends in returns to an additional year of education and experience across consumption quintiles is, to our knowledge, the first in the Indian context to use panel data and address selection bias.

5.8. FINDINGS

IHDS I:

The estimates for private returns to an additional year of education, in the OLS model range between the lowest for APL1 (5.3%) and highest for APL5(13%). The private returns to an additional year of experience in the OLS model range between the lowest for BPL (2%) and highest for APL5(6.2%). Estimates for Networking Intensity in the OLS model range between the lowest of APL5 (- 4.6%) and the highest for BPL(4%).

The private returns to an additional year of education in the Heckman model (MLE) range between the lowest BPL (5.2%) and highest for APL5(12.9%). The private returns to an additional year of experience in the Heckman model (MLE) range between the lowest for APL1 (2.6%) and the highest for APL5(6.4%). Estimates for networking intensity in the Heckman model (MLE) range between the lowest of APL5 (- 4.6%) and the highest for BPL(3.3%). The overall models are significant across quintiles, with a significant Chi-squared value. The model also estimates IMR.

The private returns to an additional year of education in the Heckman model (TSE) range between the lowest BPL (5 %) and highest for APL5(13%). The private returns to an additional year of experience in the Heckman model (TSE) range between the lowest for BPL and APL1 (2.6%) and the highest for APL5(6.5%). Estimates for networking intensity in the Heckman model (TSE) range between the lowest of APL5 (- 4.8%) and the highest for BPL(3.6%). The overall models are significant across quintiles, with a significant Chi-squared value. The model also estimates IMR. The coefficient for IMR was significant across all consumption quintiles, indicating the presence of selection bias. The instruments used for the first stage selection equation; were Married, Nchildren, and Nadults.

IHDS II:

The estimates for private returns to an additional year of education, in the OLS model range between the lowest for APL1 (3.9%) and highest for APL5 (12%). The private returns to an additional year of experience in the OLS model range between the lowest for BPL (1.9%) and highest for APL5(4.5%). Estimates for Networking Intensity in the OLS model range between the lowest of APL5 (- 2.1%) and the highest for BPL(7.3%). The private returns to an additional year of education in the Heckman model range between the lowest APL1 (3.9%) and highest for APL5(12.3%). The private returns to an additional year of experience in the Heckman model range between the lowest for APL1 (1.9%) and the highest for APL5(4.6%). Estimates for networking intensity in the Heckman model range between the lowest of APL5 (- 2.2%) and the highest for BPL(6.9%). The overall models are significant across quintiles, with a significant Chi-squared value. The model also estimates IMR.

We find that the returns to an additional year of education are lowest for APL1, followed by BPL, and highest for APL5, followed by APL4. We find that the returns to an additional year of experience are lowest for APL1 and highest for APL5. We also find that an inverted U-shaped relationship exists between earnings and experience for all quintiles. Last, we find that social capital, as measured by networking intensity, bears a positive impact on BPL and APL1 to APL4 and a negative impact on APL5.

Panel Data:

We find that the returns to an additional year of education are lowest for APL1, followed by BPL, and highest for APL5, followed by APL4. We find that the returns to an additional year of experience are lowest for BPL & APL1 and highest for APL5. We also find that an inverted U-shaped relationship exists between earnings and experience for all quintiles. Last, we find that social capital, as measured by networking intensity, bears a positive impact on BPL and APL1 and a negative impact on APL3, APL4, and APL5. We also find evidence of selection bias across quintiles.

5.9. CONCLUSION

We believe no study has used a panel data set in the Indian context that has examined the role of education, experience, and social capital in wage determination across consumption quintiles with a nationwide representative sample while integrating individual-level data with household and village attributes and addressing selection bias. These results, therefore, update all earlier estimates in this context.

We conclude that Education has a positive and significant impact on earnings across all quintiles. Experience, too, has positive and significant across consumption quintiles. However, the returns to the lower economic strata are comparatively lower. The squared effect was significant and negative. Networking intensity is significant and positive for lower quintiles. We conclude that private returns to education are lower for those below the poverty line and just above the poverty line. We find a similar trend in private returns to experience. However, social capital has significant positive bearing for the lower strata.

Additionally, our projected earnings from panel estimates lead us to believe that earnings across consumption quintiles provide evidence of economic disparities, though the private returns to education and experience are positive. This is illustrated in figure 5.1., that shows divergence in earnings across consumption quintiles.

The next chapter (Chapter 6) discusses private returns to education disaggregated by gender using the Heckman model. This, we believe will yield results pertaining to the gender wage gap and are crucial in the context of SDG4 and SDG 5 (Education and Gender Equality).

CHAPTER 6: TRENDS IN RETURNS TO HUMAN CAPITAL INVESTMENT ACROSS GENDER

In the previous chapter, we summarized the estimates for returns of an additional year of education by class (consumption quintiles). We used the models derived from Chapter 4 to compare estimates of returns to an additional year of education, experience, and networking intensity across Consumption quintiles (BPL, APL1 to APL5). We further illustrated projected earnings by class to show the divergence between BPL and APL classes. In this context, the next chapter will explore the same idea across genders.

6.1. INTRODUCTION

In this chapter, we will examine private returns to an additional year of education, and experience, and networking intensity across gender. Social groups in the Indian context can broadly be based on gender, caste, and religion. This chapter aims to comparatively analyze trends in the returns to investment in additional years of education by individuals across genders. We estimate the returns to investment in additional years of education using cross-sectional data from the Indian Human Development Survey (IHDS: 2004-05 and 2011-12) and the panel data for these two rounds of IHDS. The chapter summarizes estimates for these returns to private investment on additional years of education using continuous Years of Educational attainment. These models are then estimated using Ordinary Least Squares and the Heckman Model for Correction of Selection Bias.

6.2. FRAMEWORK

For the purpose of establishing a basic framework to achieve the objective mentioned above, this paper uses a modified version of the Mincer earnings function (Jacob Mincer, 1974). The original function, as developed by Jacob Mincer in his book

Schooling, experience and Earnings, has come to become a benchmark for studies in the field of human capital and labor economics, though the basic earning variation by schooling and age has existed since Miller (1955).

6.2.1. Models

We use Two Extended Econometric Models for Mincer's Earnings Function, which can then be stated as follows:

Model 1: Ordinary Least Squares using Continuous Years of Educational attainment by Gender

$$LnY = a_0 + a_1\,Educ + a_2\,Exp + a_3\,Exp^2 + a_4\,(Networkingintensity) + u \qquad (6.1)$$

Where:

LnY = Natural Log of Hourly Earnings

Educ = Years of Schooling

Exp = Potential Experience (Age - Schooling – 5 years)

Exp^2 = Square of Potential Experience

Networking Intensity = Number of social or economic groups households were members of or associated with.

u = Error term

This model is then estimated using the Heckman Model for Correction of selection Bias. The Heckman Model for Correction of selection Bias would then be estimated as follows:

Model 2: The Heckman's Two-Step Model (selection Bias) Continuous Years of Educational Attainment: by Gender

Main Equation:

$$\mathbf{Ln\_Y_i = \beta_0 + \beta_1 Educ_i + \beta_2 Exp_i + \beta_3 Exp^2_i + \beta_4}\ \boldsymbol{Networkingintensity} + \boldsymbol{\beta_5 \lambda} + \mathbf{e_i} \quad \textit{(6.2)}$$

Where:

LnY = Natural Log of Hourly Earnings

Educ = Years of Schooling

Exp = Potential Experience

Exp^2 = Square of Potential Experience

$\boldsymbol{\lambda}$ = Lambda (Inverse Mills Ratio)

e = Error term

Selection Equation:

$$\mathbf{Labour\ Force\ Participation_{it} = \alpha_0 + \alpha_1 Married_{it} + \alpha_2 NChildren_{it} + \alpha_3 Nadults_{it} + \alpha_4 (X)_{it} + v_{it}} \quad \textit{(6.3)}$$

Where:

Labourforce participation = Dummy variable for Labourforce participation

Educ = Years of Schooling

Exp = Potential Experience

Exp^2 = Square of Potential Experience

Z = Selection Instruments from the selection equation 4 (*Married*, *NChildren*, *Nadults*)

X = Set of exogenous explanatory variables of the outcome equation indicated in equation

v = Error term

Equations (6.2) and (6.3) are the main and selection equations, respectively, that are used in the Heckman selection model to be used in this chapter. The model also estimates a value for Lambda (**λ**) (Inverse Mills Ratio) to allow for testing the hypothesis of whether selection bias is an issue in the estimation. A significant value for Lambda would imply the existence of selection Bias and, as such, the use of Heckman's Two Step model for estimation.

6.3. ANALYSIS

6.4. CROSS-SECTIONAL ANALYSIS IHDS I

6.4.1. Ordinary Least Square Estimates- IHDS I

Table 6. 1: Ordinary Least Squares Estimates by Gender for IHDS I

Dependent Variable: lnHourly_Earnings	**Male**	**Female**
Education	0.0929***	0.102***
	(0.00)	(0.00)
Experience	0.0530***	0.0395***
	(0.00)	(0.00)
Experiencesqr	-0.000678***	-0.000480***
	(0.00)	(0.00)
Networking Intensity	0.0165***	-0.00873
	(0.00)	(0.01)
Constant	1.513***	1.343***
	(0.02)	(0.05)
Number of Observations	24421	9592
R square	0.305	0.292
Adj R square	0.305	0.291
F	2203.4	493.0
P value	0.0000	0.0000

Source: Author's calculation using IHDS I Data

Robust standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

We estimated the disaggregated model by two groups – male and female, as information was provided for only these two groups. Table 6.1 provides the estimates by gender. *Education* was positive and significant for both males and females. *Experience* was positive and significant for females and males and the square term (*Experiencesqr*) is negative and significant for males and females. This suggests an inverted U-shaped relation between earnings and experience for both genders. Networking intensity was positive and significant for males only. The R square for Males (0.31) and females (0.29).

6.4.2. Heckman Selection Model (Maximum Likelihood Estimates)- IHDS I

Table 6. 2: Heckman selection model estimates (using maximum likelihood estimator) by gender

Outcome Equation Dependent Variable: lnHourly_Earnings		
	Male	**Female**
Education	0.0919***	0.0782***
	(0.00)	(0.00)
Experience	0.0603***	0.0451***
	(0.00)	(0.00)
Experiencesqr	-0.000794***	-0.000644***
	(0.00)	(0.00)
Networking Intensity	0.0144***	-0.0110**
	(0.00)	(0.00)
Constant	1.169***	0.139***
	(0.06)	(0.05)
Selection Equation Dependent Variable: Labourforce Participation		
Married	0.401***	0.213***
	(0.01)	(0.01)
Number of Children	0.0509***	0.0312***
	(0.00)	(0.00)
Number of Adults	-0.118***	-0.118***
	(0.00)	(0.01)
Constant	-0.229***	-0.863***
	(0.01)	(0.02)
athrho	0.450***	1.489***
	(0.08)	(0.07)
insigma	-0.405***	-0.0721***
	(0.02)	(0.03)
Total obs	60030	61435
Selected obs	24421	9592
Non selected obs	35609	51843
rho	0.422	0.903
Lambda (IMR)	0.281	0.840
chi2	8692.4	405.4
P value	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 6.2 provides estimates from Heckman (MLE) model. *Education* was positive and significant for both males and females. *Experience* was positive and significant for females and males and the square term (*Experiencesqr*) is negative and significant for males and females. This suggests an inverted U-shaped relation between earnings and experience for both genders. Networking intensity was positive and significant for males, whereas it was negative and significant for females.

The coefficient for *Education* for Males (0.092), and Females (0.78). The coefficient for *Experience* for Males (0.060), and Females (0.045). The coefficient for *Networking Intensity* for Males (0.014), and Females (-0.011). IMR for males (0.28), amd females (0.84). Both models are significant overall.

6.4.3. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) - IHDS I

Table 6. 3: Heckman two-step selection model estimates (two-step estimator) by gender

Outcome Equation Dependent Variable: lnHourly_Earnings		
	Male	**Female**
Education	0.0925***	0.100***
	(0.00)	(0.00)
Experience	0.0587***	0.0442***
	(0.00)	(0.00)
Experiencesqr	-0.000767***	-0.000572***
	(0.00)	(0.00)
Networking Intensity	0.0148***	-0.0118**
	(0.00)	(0.01)
Constant	1.245***	0.906***
	(0.04)	(0.07)
Selection Equation Dependent Variable: Labourforce Participation		
	Male	**Female**
Married	0.410***	0.241***
	(0.01)	(0.02)

Number of Children	0.0468***	0.0400***
	(0.00)	(0.00)
Number of Adults	-0.120***	-0.167***
	(0.00)	(0.00)
Constant	-0.224***	-0.757***
	(0.01)	(0.02)
IMR	0.215***	0.267***
	(0.03)	(0.03)
Total obs	60030	61435
Selected obs	24421	9592
Non selected obs	35609	51843
rho	0.330	0.435
Lambda (IMR)	0.215	0.267
chi2	10766.4	3732.4
P value	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 6.3 provides estimates from Heckman (TSE) model. *Education* was positive and significant for both males and females. *Experience* was positive and significant for females and males and the square term (*Experiencesqr*) is negative and significant for males and females. This suggests an inverted U-shaped relation between earnings and experience for both genders. Networking intensity was positive and significant for males, whereas it was negative and significant for females.

The coefficient for *Education* for Males (0.093), and Females (0.10). The coefficient for *Experience* for Males (0.059), and Females (0.044). The coefficient for *Networking Intensity* for Males (0.015), and Females (-0.012). IMR for males (0.215), and females (0.27) was significant. this significant value of IMR indicates presence of selection bias. Both models are significant overall.

6.5. CROSS-SECTIONAL ANALYSIS IHDS II

6.5.1. Ordinary Least Square Estimates- IHDS II

Table 6. 4: Ordinary Least Squares Estimates by Gender for IHDS II

Dependent Variable: lnHourly_Earnings	Male	Female
Education	0.0726***	0.0794***
	(0.00)	(0.00)
Experience	0.0341***	0.0352***
	(0.00)	(0.00)
Experiencesqr	-0.000364***	-0.000383***
	(0.00)	(0.00)
Networking Intensity	0.0522***	0.0134***
	(0.00)	(0.01)
Constant	2.054***	1.693***
	(0.02)	(0.04)
Number of Observations	33000	13488
R square	0.206	0.199
Adj R square	0.206	0.199
F	1614.6	449.8
P value	0.0000	0.0000

Source: Author's calculation using IHDS II Data
Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

We estimated the disaggregated model by two groups – male and female, as information was provided for only these two groups. Table 6.4 provides the estimates by gender. *Education* was positive and significant for both males and females. *Experience* was positive and significant for females and males and the square term (*Experiencesqr*) is negative and significant for males and females. This suggests an inverted U-shaped relation between earnings and experience for both genders. Networking intensity was positive and significant for males and females. The R square for Males (0.21) and females (0.20).

6.5.2. **Heckman Selection Model (Maximum Likelihood Estimates)- IHDS II**

Table 6. 5: Heckman selection model estimates (using maximum likelihood estimator) by gender

Outcome Equation Dependent Variable: lnHourly_Earnings		
	Male	**Female**
Education	0.0721***	0.0790***
	(0.00)	(0.00)
Experience	0.0354***	0.0355***
	(0.00)	(0.00)
Experiencesqr	-0.000386***	-0.000391***
	(0.00)	(0.00)
Networking Intensity	0.0511***	0.0122**
	(0.00)	(0.01)
Constant	1.962***	1.602***
	(0.02)	(0.05)
Selection Equation Dependent Variable: Labourforce Participation		
	Male	**Female**
Married	0.124***	-0.119***
	(0.01)	(0.01)
Number of Children	0.0292***	0.0131***
	(0.00)	(0.00)
Number of Adults	-0.140***	-0.191***
	(0.00)	(0.00)
Constant	0.655***	-0.0386**
	(0.02)	(0.02)
athrho	0.203***	0.121**
	(0.03)	(0.05)
insigma	-0.427***	-0.482***
	(0.01)	(0.01)
Total obs	53877	57792
Selected obs	33000	13488
Non selected obs	20877	44304
rho	0.200	0.120
Lambda (IMR)	0.131	0.0742
chi2	6356.9	1736.1
P value	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 6.5 provides estimates from Heckman (MLE) model. *Education* was positive and significant for both males and females. *Experience* was positive and significant for females and males and the square term (*Experiencesqr*) is negative and significant for males and females. This suggests an inverted U-shaped relation between earnings and experience for both genders. Networking intensity was positive and significant for males and females.

The coefficient for *Education* for Males (0.072), and Females (0.79). The coefficient for *Experience* for Males (0.035), and Females (0.035). The coefficient for *Networking Intensity* for Males (0.051), and Females (0.012). IMR for males (0.13), and females (0.74). Both models are significant overall.

6.5.3. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) - IHDS II

Table 6. 6: Heckman two-step selection model estimates (two-step estimator) by gender

Outcome Equation Dependent Variable: lnHourly_Earnings		
	Male	**Female**
Education	0.0720***	0.0790***
	(0.00)	(0.00)
Experience	0.0358***	0.0355***
	(0.00)	(0.00)
Experiencesqr	-0.000392***	-0.000391***
	(0.00)	(0.00)
Networking Intensity	0.0508***	0.0122**
	(0.00)	(0.00)
Constant	1.931***	1.603***
	(0.03)	(0.04)
Selection Equation Dependent Variable: Labourforce Participation		
	Male	**Female**
Married	0.123***	-0.122***
	(0.01)	(0.01)
Number of Children	0.0250***	0.0119***
	(0.00)	(0.00)
Number of Adults	-0.140***	-0.190***
	(0.00)	(0.00)
Constant	0.660***	-0.0357**
	(0.02)	(0.02)

IMR	0.174***	0.0738***
	(0.03)	(0.03)
Total obs	53877	57792
Selected obs	33000	13488
Non selected obs	20877	44304
rho	0.264	0.119
Lambda (IMR)	0.174	0.0738
chi2	8485.6	3254.1
P value	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 6.6 provides estimates from Heckman (TSE) model. *Education* was positive and significant for both males and females. *Experience* was positive and significant for females and males and the square term (*Experiencesqr*) is negative and significant for males and females. This suggests an inverted U-shaped relation between earnings and experience for both genders. Networking intensity was positive and significant for males, and females.

The coefficient for *Education* for Males (0.072), and Females (0.79). The coefficient for *Experience* for Males (0.036), and Females (0.036). The coefficient for *Networking Intensity* for Males (0.050), and Females (0.012). IMR for males (0.17), and females (0.738) was significant. this significant value of IMR indicates presence of selection bias. Both models are significant overall.

6.6. PANEL DATA ANALYSIS IHDS I AND II

6.6.1. Descriptive Statistics

Table 6. 7: Descriptive Statistics:

Variable	Description		Mean	Std Dev	Min	Max	Obs.
ln_Hourly _Earnings	Natural logarithm of hourly earnings observed	overall	3.02	0.82	-0.84	6.92	21413
		between		0.77	-0.84	6.71	15601
		within		0.28	1.10	4.94	1.37
Education	Completed years of education	overall	7.90	4.44	0	16	52134
		between		4.03	0	16	29878
		within		1.63	-0.10	15.90	1.74
Experience	Potential experience: {(Age - Education - 5) Years}	overall	19.51	13.53	0	45	52378
		between		13.62	0	45	29889
		within		2.71	-1.99	41.01	1.75
Experiencesqr	Square of potential experience	overall	563.77	629.60	0	2025	52378
		between		627.80	0	2025	29889
		within		138.98	-446.73	1574.27	1.75
Networking Intensity		overall	0.71	1.08	0	7	52225
		between		0.90	0	7	29870
		within		0.62	-2.79	4.21	1.75
Lf_Participation	Labor force participation dummy variable ('1' if hourly earnings are observed, '0' otherwise)	overall	0.41	0.49	0	1	52378
		between		0.43	0	1	29889
		within		0.25	-0.09	0.91	1.75
Married	Married dummy variable ('1' if the	overall	0.62	0.49	0	1	52378
		between		0.46	0	1	29889

	individual is married, '0' otherwise)	within		0.18	0.12	1.12	1.75
Nchildren	Number of children in the household	overall	1.54	1.62	0	18	52378
		between		1.41	0	15	29889
		within		0.84	-4.96	8.04	1.75
Nadults	Number of adults in the household	overall	3.47	1.74	0	18	52378
		between		1.50	0	18	29889
		within		0.90	-4.53	11.47	1.75
COPC_BPL_APL	Consumption quintiles: '0' Below the poverty line, '1' Above the poverty line lowest quintile, and '5' Above the poverty line highest quintile	overall	2.66	1.75	0	5	52365
		between		1.54	0	5	29889
		within		0.89	0.16	5.16	1.75
Female	A female dummy variable ('1' if the individual is female, '0' otherwise)	overall	0.39	0.49	0	1	52378
		between		0.49	0	1	29889
		within		0.03	-0.11	0.89	1.75
Caste	Caste Categories: '1' General/Forward/Others, '2' Other Backward Classes, '3' Scheduled Castes and '4' Scheduled Tribes	overall	2.11	1.10	1	4	52310
		between		0.81	1	4	29888
		within		0.78	0.61	3.61	1.75
Religion	Religion Categories: '1' Hindu, '2' Other Minorities, '3' Muslims	overall	1.33	0.69	1	3	52378
		between		0.70	1	3	29889
		within		0.10	0.33	2.33	1.75

Source: Author's calculation from IHDS I and II

In table 6.7 we find that the (natural log of) average hourly earnings of an individual (overall) was 3.02 with a standard deviation of 0.8 (see Table 6.4.3.1). Average education was 7.9 years with a standard deviation of 4.4 years (overall), and the average experience was 19.5 years with a standard deviation of 13.5 years (overall). Of the entire sample, 62% were married, and 39% were female. The average number of children was 1.54 and adults was 3.47.

6.6.2. Composition of Gender in IHDS Panel Data

Table 6. 8: Composition of Gender in IHDS Panel Data

	Overall		Between		Within
Gender Categories Panel Data	Frequency	Percent	Frequency	Percent	Percent
Male	32079	61.25	18360	61.43	99.77
Female	20299	38.75	11615	38.86	99.63
Total	52378	100.00	29975	100.29	99.71

Source: Author's calculation from IHDS I and II

Table 6.8 provides the frequency and percentage of observations from the specified sample for males and females. We observe that overall, males comprise 61.25 % of the sample, and females comprise 38.75 percent of the sample.

6.6.3. Random And Fixed Effects Model Using Panel Data

Table 6. 9: Random and fixed effects model using panel data

Dependent Variable: ln Hourly Earnings	**Random**	**Fixed**
Education	0.101***	0.0666***
	(0.0013)	(0.0033)
Experience	0.0374***	0.0608***
	(0.0016)	(0.0046)
Experiencesqr	-0.000381***	-0.000575***
	(0.0000)	(0.0001)
Networking Intensity	0.0191***	0.0174**
	(0.0044)	(0.0074)
Constant	1.656***	1.570***
	(0.0198)	(0.0508)
Observations	21291	21291
R-Square	0.191	0.207

Source: Author's calculation from IHDS I and II
Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 6.9, illustrates the estimates from the generalized estimator with random effects and the within estimator with fixed effects. This summarizes the coefficients obtained from the two estimators. Next, we conduct a Hausman test to choose between the random and fixed effect models.

6.6.4. Hausman Test

Table 6. 10: Estimates for comparison of random and fixed effects model

Variables	Fixed	Random	Difference	S.E.
Education	0.0666	0.1015	-0.0349	0.0027
Experience	0.0608	0.0374	0.0234	0.0042
Experiencesqr	-0.0006	-0.0004	-0.0002	0.0001
Networking Intensity	0.0174	0.0191	-0.0018	0.0053

Source: Authors' calculations using IHDS panel data

Test: Ho: Differences in coefficients are not systematic

chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 186.29

Prob>chi2 = 0.0000

The results from the Hausman test (table 6.10)provide a chi-square value of 186.29 (p-value 0.0000). Indicating the significance of the fixed effects model over the random effects model. This indicates rejection of the null hypothesis (of RE), so we accept the alternate hypothesis that the FE model is better suited to our data.

6.6.5. Estimates From The Fixed Effects Model By Gender

We then estimate the fixed effects model by Gender using the within estimator to discuss the returns to additional years of education across genders.

Table 6. 11: Estimates for Fixed effects model

Dependent Variable: ln Hourly Earnings	Male	Female
Education	0.0670***	0.0632***
	(0.0036)	(0.0081)
Experience	0.0642***	0.0365**
	(0.0049)	(0.0144)
Experiencesqr	-0.000642***	-0.0000830
	(0.0001)	(0.0003)
NetworkingIntensity	0.0178**	0.0163
	(0.0079)	(0.0198)
Constant	1.600***	1.537***
	(0.0539)	(0.1534)
Number of Observations	17225	4066
R square within	0.214	0.179

Source: Authors' calculations using IHDS panel data
Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

We estimated the model by two groups – male and female, as information was provided for only these two groups (table 6.11). *Education* was positive and significant for both males and females. *Experience* was positive and significant for females and males and the square term (*Experiencesqr*) is negative and significant for males (jointly significant for females). This suggests an inverted U-shaped relation between earnings and experience for males. Networking intensity was positive and significant for males. The R square is within for Males (0.21) and females (0.17). However, the literature suggests that models relying solely on traditional estimation methods suffer from at least two sources of bias. Namely selection bias and Endogeneity bias. Next,, we estimate the Heckman selection model with fixed effects for correction of selection bias. Observe that the number of observations is restricted to individuals whose hourly earnings are observed, hence raising the question of selection bias.

6.6.6. Heckman Selection Model With Fixed Effects By Gender

We then estimate the fixed effects of the Heckman selection model (xtheckmanfe) using Stata 16.1 to account for selection bias in labor force participation.

Table 6. 12: Estimates for Heckman Selection model with fixed effects by Gender

Dependent Variable: In Hourly Earnings	Male	Female
Education	0.0369***	0.0524***
	(0.0052)	(0.0113)
Experience	0.00831	0.0552***
	(0.0067)	(0.0192)
Experiencesqr	0.0000166	-0.000717**
	(0.0001)	(0.0004)
Networking Intensity	0.0220***	0.00771
	(0.0084)	(0.0159)
2012. year	0.232***	0.383**
	(0.0473)	(0.1602)
IMR 2005	-0.324***	0.247*
	(0.0513)	(0.1311)
IMR 2012	-0.449***	0.0431
	(0.1063)	(0.1865)
Constant	1.687***	0.709***
	(0.0541)	(0.2027)

Source: Author's calculation from IHDS I and II
Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

We estimated the model by two groups – male and female, as information was provided for only these two groups. *Education* was positive and significant for both males and females. *Experience* was positive and significant for females (jointly significant for males) and the square term (*Experiencesqr*) is negative. This suggests an inverted U-shaped relation between earnings and experience for females. Networking intensity was positive and significant for males. The IMR (Inverse Mills Ratio) is significant for both groups in 2005 and for males in 2012, indicating the presence of selection bias.

Selection equation results are reported below.

6.6.7. Estimates For Selection Equation

Table 6. 13: Estimates for selection equation

Selection Equation	Male	Female
Year Specific Education		
2005	0.0154	0.0286
	(0.0119)	(0.0200)
2012	-0.0153	0.0363**
	(0.0096)	(0.0143)
Year Specific Experience		
2005	-0.0110	-0.0387
	(0.0181)	(0.0241)
2012	0.0677***	0.105***
	(0.0113)	(0.0169)
Year Specific Experience_square		
2005	-0.000000567	0.000193
	(0.0003)	(0.0004)
2012	-0.00144***	-0.00138***
	(0.0002)	(0.0003)
Year Specific Networking Intensity		
2005	-0.0373**	-0.0109
	(0.0147)	(0.0285)
2012	0.0258	0.0269
	(0.0209)	(0.0226)
Year Specific Married		
2005	-0.0546	-0.454***
	(0.0676)	(0.1387)
2012	0.186***	-0.234*
	(0.0568)	(0.1237)
Year Specific Nchildren		
2005	-0.0759***	-0.00574
	(0.0168)	(0.0249)
2012	0.00219	-0.00681
	(0.0142)	(0.0210)
Year Specific Nadults		
2005	-0.0182	-0.0515*

	(0.0154)	(0.0275)
2012	-0.0271**	-0.0480**
	(0.0115)	(0.0194)
Year-Specific Individual Average Education		
2005	-0.0315**	-0.0413*
	(0.0134)	(0.0226)
2012	-0.0214**	-0.0437***
	(0.0100)	(0.0136)
Year-Specific Individual Average Experience		
2005	0.164***	0.156***
	(0.0212)	(0.0251)
2012	-0.00994	-0.0595***
	(0.0131)	(0.0170)
Year Specific Individual Average Experience_square		
2005	-0.00274***	-0.00232***
	(0.0003)	(0.0004)
2012	0.0000303	0.000364
	(0.0002)	(0.0003)
Year-Specific Individual Average Networking Intensity		
2005	0.0350*	0.0955***
	(0.0180)	(0.0307)
2012	-0.0284	0.103***
	(0.0232)	(0.0289)
Year Specific Individual Average Married		
2005	0.356***	0.160
	(0.0714)	(0.1479)
2012	-0.0722	-0.0911
	(0.0684)	(0.1322)
Year Specific Individual Average Nchildren		
2005	0.0697***	-0.0219
	(0.0192)	(0.0294)
2012	-0.0345**	-0.00926
	(0.0160)	(0.0221)
Year Specific Individual Average Nadults		
2005	-0.130***	-0.0979***
	(0.0151)	(0.0320)
2012	-0.104***	-0.0809***

	(0.0137)	(0.0193)
Year		
2012	1.899***	1.010***
	(0.0870)	(0.1177)
Constant	-1.203***	-1.541***
	(0.0885)	(0.1243)
N	31857	20132

Source: Author's calculation
Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

We find that selection instruments are significant in the first stage equation of the Heckman model.

6.6.8. Comparison Of Fixed Effects Estimates With Heckman Estimates By Gender For IHDS Panel Data

Table 6. 14: Comparison of Fixed effects estimates with Heckman Estimates by Gender for IHDS Panel Data

	Heckman fixed estimates		Fixed effects estimates	
Outcome Equation	**Male**	**Female**	**Male**	**Female**
Dependent Variable: ln Hourly Earnings				
Education	0.0369***	0.0524***	0.0670***	0.0632***
	(0.0064)	(0.0110)	(0.0036)	(0.0081)
Experience	0.00831	0.0552***	0.0642***	0.0365**
	(0.0074)	(0.0178)	(0.0049)	(0.0144)
Experiencesqr	0.0000166	-0.000717**	-0.000642***	-0.00008
	(0.0002)	(0.0003)	(0.0001)	(0.0003)
NetworkingIntensity	0.0220**	0.00771	0.0178**	0.0163
	(0.0097)	(0.0159)	(0.0079)	(0.0198)
Mean of Education	0.0726***	0.0732***		
	(0.0072)	(0.0122)		
Mean of Experience	0.0306***	0.00701		
	(0.0063)	(0.0149)		
Mean of Experiencesqr	-0.000341***	-0.0000592		
	(0.0001)	(0.0002)		
Mean of Networkingintensity	0.0177	-0.00287		
	(0.0119)	(0.0264)		
Mean of Married	-0.117***	-0.160***		
	(0.0241)	(0.0601)		
Mean of NChildren	-0.0283***	-0.0276**		

	(0.0051)	(0.0109)		
Mean of NAdults_	0.0423***	0.00482		
	(0.0081)	(0.0209)		
2012. year	0.232***	0.383**		
	(0.0433)	(0.1678)		
Inverse Mills Ratio				
IMR 2005	-0.324***	0.247*		
	(0.0598)	(0.1370)		
IMR 2012	-0.449***	0.0431		
	(0.1052)	(0.1859)		
Constant	1.687***	0.709***	1.600***	1.537***
	(0.0550)	(0.1955)	(0.0539)	(0.1534)
Number of Observations	31857	20132	17225	4066
R square within	-	-	0.214	0.179

Source: Authors' calculations using IHDS I and II
Robust standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 6.14 summarizes the coefficients from the Within estimator and Heckman estimator with fixed effects by gender in order to discuss the differences obtained by these estimation techniques. While Within estimates for returns to an additional year of education were larger than the Heckman estimates. This is in line with literature that suggests that traditional estimation techniques overestimate the returns to additional years of education due to selection bias. However, my estimates suggest that once labor force participation for women is accounted for, the returns to an additional year of education for women (0.052) are higher than for men (0.037). Also, Experience and Experiencesqr were significant for only women in the labor force; their averages were significant for men. Networking intensity was significant for men. The significance of IMR (Inverse Mills Ratio) confirms the presence of selection bias in the within estimator.

6.7. PROJECTED EARNINGS BY GENDER WITH EDUCATION AND EXPERIENCE AS COVARIATES

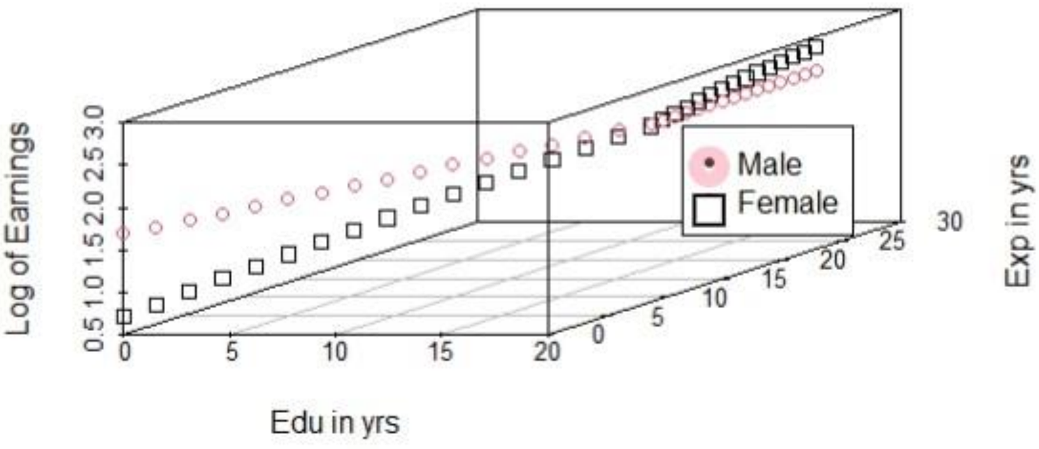

Figure 6. 1: Projected earnings by gender with education and experience as covariates.

Returns to females (5.2%) are much higher than the returns to males (3.7%). The received literature on gender differentials in earnings has universally found that women earn less than men for a given level of education and experience. We did not find evidence contrary to this. However, we did find that there is a tendency towards convergence in earnings between males and females for given levels of education and experience. This has been noted by other authors as well, and it can be attributed in part to structural and technological change (Cortes, Oliveira, & Salomons, 2020). As the workforce has moved from manufacturing to service-oriented employment, women are less likely to be disadvantaged in such work situations. Moreover, their soft skills at the workplace may be more sought after than those of men. Our findings provide evidence of convergence in the long run in the Indian context (Figure 6.1). We found that the initial value of earnings (intercept value) for males is much higher than that of females.

However, the marginal returns to an additional year of education are much higher for women than men.

6.8. FINDINGS

IHDS I :

The private returns to an additional year of *Education* in the OLS model range between the lowest of males (9.29%) and females (10.2%). The private returns to an additional year of *Education* in the Heckman (MLE) model range between the lowest of females (7.8%) and males (9.19%). The private returns to an additional year of *Education* in the Heckman (TSE) model range between the lowest of females (10%) and males (9.3%).

The private returns to an additional year of *Experience* in the OLS model range between the lowest of females (3.95%) and males (5.3%). The private returns to an additional year of *Experience* in the Heckman (MLE) model range between the lowest of females (4.5%) and males (6%). The private returns to an additional year of *Experience* in the Heckman (TSE) model range between the lowest of females (4.4%) and males (5.9%).

The private returns to *Networking Intensity* in the OLS model range between the lowest of females (- 0.8%) (not significant) and males(1.65%). The estimates for *Networking Intensity* in the Heckman (MLE) model range between the lowest of females (- 1.1%) and males(1.4%). The estimates for *Networking Intensity* in the Heckman (TSE) model range between the lowest of females (- 1.2%) and males(1.5%).

IHDS II

The private returns to an additional year of *Education* in the OLS model range between the lowest of males (7.3%) and females (7.9%). The private returns to an additional year of *Education* in the Heckman (MLE) model range between the lowest of males

(7.2%) and females (7.9%). The private returns to an additional year of *Education* in the Heckman (TSE) model range between the lowest of males (7.2%) and females (7.9%).

The private returns to an additional year of *Experience* in the OLS model range between the lowest of males (3.4%) and females (3.5%). The private returns to an additional year of *Experience* in the Heckman (MLE) model range between the lowest of males (3.54%) and females (3.55%). The private returns to an additional year of *Experience* in the Heckman (TSE) model are similar for males (3.6%) and females (3.6%).

The estimates for *Networking Intensity* in the OLS model range between the lowest of females (1.34%) (not significant) and males(5.22%). The estimates for *Networking Intensity* in the Heckman (MLE) model range between the lowest of females (1.22%) and males(5.11%). The estimates for *Networking Intensity* in the Heckman (TSE) model range between the lowest of females (1.2%) and males(5%).

Panel data IHDS I & II:

Education was positive and significant for both males and females. The experience was positive and significant for females and jointly significant for males. The square term (Experiencesqr) was negative. This suggests an inverted U-shaped relation between earnings and experience for females. The implication is that while women's earnings increase with experience, the rate of increase in earnings—caused by a rise in experience—occurs at a decreasing rate. Networking intensity was positive and significant for males. The IMR was significant for both groups in 2005 and for males in 2012. Returns to females (5.2%) are much higher than the returns to males (3.7).

6.9. CONCLUSION

We believe no study has used a panel data set in the Indian context that has examined the role of education, experience, and social capital in wage determination across genders with a nationwide representative sample while integrating individual-level data with household and village attributes and addressing selection bias. These results, therefore, update all earlier estimates in this context.

We conclude that Education and Experience have a positive and significant impact on earnings for men and women. However, Social capital has a significant positive bearing only for men in the panel data model. Though the finding for social capital for women was unanticipated, one possible explanation for the insignificance of social capital for women could be the social barriers to women's participation in the economy. Results for cross-sectional data provide mixed results.

Using the panel estimates we conclude that returns to females (5.2%) are much higher than the returns to males (3.7%), this we believe is evidence on possible long-run convergence in earnings as illustrated in figure 6.1. Our finding are significant in linking SDG4 (quality education) and SDG5 (gender equality).

The next chapter (Chapter 7) discusses private returns to education disaggregated by caste using the Heckman models. This, we believe, will yield insightful results pertaining to socio-economic inequalities. With the need for affirmative action being well understood in the Indian context, the next chapter helps to gain an insight into the private returns to education across castes while examining the role of education, experience, and social networking in breaking social barriers.

CHAPTER 7: TRENDS IN RETURNS TO HUMAN CAPITAL INVESTMENT ACROSS CASTE

In the previous chapter, we saw empirical evidence of earnings differentials across genders. We gained an understanding of the private returns to education for men and women. In this chapter, we provide disaggregated estimates of private return to education by caste (General/ forward and others, Other backward classes (OBC), Scheduled Castes (SC), and Scheduled Tribes (ST)).

7.1. INTRODUCTION

Social groups in the Indian context are broadly based on caste and religion. In this chapter, we try to examine the impact of education as a form of human capital, experience, and social capital (as measured by Networking Intensity) on the hourly earnings of individuals. The chapter aims to comparatively analyze trends in the private returns to investment in additional years of education and experience by individuals across Caste. In the chapter we estimated returns for human capital investment for four caste categories, namely General, Other Backward Castes (OBC), Scheduled Castes (SC), and Scheduled Tribes (ST). The scope of the chapter is restricted to one of the major social groups in the Indian context, namely, caste. Our disaggregated estimates for private returns, we believe, would provide a deeper understanding of the differences in returns to *Education* across castes.

We estimate the returns to investment in additional years of education using cross-sectional and panel data from two rounds of IHDS. The chapter summarizes estimates for these returns to private investment on additional years of education using Mincer's earnings framework. We estimate the OLS, and Heckman models to address selection Bias.

Returns to education across social groups have been studied extensively in the Indian However, most papers use cross-sectional data or pooled data. The chapter overcomes this scarce literature on trends in private returns to investment in additional years of education by providing an in-depth analysis, first using cross-sectional and then from an elaborate panel data set.

7.2. FRAMEWORK

For the purpose of establishing a basic framework to achieve the objective mentioned earlier, this paper uses a modified version of the Mincer earnings function (Jacob Mincer, 1974). The original function, as developed by Jacob Mincer in his book Schooling, experience and Earnings, has come to become a benchmark for studies in the field of human capital and labor economics, though the basic earning variation by schooling and age has existed since Miller (1955)

7.2.1. Models

We use an Extended Econometric Model for Mincer's Earnings Function, which can then be stated as follows:

Model 1: Ordinary Least Squares using Continuous Years of Educational attainment.

$$\boldsymbol{Ln\ Hourly\ Earnings_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \beta_3 Experience_i^2 + \beta_4 Networking\ Intensity_i + u_i} \quad \textbf{(7.1)}$$

Where Ln Hourly Earnings is dependent on Education and Experience and their squares (to test for non-linearity), and u is the stochastic error term. β1 = the average private returns to an additional year of education. However, as pointed out in Chapter 4, the estimation of this suffers from at least two sources of endogeneity biases—namely, selection Bias and ability Bias. Hence, we proceed to estimate our model using Heckman's Two-step procedure to address selection bias.

The Heckman Model then estimates this model for the Correction of selection Bias. The Heckman Model for Correction of selection Bias would then be estimated as follows:

Model 2: The Heckman's selection model by Religion

First stage equation:

$$\boldsymbol{Labourforce\ Participation_i = a_0 + a_1 Married_i + a_2 Nchildren_i + a_3 Nadults_i + a_4 X_i + e_i} \quad \textbf{(7.2)}$$

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 7.3 and ε = stochastic error term.

Second stage equation:

$$\boldsymbol{Ln\ Hourly\ Earnings_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \beta_3 Experience_i^2 + \beta_4 Networking\ Intensity_i + \beta_5 IMR_i + \beta_6 Z_i + u_i} \quad \textbf{(7.3)}$$

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (7.2) (Married, Nchildren, and Nadults), u= error term.

The model also estimates a value for IMR to allow for testing the hypothesis of whether selection bias is an issue in the estimation. A significant value for Lambda would imply the existence of selection Bias and, as such, the use of Heckman's Two Step model for estimation.

7.3. ANALYSIS

7.4. CROSS-SECTIONAL ANALYSIS IHDS I

7.4.1. Descriptive Statistics IHDS I

Table 7. 1: Descriptive statistics:

Variables	Description	Obs	Mean	Std. Dev.	Min	Max
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	35,519	2.70	0.78	-2.22	7.25
Education	Completed years of education	214,399	4.67	4.68	0.00	15.00
Experience	Potential experience: {(Age - Education - 5) Years}	269,335	22.34	18.53	0.00	45.00
Experiencesqr	Square of potential experience	269,335	842.53	878.68	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups households were members of or associated with.	179,682	0.61	1.02	0.00	7.00
Age	Age in years	215,754	27.35	19.35	0.00	116.00
Labourforce Participation	Labor force participation dummy variable ('1' if hourly earnings are observed, '0' otherwise)	269,335	0.18	0.38	0.00	1.00
Married	Married dummy variable ('1' if the individual is married, '0' otherwise)	269,335	0.36	0.48	0.00	1.00

Number of Children	Number of children in the household	215,754	2.20	1.86	0.00	17.00
Number of Adults	Number of adults in the household	215,754	3.27	1.72	0.00	18.00
Caste Categories:	Caste Categories: '1'General/Forward/Others, '2' Other Backward Classes, '3' Scheduled Castes and '4' Scheduled Tribes	180,287	2.16	1.26	1.00	4.00

Source: Author's calculation using IHDS I Data

Table 7.1 provides the summary statistics for IHDS I. We found that an individual's (natural log of) average hourly earnings were 2.70 with a standard deviation of 0.78. Average education was 4.67 years, with a standard deviation of 4.7 years. The average experience was 22.34 years, with a standard deviation of 18.5 years. Of the total sample, 36% were married. The average number of children was 2.2, and the adults were 3.3. Caste Categories comprised of '1' General/Forward castes (Including Brahmins and others), '2' Other Backward castes, '3' Scheduled castes, and '4' Scheduled tribes.

Next, we proceed to estimate the models.

7.4.2. Ordinary Least Squares estimates – IHDS I

Table 7. 2: OLS model estimates disaggregated by caste

Outcome Equation Dependent Variable: lnHourly_Earnings				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Education	0.104***	0.0847***	0.107***	0.115***
	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0439***	0.0425***	0.0490***	0.0545***
	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.000520***	-0.0005***	-0.0006***	-0.00064***
	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0157***	-0.00266	0.0472***	-0.0553***
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.437***	1.621***	1.281***	1.380***
	(0.03)	(0.03)	(0.06)	(0.04)
Total obs	14255	9145	4106	6507
R square	0.337	0.228	0.322	0.406
Adj r square	0.337	0.227	0.321	0.406
F	1467.2	475.6	253.5	1094.8
P value	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 7.2 summarizes the estimates from the OLS model for IHDS I. In the OLS model, the coefficients for Education were positive and significant across all caste groups, namely, General (0.104), OBC (0.0847), SC (0.107), and ST (0.115). The coefficients for Experience were positive and significant across all caste groups, namely, General (0.0493), OBC (0.0425), SC (0.049), and ST (0.055). The coefficients for Experiencesqr was negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for all caste groups, namely, General (0.0157), SC (0.0472), and negative and significant for ST (- 0.0553) except OBC (-0.00807), which was not significant. The R squares were General (0.34), OBC (0.23), SC (0.32), and ST (0.40).

7.4.3. Heckman Selection Model (Maximum Likelihood Estimates)- IHDS I

Table 7. 3:Heckman selection model estimates (using maximum likelihood estimator) disaggregated by caste

Outcome Equation Dependent Variable: lnHourly_Earnings				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Education	0.0996***	0.0753***	0.0837***	0.114***
	(0.00)	(0.00)	(0.01)	(0.00)
Experience	0.0493***	0.0455***	0.0527***	0.0568***
	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.000621***	-0.000639***	-0.000761***	-0.000679***
	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0128**	-0.00807	0.0322***	-0.0574***
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.752***	1.075***	0.781***	0.970***
	(0.07)	(0.04)	(0.06)	(0.11)
Selection Equation Dependent Variable: Labourforce Participation				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Married	0.118***	0.137***	0.226***	0.0168
	(0.01)	(0.02)	(0.03)	(0.02)
Number of Children	0.0267***	0.0267***	0.0582***	0.0347***
	(0.00)	(0.00)	(0.01)	(0.01)
Number of Adults	-0.0941***	-0.0824***	-0.0754***	-0.0862***
	(0.00)	(0.01)	(0.01)	(0.01)
Constant	-0.365***	-0.0533**	-0.0995***	-0.528***
	(0.02)	(0.02)	(0.03)	(0.02)
athrho	0.878***	1.189***	1.384***	0.459***
	(0.10)	(0.06)	(0.11)	(0.12)
lnsigma	-0.219***	-0.241***	-0.192***	-0.347***
	(0.04)	(0.02)	(0.03)	(0.04)
Total obs	47889	20668	8440	28538
Selected obs	14255	9145	4106	6507
Non selected obs	33634	11523	4334	22031
rho	0.706	0.830	0.882	0.429
Lambda (IMR)	0.567	0.653	0.728	0.303
chi2	3412.3	1125.1	236.2	3896.7
P value	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table 7.3 we summarize the estimates derived from Heckman (MLE) model. In the Heckman Selection model, the coefficients for Education were positive and highly significant across all caste groups, namely, General (0.0996), OBC (0.0753), SC (0.0837), and ST (0.114). The coefficients for Experience were positive and highly significant across all caste groups, namely, General (0.0493), OBC (0.0455), SC (0.0527), and ST (0.0568). The coefficients for Experiencesqr were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for all caste groups, namely, General (0.0128), SC (0.0322), and negative and significant for ST (-0.0574) except OBC (-0.00807), which was not significant.

7.4.4. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) - IHDS I

Table 7. 4: Heckman two-step selection model estimates (two-step estimator) disaggregated by caste

Outcome Equation Dependent Variable: lnHourly_Earnings				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Education	0.103***	0.0836***	0.106***	0.114***
	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0470***	0.0481***	0.0540***	0.0571***
	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.000573***	-0.000653***	-0.000731***	-0.000684***
	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0139***	-0.00577	0.0459***	-0.0579***
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.104***	1.134***	1.027***	0.895***
	(0.07)	(0.08)	(0.09)	(0.12)
Selection Equation Dependent Variable: Labourforce Participation				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Married	0.123***	0.150***	0.259***	0.0195
	(0.02)	(0.02)	(0.03)	(0.02)
Number of Children	0.0284***	0.0256***	0.0805***	0.0324***
	(0.00)	(0.01)	(0.01)	(0.01)
Number of Adults	-0.105***	-0.0933***	-0.123***	-0.0857***
	(0.00)	(0.01)	(0.01)	(0.01)
Constant	-0.338***	-0.0267	-0.0225	-0.528***
	(0.02)	(0.03)	(0.04)	(0.02)
IMR	0.265***	0.483***	0.249***	0.357***
	(0.05)	(0.07)	(0.07)	(0.08)

Total obs	47889	20668	8440	28538
Selected obs	14255	9145	4106	6507
Non selected obs	33634	11523	4334	22031
rho	0.389	0.688	0.399	0.492
Lambda (IMR)	0.265	0.483	0.249	0.357
chi2	7011.9	2625.9	1909.9	4241.8
P value	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table 7.4, we report the results for Heckman (TSE) model. We observe that the coefficient for Education for General and forward castes (0.103), OBC (0.0836), SC (0.106), and ST (0.114). The coefficients were positive and significant all castes

We also observe the coefficient for Experience for General and Forward castes (0.047), OBC (0.048), SC (0.054), and ST (0.057). The coefficients were positive and significant for all castes

The coefficient for Square of Experience was negative and significant implying non-linear effects.

The coefficient for Networking Intensity for General and Forward castes (0.0139), SC (0.0459), and ST (- 0.0579). The coefficients were positive and significant for General and Forward Castes, and SC and were negative for ST.

The coefficient for IMR (Inverse Mills Ratio) was significant, indicating the presence of selection bias.

We find that the returns to an additional year of education are lowest for OBC and highest for ST. We find that the returns to an additional year of experience are lowest for General and highest for SC.

We also find that an inverted U-shaped relationship exists between earnings and experience. Last, we find that social capital, as measured by networking intensity, positively impacts earnings for General/forward and SC but was negative for ST.

However, it is interesting to note that the intercept term in the main model was highest for OBC, followed by General/ forward castes, and lowest for SC. We also observe that, while the General/ Forward caste shows a significant positive impact of all explanatory variables on earnings (except experience square), the OBC show a significant positive impact on education and experience but not networking intensity. Similarly, ST show a significant positive impact on education and experience but a negative impact on networking intensity. Lastly, though SC show a significant positive impact on all explanatory variables, their intercept term seems to be the lowest among all caste categories.

7.5. CROSS-SECTIONAL ANALYSIS IHDS II

7.5.1. Descriptive Statistics IHDS II

Table 7. 5: Descriptive statistics:

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	53,351	2.98	0.75	0.00	7.82
Education	Completed years of education	2,04,336	5.32	4.88	0.00	16.00
Experience	Potential experience: {(Age - Education - 5) Years}	2,69,335	24.48	18.81	0.00	45.00
Experiencesqr	Square of potential experience	2,69,335	952.71	900.05	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups households were members of or associated with.	1,60,427	0.68	1.05	0.00	7.00
Age	Age in years	2,04,565	29.82	20.36	0.00	99.00

Labourforce Participation	Labor force participation dummy variable ('1' if hourly earnings are observed, '0' otherwise)	2,69,335	0.20	0.40	0.00	1.00
Married	Married dummy variable ('1' if the individual is married, '0' otherwise)	2,69,335	0.35	0.48	0.00	1.00
Number of Children	Number of children in the household	2,04,568	1.84	1.66	0.00	18.00
Number of Adults	Number of adults in the household	2,04,568	3.41	1.70	0.00	18.00
Caste Categories	Caste Categories '1'General/Forward/Others, '2' Other Backward Classes, '3' Scheduled Castes and '4' Scheduled Tribes	2,04,114	2.09	0.91	1.00	4.00

Source: Author's calculation using IHDS II Data

We found that an individual's (natural log of) average hourly earnings were 2.98 with a standard deviation of 0.75. Average education was 5.32 years, with a standard deviation of 4.88 years. The average experience was 24.48 years, with a standard deviation of 18.81 years. Of the total sample, 35% were married. The average number of children was 1.84, and the adults were 3.41. Caste Categories comprised of '1' General/Forward castes (Including Brahmins and others), '2' Other Backward castes, '3' Scheduled castes, and '4' Scheduled tribes.

Next, we proceed to estimate the Heckman Selection model for these caste categories for IHDS II (2011-12).

7.5.2. Ordinary Least Squares estimates IHDS II

Table 7. 6: OLS model estimates disaggregated by caste

Outcome Equation Dependent Variable: lnHourly_Earnings				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Education	0.0779***	0.0640***	0.0863***	0.0899***
	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0322***	0.0235***	0.0395***	0.0356***
	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.000352***	-0.000234***	-0.000484***	-0.000358***
	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0382***	0.0391***	0.0363***	0.0285***
	(0.00)	(0.01)	(0.01)	(0.01)
Constant	1.923***	2.169***	1.721***	1.877***
	(0.03)	(0.03)	(0.05)	(0.04)
Total obs	18409	12126	4990	8553
R square	0.213	0.150	0.239	0.264
Adj r square	0.213	0.149	0.238	0.263
F	933.2	353.4	189.4	657.2
P value	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 7.6 summarizes the estimates from the OLS model for IHDS II. In the OLS model, the coefficients for Education were positive and significant across all caste groups, namely, General (0.0.78), OBC (0.064), SC (0.086), and ST (0.089). The coefficients for Experience were positive and significant across all caste groups, namely, General (0.032), OBC (0.0235), SC (0.0395), and ST (0.0356). The coefficients for Experiencesqr were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for all caste groups, namely, General (0.0382), OBCC (0.0391), SC (0.0363), and ST (0.0285). The R squares were General (0.21), OBC (0.15), SC (0.30), and ST (0.26).

7.5.3. Heckman Selection Model (Maximum Likelihood Estimates)- IHDS II

Table 7. 7: Heckman selection model estimates (using maximum likelihood estimator) disaggregated by caste

Outcome Equation Dependent Variable: lnHourly_Earnings				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Education	0.0772***	0.0633***	0.0848***	0.0894***
	(0.00)	(0.00)	(0.01)	(0.00)
Experience	0.0335***	0.0263***	0.0424***	0.0357***
	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.000372***	-0.000280***	-0.000535***	-0.000361***
	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0374***	0.0382***	0.0308**	0.0281***
	(0.00)	(0.01)	(0.01)	(0.01)
Constant	1.694***	2.003***	1.552***	1.747***
	(0.05)	(0.05)	(0.23)	(0.06)
Selection Equation Dependent Variable: Labourforce Participation				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Married	0.0680***	0.218***	0.183***	-0.0698***
	(0.01)	(0.02)	(0.04)	(0.02)
Number of Children	-0.0125***	0.00609	0.0412***	0.0139**
	(0.00)	(0.01)	(0.01)	(0.01)
Number of Adults	-0.101***	-0.0985***	-0.124***	-0.0840***
	(0.00)	(0.01)	(0.02)	(0.01)
Constant	0.136***	0.436***	0.536***	-0.0453*
	(0.02)	(0.03)	(0.07)	(0.02)
athrho	0.364***	0.378***	0.421	0.184***
	(0.08)	(0.09)	(0.68)	(0.07)
insigma	-0.374***	-0.484***	-0.463***	-0.350***
	(0.02)	(0.02)	(0.13)	(0.01)
Total obs	42937	19796	7908	24259
Selected obs	18409	12126	4990	8553
Non selected obs	24528	7670	2918	15706
rho	0.349	0.361	0.398	0.182
Lambda (IMR)	0.240	0.223	0.250	0.128
chi2	3386.2	1317.2	416.4	2550.2
P value	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table 7.7 we summarize the estimates derived from Heckman (MLE) model. In the Heckman Selection model, the coefficients for Education were positive and highly significant across all caste groups, namely, General (0.077), OBC (0.063), SC (0.084), and ST (0.089). The coefficients for Experience were positive and highly significant across all caste groups, namely, General (0.036), OBC (0.026), SC (0.042), and ST (0.035). The coefficients for Experiencesqr were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for Networking intensity were positive and highly significant for all caste groups, namely, General (0.037), OBC (0.038), SC (0.030), and ST (0.028), which was not significant.

7.5.4. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) - IHDS II

Table 7. 8: Heckman two-step selection model estimates (two-step estimator) disaggregated by caste

Outcome Equation Dependent Variable: lnHourly_Earnings				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Education	0.0775***	0.0638***	0.0858***	0.0891***
	(0.00)	(0.00)	(0.00)	(0.00)
Experience	0.0337***	0.0273***	0.0422***	0.0358***
	(0.00)	(0.00)	(0.00)	(0.00)
Experiencesqr	-0.000372***	-0.000294***	-0.000528***	-0.000362***
	(0.00)	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0374***	0.0379***	0.0332***	0.0278***
	(0.00)	(0.01)	(0.01)	(0.01)
Constant	1.671***	1.941***	1.579***	1.650***
	(0.05)	(0.06)	(0.07)	(0.09)

Selection Equation Dependent Variable: Labourforce Participation				
	General/ Forward/Others	**OBC**	**SC**	**ST**
Married	0.0711***	0.232***	0.188***	-0.0654***
	(0.01)	(0.02)	(0.03)	(0.02)
Number of Children	-0.0184***	-0.00171	0.0352***	0.0118**
	(0.00)	(0.01)	(0.01)	(0.01)

Number of Adults	-0.0984***	-0.0932***	-0.127***	-0.0835***
	(0.00)	(0.01)	(0.01)	(0.01)
Constant	0.134***	0.421***	0.552***	-0.0475**
	(0.02)	(0.03)	(0.04)	(0.02)
IMR	0.260***	0.295***	0.197**	0.223***
	(0.05)	(0.06)	(0.08)	(0.08)
Total obs	42937	19796	7908	24259
Selected obs	18409	12126	4990	8553
Non selected obs	24528	7670	2918	15706
rho	0.375	0.466	0.318	0.309
Lambda (IMR)	0.260	0.295	0.197	0.223
chi2	4899.7	2108.9	1532.3	2900.7
P value	0.0000	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

In table 7.8, we report the results for Heckman (MLE) model. We observe that the coefficient for Education for General and forward castes (0.076), OBC (0.064), SC (0.086), and ST (0.089). The coefficients were positive and significant in all castes.

We also observe the coefficient for Experience for General and Forward castes (0.034), OBC (0.027), SC (0.042), and ST (0.035). The coefficients were positive and significant for all castes. The coefficient for the Square of Experience was negative and significant, implying non-linear effects.

The coefficient for Networking Intensity for General and Forward castes (0.037), OBC (0.038), SC (0.033), ST (0.028). The coefficients were positive and significant for all castes. The coefficient for IMR (Inverse Mills Ratio) was significant, indicating the presence of selection bias.

We find that the returns to an additional year of education are lowest for OBC and highest for ST. We find that the returns to an additional year of experience are lowest for OBC and highest for SC.

We also find that an inverted U-shaped relationship exists between earnings and experience. Last, we find that social capital, as measured by networking intensity, positively impacts earnings for all castes

7.6. PANEL DATA ANALYSIS IHDS I & II

In this section of the chapter, we proceed towards analyzing data from rounds 1 and 2 of IHDS, namely IHDS I & II (2004-05 and 2011-12). We merged the data from IHDS I and II and formed a panel data set with 150,995 individuals who were interviewed in both rounds. The analysis is restricted to individuals who were interviewed in both rounds of IHDS. We then merged this panel data set with household-level data and village-level data to form an elaborate panel dataset incorporating household and village attributes. The sub-sample selected for analysis in this chapter is restricted to individuals in the age group of 15 to 64 years as per the guidelines of Labour force participation followed by NSSO. Experience (Potential Experience) has been restricted to an upper bound of 45 years. Our analysis consists of descriptive statistics of the selected sample, followed by estimation of fixed effects multiple regression model by Consumption categories (as described in the previous section) using the Heckman Selection model addressing selection bias.

The model can then be specified as follows:

First stage equation:

$$\textbf{Labourforce Participation}_{it} = \alpha_0 + \alpha_1\textbf{Married}_{it} + \alpha_2\textbf{NChildren}_{it} + \alpha_3\textbf{Nadults}_{it} + \alpha_4(\textbf{X})_{it} + \varepsilon_{it} \qquad \textbf{(7.4)}$$

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 7.5 and e = stochastic error term.

Second stage equation:

$$\boldsymbol{Ln\ Hourly\ Earnings_{it} = \beta_0 + \beta_1 Education_{it} + \beta_2 Experience_{it} + \beta_3 Experience_{it}^2 + \beta_4 Networking\ Intensity_{it} + \beta_5 IMR_{it} + \beta_6 Z_{it} + u_{it}} \quad \textbf{(7.5)}$$

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (eq 7.4) (Married, Nchildren, and Nadults), u= error term.

7.6.1. Descriptive Statistics

Table 7. 9: Descriptive Statistics:

Variable	Description		Mean	SD	Min	Max	Obs
lnHourly_Earnings	Natural logarithm of hourly earnings observed	overall	3.02	0.82	-0.84	6.92	21413
		between		0.77	-0.84	6.71	15601
		within		0.28	1.10	4.94	1.37
Education	Completed years of education	overall	7.90	4.44	0	16	52134
		between		4.03	0	16	29878
		within		1.63	-0.10	15.90	1.74
Experience	Potential experience: {(age - education - 5) years}	overall	19.51	13.53	0	45	52378
		between		13.62	0	45	29889
		within		2.71	-1.99	41.01	1.75
Experience square	Square of potential experience	overall	563.77	629.60	0	2025	52378
		between		627.80	0	2025	29889
		within		138.98	-446.7	1574.27	1.75
Networking intensity	Number of social or economic committees that household members were associated with	overall	0.71	1.08	0	7	52225
		between		0.90	0	7	29870
		within		0.62	-2.79	4.21	1.75
Labor force participation	Binary variable ('1' if	overall	0.41	0.49	0	1	52378
		between		0.43	0	1	29889

	hourly earnings are observed, '0' otherwise)	within		0.25	-0.09	0.91	1.75
Married	Binary variable ('1' if the individual is married, '0' otherwise)	overall between within	0.62	0.49 0.46 0.18	0 0 0.12	1 1 1.12	52378 29889 1.75
Children	Number of children in the household	overall between within	1.54	1.62 1.41 0.84	0 0 -4.96	18 15 8.04	52378 29889 1.75
Adults	Number of adults in the household	overall between within	3.47	1.74 1.50 0.90	0 0 -4.53	18 18 11.47	52378 29889 1.75
Caste	Categories: '1' if general /others, '2' if other backward classes, '3' if scheduled castes, and '4' if scheduled tribes	overall between within	2.11	1.10 0.81 0.78	1 1 0.61	4 4 3.61	52310 29888 1.75

Source: Author's calculations based on IHDS I and II

Table 9.9 provides the descriptive statistics for the merged panel data. We found that the (natural log of) average hourly earnings of an individual (overall) was 3.02 with a standard deviation of 0.8. Average education was 7.9 years, with a standard deviation of 4.4 years (overall). The average experience was 19.5 years, with a standard deviation of 13.5 years (overall). Of the total sample, 62% were married, and 39% were female. The average number of children was 1.54, and OBCthe average number of adults was 3.47.

7.6.2. Heckman Estimates

Table 7. 10: Estimates from the Fixed-effects Heckman Selection Model disaggregated by Caste (dependent variable: lnHourly_Earnings).

Main Equation	General	OBC	SC	ST
Education	0.0466***	0.0191**	0.0599***	0.0152*
	(0.0000)	(0.0448)	(0.0001)	(0.0837)
Experience	0.0305***	-0.00834	0.0267*	0.00349
	(0.0020)	(0.4257)	(0.0780)	(0.7966)
Experience_square	-0.000341**	0.000156	-0.000249	-0.0000310
	(0.0409)	(0.5025)	(0.3933)	(0.9103)
Networking Intensity	0.0108	0.0161	-0.00135	0.0267
	(0.2345)	(0.2150)	(0.9552)	(0.1013)
2012	0.215***	0.363***	0.179	0.374***
	(0.0048)	(0.0000)	(0.2410)	(0.0022)
IMR				
2005	-0.203***	-0.186***	0.0892	-0.276***
	(0.0013)	(0.0025)	(0.4125)	(0.0024)
2012	-0.336***	-0.341***	0.0697	-0.440***
	(0.0002)	(0.0006)	(0.7914)	(0.0001)
Constant	1.582***	1.834***	1.122***	1.681***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Source: Author's calculations based on IHDS I and II
Note: Standard errors are in parentheses.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 7.10 provides the disaggregated estimates derived from the Heckman fixed effects model. We observe that the coefficient for Education for General and forward castes (0.0466), OBC(0.0191), SC (0.0599), and ST (0.0152). The coefficients were positive and significant for General / Forward Castes, Other Backward Castes, SC, and ST.

We also observe the coefficient for Experience, General, and Forward castes (0.0305) and SC (0.0267). The coefficients were positive and significant for General and Forward Castes and SC.

The coefficient for the Square of Experience was negative and significant for only General and Forward Castes.

The coefficient for Networking Intensity was not significant for any caste categories.

The coefficient for IMR (Inverse Mills Ratio) was significant, indicating the presence of selection bias. Hence, it justifies the use of Heckman's Two Step Correction model to address selection bias.

We find that the returns to an additional year of education are lowest for ST and highest for SC, followed by General/Forward castes. We find that the returns to an additional year of experience are lowest for SC and highest for General/forward castes.

We also find that an inverted U-shaped relationship exists between earnings and experience only for general and forward castes. Last, we find that social capital, as measured by networking intensity, did not illustrate significance in explaining earnings for all caste categories in the panel model.

However, it is interesting to note that the intercept term in the main model was highest for OBC, followed by STs and General/ Forward, and lowest for SCs.

Table 7. 11: Selection equation estimates from Fixed Effects- Heckman Selection model by Caste (dependent variable: Labourforce Participation).

Selection Equation	General	OBC	SC	ST
Year Specific Education				
2005	0.0644***	0.0223	0.101**	0.00978
	(0.0000)	(0.1593)	(0.0103)	(0.6146)
2012	-0.0133	-0.00287	-0.0685**	-0.0500***
	(0.2086)	(0.8642)	(0.0271)	(0.0017)
Year Specific Experience				
2005	0.00295	0.0113	0.0544	-0.0137
	(0.9000)	(0.6741)	(0.2737)	(0.6438)
2012	0.0728***	0.0292	0.0406	0.0682***
	(0.0000)	(0.1625)	(0.1730)	(0.0007)
Year Specific Experience_square				
2005	-0.000244	-0.000822	-0.00109	-0.0000804
	(0.5196)	(0.1072)	(0.1637)	(0.8666)
2012	-0.000956***	-0.000637	-0.000918	-0.00152***
	(0.0009)	(0.1497)	(0.1465)	(0.0001)
Year Specific Networking Intensity				
2005	-0.0276	0.0122	-0.0967**	-0.0444*
	(0.1817)	(0.6698)	(0.0368)	(0.0732)
2012	0.0380**	0.0163	-0.0638	0.0397
	(0.0305)	(0.5493)	(0.1290)	(0.1321)
Year Specific Married				
2005	-0.756***	-0.886***	-0.711***	-0.930***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
2012	0.778***	1.209***	0.683***	0.846***
	(0.0000)	(0.0000)	(0.0005)	(0.0000)
Year Specific Nchildren				
2005	-0.0543***	-0.0929***	-0.0619	-0.0362*
	(0.0037)	(0.0002)	(0.1166)	(0.0938)
2012	-0.00876	-0.0460*	0.0147	0.00426
	(0.6676)	(0.0807)	(0.7356)	(0.8418)
Year Specific Nadults				
2005	-0.0353*	0.00975	-0.0828*	-0.0465**
	(0.0613)	(0.7161)	(0.0642)	(0.0146)
2012	-0.0195	0.0193	-0.0450	-0.0129
	(0.2544)	(0.4724)	(0.2581)	(0.5731)
Year-Specific Individual Average Education				
2005	-0.0407**	0.00474	-0.0872**	0.0312
	(0.0143)	(0.7905)	(0.0244)	(0.1139)
2012	0.0189*	0.0117	0.0838***	0.0690***
	(0.0911)	(0.4999)	(0.0091)	(0.0000)
Year-Specific Individual Average Experience				
2005	0.138***	0.150***	0.0936*	0.150***

	(0.0000)	(0.0000)	(0.0899)	(0.0000)
2012	-0.0115	0.0420*	0.0358	-0.00537
	(0.4628)	(0.0590)	(0.2629)	(0.7909)
Year Specific Individual Average Experience_square				
2005	-0.00210***	-0.00167***	-0.00145*	-0.00212***
	(0.0000)	(0.0039)	(0.0906)	(0.0001)
2012	-0.000316	-0.000813*	-0.000600	0.000328
	(0.2892)	(0.0725)	(0.3377)	(0.4056)
Year-Specific Individual Average Networking Intensity				
2005	0.0208	0.0122	-0.0600	0.0360
	(0.4367)	(0.7380)	(0.2992)	(0.2243)
2012	-0.0289	0.0266	0.00476	-0.0149
	(0.1277)	(0.4001)	(0.9338)	(0.6604)
Year Specific Individual Average Married				
2005	0.733***	0.746***	0.973***	0.802***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
2012	-0.988***	-1.218***	-0.764***	-1.245***
	(0.0000)	(0.0000)	(0.0012)	(0.0000)
Year Specific Individual Average Nchildren				
2005	0.0740***	0.129***	0.113**	0.0657**
	(0.0018)	(0.0001)	(0.0297)	(0.0344)
2012	-0.0244	0.0380	-0.0185	0.0000827
	(0.2549)	(0.2118)	(0.6916)	(0.9974)
Year Specific Individual Average Nadults				
2005	-0.106***	-0.178***	-0.0483	-0.0625***
	(0.0000)	(0.0000)	(0.3644)	(0.0049)
2012	-0.103***	-0.100***	-0.0794*	-0.0988***
	(0.0000)	(0.0021)	(0.0737)	(0.0000)
Year				
2012	1.595***	1.542***	1.451***	1.779***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	-1.746***	-1.678***	-1.349***	-2.177***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N	23700	9862	3813	14614

Source: Author's calculations based on IHDS I and II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 7.11 above summarizes the results of the first stage equation of the Heckman model with fixed effects. In addition to the instruments used to estimate labor force participation, the model also estimates interactions between these instruments and other covariates by year. The instruments used for the first stage probit model; Married, Nchildren, and NAdults were also significant in the first stage equation. These estimates were jointly significant across the models.

7.7. PROJECTED EARNINGS BY CASTE WITH EDUCATION AND EXPERIENCE AS COVARIATES

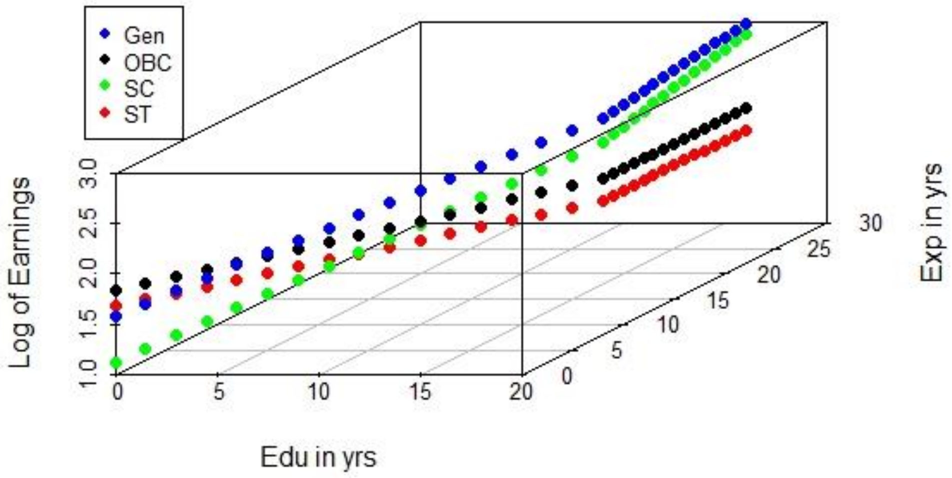

Figure 7. 1: Projected earnings by caste with education and experience as covariates.

Among the caste groups (Figure 7.1), the SCs show the highest returns to education (5.9 %), followed by the general castes (4.6%). However, general castes have a higher intercept, which elevates their earnings above other castes for almost the entire range of years. While we see that SCs projections could possibly hint at some positive dividends for investment in education in india, the projections for OBC and ST seem to illustrate divergence.

7.8. FINDINGS

Cross-sectional IHDS I :

The private returns to an additional year of *Education* in the OLS model range between the lowest for OBC (8.5%) and highest for ST (11.5%). The private returns to an additional year of *Education* in the Heckman model (MLE) range between the lowest for OBC (7.6%) and highest for ST (11.4%). The private returns to an additional year of *Education* in the Heckman model (TSE) range between the lowest for OBC (8.4%) and highest for ST (11.4%).

The private returns to an additional year of *Experience* in the OLS model range between the lowest for OBC (4.25%) and highest for ST (5.45%). The private returns to an additional year of *Experience* in the Heckman model (MLE) range between the lowest of OBC (4.6%) and highest for ST (5.7%). The private returns to an additional year of *Experience* in the Heckman model (TSE) range between the lowest of General and forward (4.7%) and highest for ST (5.7%).

The estimate for *Networking Intensity* in the OLS model range between the lowest for ST (-5.5%) and the highest for SC (4.7%). The estimate for *Networking Intensity* in the Heckman model (MLE) range between the lowest for ST (- 5.7%) and highest for SC (3.2%). The estimate for *Networking Intensity* in the Heckman model (TSE) range between the lowest for ST (- 5.8%) and highest for SC (4.6%).

Cross-sectional IHDS II :

The private returns to an additional year of *Education* in the OLS model range between the lowest for OBC (6.4%) and highest for ST (8.99%). The private returns to an additional year of *Education* in the Heckman model (MLE) range between the lowest of OBC (6.3%) and highest for ST (8.9 %). The private returns to an additional year of

Education in the Heckman model (TSE) range between the lowest of OBC (6.38%) and highest for ST (8.9 %).

The private returns to an additional year of *Experience* in the OLS model range between the lowest for OBC (2.35%) and the highest for SC (3.95%). The private returns to an additional year of *Experience* in the Heckman model (MLE) range between the lowest of OBC (2.63%) and highest for SC (4.2%). The private returns to an additional year of *Experience* in the Heckman model (TSE) range between the lowest of OBC (2.7%) and highest for SC (4.2%).

The estimate for *Networking Intensity* in the OLS model range between the lowest of ST (2.85%) and the highest for General (3.82%). The estimate for *Networking Intensity* in the Heckman model (MLE) range between the lowest for ST (2.8%) and highest for OBC (3.8%). The estimate for *Networking Intensity* in the Heckman model (TSE) range between the lowest for ST (2.8%) and highest for OBC (3.8%).

Panel data IHDS I & II:

We found that *Education* has a positive and significant impact on all castes. The *Experience* was positive and significant for general castes, and SC. *Experience*sqr was negative and significant for general. This suggests a U-shaped relation between earnings and *Experience* for (general). The IMR was significant for most of the values, indicating the presence of a selection bias.

7.9. CONCLUSIONS

We conclude that Education and Experience have a positive and significant impact on earnings for all castes. Panel data results show that SCs, realise highest private returns to education followed by General. A possible explanation could be that reservations allow for a conducive environment for individuals to enter the job market. This would result in higher marginal earnings. It could be compared as such to a neoclassical model, where it is expected that countries with limited capital will experience a higher rate of return on capital as compared to richer countries. Estimates from cross-sectional data and panel data lead us to conclude that affirmative action has helped to some extent in reducing social disparities and generating opportunities. However, the results are not uniform across social groups. Our earnings projection using panel data estimates lead us to conclude that while some social groups may have benefitted (like SCs) , others like OBCs and STs seem to need further targeted intervention. Education (SDG 4), then becomes crucial in context to breaking social barriers and reduce inequalities (SDG10).

In the next chapter, we cover another major social group in the Indian context. We analyze trends in private returns to education and experience across religions. This disaggregation, we believe, will provide more meaningful insight in to social disparities and add to the findings from the current chapter.

CHAPTER 8: TRENDS IN RETURNS TO HUMAN CAPITAL INVESTMENT ACROSS RELIGION

The previous chapter showed empirical evidence of earnings differentials across caste. We understood differences in the private returns to education across caste categories.

In this chapter, we provide disaggregated estimates of private return to education by religions. In this particular chapter, we aim to furnish a comprehensive statistical analysis that delivers disaggregated estimates of private returns to education based on different religious affiliations.

8.1. INTRODUCTION

In this chapter, we estimate returns for human capital investment for three religious groups, namely Hindus, Other Minorities (Christian, Sikh, Jain, Buddhist, and others), and Muslims. We compare the returns to investment in human capital for these religious groups in an attempt to highlight the direction and magnitude of their private returns. These disaggregated private returns, we believe, would provide a deeper understanding of the trajectory that their marginal earnings follow, given their level of education and experience. We are also interested in observing the role of social capital in determining earnings across these religious groups.

We estimate the returns to investment in additional years of education using cross-sectional data from the Indian Human Development Survey (IHDS: 2004-05 and 2011-12) and the panel data for these two rounds of IHDS. The chapter summarises estimates for these returns to private investment on additional years of education with continuous Years of Educational attainment. This model is estimated using Ordinary Least Squares and the Heckman Models for correction of selection bias.

The aim of this chapter is to provide a comprehensive understanding of the role of education in shaping the economic prospects of individuals, while also taking into account the potential influence of religious affiliation. This, we believe, would offer new insights into the relationship between earnings, education, and religion.

8.2. FRAMEWORK

For the purpose of establishing a basic framework to achieve the objective mentioned above, this paper uses a modified version of the Mincer earnings function (Jacob Mincer, 1974). The original function, as developed by Jacob Mincer in his book Schooling, experience and Earnings, has come to become a benchmark for studies in the field of human capital and labor economics, though the basic earning variation by schooling and age has existed since Miller (1955).

8.2.1. Models

We use an Extended Econometric Model for Mincer's Earnings Function, which can then be stated as follows:

Model 1: Ordinary Least Squares using Continuous Years of Educational attainment.

$$\boldsymbol{Ln\ Hourly\ Earnings_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \beta_3 Experience_i^2 + \beta_4 Networking\ Intensity_i + u_i} \quad \textbf{(8.1)}$$

Where Ln Hourly Earnings is dependent on Education and Experience and their squares (to test for non-linearity), and u is the stochastic error term.

However, as pointed out in Chapter 4, the estimation of this suffers from at least two sources of endogeneity biases—namely, selection Bias and ability Bias. Hence, we proceed to estimate our model using Heckman selection models to address selection bias in addition to OLS.

The Heckman Model for Correction of selection Bias would then be estimated as follows:

Model 2: The Heckman selection model to estimate private returns to education by Religion

First stage equation:

$$\boldsymbol{Labourforce\ Participation_i = a_0 + a_1 Married_i + a_2 Nchildren_i + a_3 Nadults_i^2 + a_4 X_i + e_i} \quad \textbf{(8.2)}$$

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 8.3 and ε = stochastic error term.

Second stage equation:

$$\boldsymbol{Ln\ Hourly\ Earnings_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \beta_3 Experience_i^2 + \beta_4 Networking\ Intensity_i + \beta_5 IMR_i + \beta_6 Z_i + u_i} \quad \textbf{(8.3)}$$

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (8.2) (Married, Nchildren, and Nadults), u= error term.

The model also estimates a value for IMR to allow for testing the hypothesis of whether selection bias is an issue in the estimation. A significant value for Lambda would imply the existence of selection Bias and, as such, the use of Heckman's Two Step model for estimation.

8.3. ANALYSIS

8.4. CROSS-SECTIONAL ANALYSIS IHDS I

8.4.1. Descriptive Statistics IHDS I

Table 8. 1: Descriptive statistics:

Variables	Description	Obs	Mean	Std. Dev.	Min	Max
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	35,519	2.70	0.78	-2.22	7.25
Education	Completed years of education	214,399	4.67	4.68	0.00	15.00
Experience	Potential experience: {(Age - Education - 5) Years}	269,335	22.34	18.53	0.00	45.00
Experiencesqr	Square of potential experience	269,335	842.53	878.68	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups households were members of or associated with.	179,682	0.61	1.02	0.00	7.00
Age	Age in years	215,754	27.35	19.35	0.00	116.00
Labourforce Participation	Labor force participation dummy variable ('1' if hourly	269,335	0.18	0.38	0.00	1.00

	earnings are observed, '0' otherwise)					
Married	Married dummy variable ('1' if the individual is married, '0' otherwise)	269,335	0.36	0.48	0.00	1.00
Number of Children	Number of children in the household	215,754	2.20	1.86	0.00	17.00
Number of Adults	Number of adults in the household	215,754	3.27	1.72	0.00	18.00
Religion Categories	Religion Categories: '1' Hindu, '2' Other Minorities, '3' Muslims	269,335	1.56	0.65	1.00	3.00

Source: Author's calculation using IHDS I Data

We found that an individual's (natural log of) average hourly earnings were 2.70 with a standard deviation of 0.78. Average education was 4.67 years, with a standard deviation of 4.7 years. The average experience was 22.34 years, with a standard deviation of 18.5 years. Of the total sample, 36% were married. The average number of children was 2.2, and the adults were 3.3. Religion Categories comprised of '1' Hindu, '2' Other Minorities, and '3' Muslims.

Next, we proceed to estimate our models by religion.

8.4.2. **Ordinary Least Squares Estimates IHDS I**

Table 8. 2: Ordinary Least Squares estimates disaggregated by religion

Outcome Equation Dependent Variable: lnHourly_Earnings			
	Hindu	**Other Minorities**	**Muslim**
Education	0.107***	0.115***	0.0807***
	(0.00)	(0.00)	(0.00)
Experience	0.0493***	0.0364***	0.0459***
	(0.00)	(0.01)	(0.00)
Experiencesqr	-0.000606***	-0.000285**	-0.000657***
	(0.00)	(0.00)	(0.00)
Networking Intensity	0.00249	-0.0410***	-0.00241
	(0.00)	(0.02)	(0.01)
Constant	1.353***	1.677***	1.774***
	(0.02)	(0.09)	(0.06)
Total obs	27883	1536	3335
R square	0.361	0.329	0.217
Adj R square	0.361	0.327	0.216
F	3074.2	208.2	170.2
P value	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 8.2 provides OLS estimates disaggregated by religion. We observe that the coefficient for Education for Hindus (0.107), Other Minorities (0.115), and Muslims (0.0807). The coefficients were positive and significant for Hindus, Other Minorities, and Muslims. We also observe the coefficient for Experience for Hindus (0.0493), Other Minorities (0.364), and Muslims (0.0459). The coefficients were positive and significant for Hindus, Other Minorities, and Muslims.

The coefficient for Square of Experience was negative and significant for Hindus, Other Minorities, and Muslims. The coefficient for Networking Intensity for Other Minorities (-0.041) was negative and significant.

We find that the private returns to an additional year of education are lowest for Muslims and highest for Other Minorites. We find that the returns to an additional year of experience are lowest for Other Minorities and highest for Hindus. We also find that social capital, as measured by networking intensity, bears a negative and significant impact on earnings for Other Minorities.

8.4.3. Heckman Selection Model (Maximum Likelihood Estimates)- IHDS I

Table 8. 3: Heckman selection model estimates (using maximum likelihood estimator) disaggregated by religion

Outcome Equation Dependent Variable: lnHourly_Earnings			
	Hindu	**Other Minorities**	**Muslim**
Education	0.103***	0.116***	0.0778***
	(0.00)	(0.00)	(0.00)
Experience	0.0536***	0.0400***	0.0469***
	(0.00)	(0.01)	(0.00)
Experiencesqr	-0.000688***	-0.000339**	-0.000678***
	(0.00)	(0.00)	(0.00)
Networking Intensity	-0.00195	-0.0408***	-0.00452
	(0.00)	(0.02)	(0.01)
Constant	0.907***	1.449***	1.380***
	(0.06)	(0.21)	(0.09)
Selection Equation Dependent Variable: Labourforce Participation			
	Hindu	**Other Minorities**	**Muslim**
Married	0.0895***	0.605***	0.0127
	(0.01)	(0.03)	(0.03)
Number of Children	0.0366***	-0.0207**	0.00927
	(0.00)	(0.01)	(0.01)
Number of Adults	-0.142***	-0.0959***	-0.105***
	(0.01)	(0.01)	(0.01)
Constant	0.0708***	-1.518***	-0.154***
	(0.02)	(0.04)	(0.04)
athrho	0.737***	0.135	0.569***
	(0.11)	(0.11)	(0.12)
lnsigma	-0.329***	-0.333***	-0.351***
	(0.04)	(0.03)	(0.04)
Total obs	70253	22288	10367
Selected obs	27883	1536	3335
Non selected obs	42370	20752	7032
rho	0.627	0.134	0.515
Lambda (IMR)	0.451	0.0958	0.362
chi2	5400.8	840.3	543.0
P value	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 8.3 provides the Heckman selection model (MLE), estimates disaggregated by religion. The coefficients for *Education* were positive and significant across all religious groups, namely, Hindus (0.0103), Other minorities (0.116), and Muslims (0.0778). The coefficients for *Experience* were positive and significant across all religious groups, namely, Hindus (0.0536), Other minorities (0.04), and Muslims (0.0469). The coefficients for *Experiencesqr* were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for *Networking intensity* were negative and significant only for Other minorities (-0.0408).

8.4.4. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) - IHDS I

Table 8. 4: Heckman two-step selection model estimates (two-step estimator) disaggregated by religion

Outcome Equation Dependent Variable: lnHourly_Earnings			
	Hindu	**Other Minorities**	**Muslim**
Education	0.106***	0.116***	0.0776***
	(0.00)	(0.00)	(0.00)
Experience	0.0514***	0.0400***	0.0483***
	(0.00)	(0.01)	(0.00)
Experiencesqr	-0.000643***	-0.000339**	-0.000698***
	(0.00)	(0.00)	(0.00)
Networking Intensity	0.000445	-0.0408**	-0.00621
	(0.00)	(0.02)	(0.01)
Constant	1.150***	1.448***	1.177***
	(0.03)	(0.23)	(0.13)
Selection Equation Dependent Variable: Labourforce Participation			
	Hindu	**Other Minorities**	**Muslim**
Married	0.0800***	0.605***	0.0369
	(0.01)	(0.03)	(0.03)
Number of Children	0.0327***	-0.0214**	0.00781
	(0.00)	(0.01)	(0.01)
Number of Adults	-0.154***	-0.0955***	-0.100***
	(0.00)	(0.01)	(0.01)
Constant	0.124***	-1.519***	-0.182***
	(0.01)	(0.04)	(0.04)
IMR	0.198***	0.0961	0.531***
	(0.02)	(0.09)	(0.10)

Total obs	70253	22288	10367
Selected obs	27883	1536	3335
Non selected obs	42370	20752	7032
rho	0.309	0.134	0.684
Lambda (IMR)	0.198	0.0961	0.531
chi2	15154.6	746.0	822.8
P value	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS I
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 8.4 provides the Heckman selection model (TSE), estimates disaggregated by religion. The coefficients for *Education* were positive and significant across all religious groups, namely, Hindus (0.0106), Other minorities (0.116), and Muslims (0.0776). The coefficients for *Experience* were positive and significant across all religious groups, namely, Hindus (0.0514), Other minorities (0.04), and Muslims (0.0483). The coefficients for *Experiencesqr* were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for *Networking intensity* were negative and significant only for Other minorities (-0.0408). IMR was significant for Hindus and Muslims.

8.5. CROSS-SECTIONAL ANALYSIS IHDS II

8.5.1. Descriptive Statistics IHDS II

Table 8. 5: Descriptive statistics:

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Ln Hourly Earnings	Natural logarithm of hourly earnings observed	53,351	2.98	0.75	0.00	7.82
Education	Completed years of education	2,04,336	5.32	4.88	0.00	16.00
Experience	Potential experience: {(Age - Education - 5) Years}	2,69,335	24.48	18.81	0.00	45.00
Experiencesqr	Square of potential experience	2,69,335	952.71	900.05	0.00	2025.00
Networking Intensity	Generated by the number of social or economic groups households were members of or associated with.	1,60,427	0.68	1.05	0.00	7.00
Age	Age in years	2,04,565	29.82	20.36	0.00	99.00
Labourforce Participation	Labor force participation dummy variable ('1' if hourly earnings are observed, '0' otherwise)	2,69,335	0.20	0.40	0.00	1.00
Married	Married dummy variable ('1' if the individual is married, '0' otherwise)	2,69,335	0.35	0.48	0.00	1.00
Number of Children	Number of children in the household	2,04,568	1.84	1.66	0.00	18.00
Number of Adults	Number of adults in the household	2,04,568	3.41	1.70	0.00	18.00
Religion Categories	Religion Categories: '1' Hindu, '2' Other Minorities, '3' Muslims	2,69,335	1.49	0.68	1.00	3.00

Source: Author's calculation using IHDS II Data

We found that an individual's (natural log of) average hourly earnings were 2.98 with a standard deviation of 0.75. Average education was 5.32 years, with a standard deviation of 4.88 years. The average experience was 24.48 years, with a standard deviation of 18.81 years. Of the total sample, 35% were married. The average number of children was 1.84, and the adults were 3.41. Religion Categories comprised of '1' Hindu, '2' Other Minorities, and '3' Muslims.

Next, we proceed to estimate our models by religion.

8.5.2. Ordinary Least Squares Estimates IHDS II

Table 8. 6: Ordinary Least Squares estimates disaggregated by religion

Outcome Equation Dependent Variable: lnHourly_Earnings			
	Hindu	**Other Minorities**	**Muslim**
Education	0.0801***	0.0977***	0.0654***
	(0.00)	(0.00)	(0.00)
Experience	0.0302***	0.0375***	0.0469***
	(0.00)	(0.01)	(0.00)
Experiencesqr	-0.000301***	-0.000376***	-0.000681***
	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0376***	0.00168	0.0324***
	(0.00)	(0.01)	(0.01)
Constant	1.921***	1.928***	1.986***
	(0.02)	(0.07)	(0.05)
Total obs	35568	2415	4639
R square	0.231	0.244	0.145
Adj R square	0.231	0.243	0.144
F	1887.6	191.5	138.3
P value	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 8.6 provides OLS estimates disaggregated by religion. We observe that the coefficient for *Education* for Hindus (0.0801), Other Minorities (0.0977), and Muslims (0.0654). The coefficients were positive and significant for Hindus, Other Minorities, and Muslims. We also observe the coefficient for *Experience* for Hindus (0.0302), Other Minorities (0.375), and Muslims (0.0469). The coefficients were positive and significant for Hindus, Other Minorities, and Muslims.

The coefficient for Square of Experience was negative and significant for Hindus, Other Minorities, and Muslims. The coefficient for *Networking Intensity* for positive and significant for Hindu (0.0376) and Muslim (0.0324).

We find that the private returns to an additional year of education are lowest for Muslims and highest for Other Minorites. We find that the returns to an additional year of experience are lowest for Hindu and highest for Muslim. We also find that social capital, as measured by networking intensity, bears a positive and significant impact for Hindu and Muslim only.

8.5.3. **Heckman Selection Model (Maximum Likelihood Estimates)- IHDS II**

Table 8. 7: Heckman selection model estimates (using maximum likelihood estimator) disaggregated by religion

Outcome Equation Dependent Variable: lnHourly_Earnings			
	Hindu	**Other Minorities**	**Muslim**
Education	0.0784***	0.104***	0.0651***
	(0.00)	(0.00)	(0.00)
Experience	0.0324***	0.0340***	0.0469***
	(0.00)	(0.01)	(0.00)
Experiencesqr	-0.000343***	-0.000304***	-0.000679***
	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0357***	0.00741	0.0314***
	(0.00)	(0.01)	(0.01)
Constant	1.686***	2.537***	1.860***
	(0.03)	(0.11)	(0.07)

Selection Equation Dependent Variable: Labourforce Participation			
	Hindu	**Other Minorities**	**Muslim**
Married	0.0548***	-0.0208	0.0334
	(0.01)	(0.03)	(0.03)
Number of Children	0.0259***	0.0194	-0.0211***
	(0.00)	(0.01)	(0.01)
Number of Adults	-0.118***	-0.118***	-0.0721***
	(0.00)	(0.01)	(0.01)
Constant	0.261***	0.130**	-0.00538
	(0.01)	(0.05)	(0.03)
athrho	0.433***	-0.887***	0.188***

	(0.05)	(0.16)	(0.07)
insigma	-0.403***	-0.0808	-0.359***
	(0.01)	(0.07)	(0.02)
Total obs	74708	6187	11917
Selected obs	35568	2415	4639
Non selected obs	39140	3772	7278
rho	0.408	-0.710	0.186
Lambda (IMR)	0.272	-0.655	0.130
chi2	6338.6	685.6	544.8
P value	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 8.7 provides the Heckman selection model (MLE), estimates disaggregated by religion. The coefficients for *Education* were positive and significant across all religious groups, namely, Hindus (0.0784), Other minorities (0.104), and Muslims (0.0651). The coefficients for *Experience* were positive and significant across all religious groups, namely, Hindus (0.0324), Other minorities (0.0340), and Muslims (0.0469). The coefficients for *Experiencesqr* were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for *Networking intensity* were positive and significant only for Hindu (0.0357) and Muslim (0.0314).

8.5.4. Heckman Two-Step Selection Model (Two-Step Consistent Estimates) - IHDS II

Table 8. 8: Heckman two-step selection model estimates (two-step estimator) disaggregated by religion

Outcome Equation Dependent Variable: lnHourly_Earnings			
	Hindu	**Other Minorities**	**Muslim**
Education	0.0788***	0.0987***	0.0648***
	(0.00)	(0.00)	(0.00)
Experience	0.0330***	0.0359***	0.0470***
	(0.00)	(0.01)	(0.00)
Experiencesqr	-0.000351***	-0.000347***	-0.000679***
	(0.00)	(0.00)	(0.00)
Networking Intensity	0.0354***	0.00248	0.0303***
	(0.00)	(0.01)	(0.01)
Constant	1.626***	2.106***	1.719***
	(0.03)	(0.13)	(0.12)
Selection Equation Dependent Variable: Labourforce Participation			
	Hindu	**Other Minorities**	**Muslim**
Married	0.0616***	-0.0319	0.0413
	(0.01)	(0.04)	(0.03)
Number of Children	0.0186***	0.0293**	-0.0224***
	(0.00)	(0.01)	(0.01)
Number of Adults	-0.117***	-0.139***	-0.0705***
	(0.00)	(0.01)	(0.01)
Constant	0.261***	0.200***	-0.0137
	(0.01)	(0.05)	(0.03)
IMR	0.331***	-0.177	0.273**
	(0.03)	(0.11)	(0.11)
Total obs	74708	6187	11917
Selected obs	35568	2415	4639
Non selected obs	39140	3772	7278
rho	0.484	-0.231	0.376
Lambda (IMR)	0.331	-0.177	0.273
chi2	10173.1	777.0	757.0
P value	0.0000	0.0000	0.0000

Source: Author's calculations based on IHDS II
Note: Standard errors are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

Table 8.8 provides the Heckman selection model (MLE), estimates disaggregated by religion. The coefficients for *Education* were positive and significant across all religious groups, namely, Hindus (0.078), Other minorities (0.987), and Muslims

(0.0648). The coefficients for *Experience* were positive and significant across all religious groups, namely, Hindus (0.0330), Other minorities (0.0359), and Muslims (0.0470). The coefficients for *Experiencesqr* were negative and significant for all caste groups, signifying an inverted U-shaped relationship. The coefficients for *Networking intensity* were positive and significant only for Hindu (0.0354) and Muslim (0.0303).

8.6. PANEL DATA ANALYSIS IHDS I & II

In this section of the chapter, we proceed towards analyzing data from rounds 1 and 2 of IHDS, namely IHDS I & II (2004-05 and 2011-12). We merged the data from IHDS I and II and formed a panel data set with 150,995 individuals who were interviewed in both rounds. The analysis is restricted to individuals who were interviewed in both rounds of IHDS. We then merged this panel data set with household-level data and village-level data to form an elaborate panel dataset incorporating household and village attributes. The sub-sample selected for analysis in this chapter is restricted to individuals in the age group of 15 to 64 years as per the guidelines of Labour force participation followed by NSSO. Experience (Potential Experience) has been restricted to an upper bound of 45 years. Our analysis consists of descriptive statistics of the selected sample, followed by estimation of fixed effects multiple regression model by Consumption categories (as described in the previous section) using the Heckman Selection model addressing selection bias.

The model can then be specified as follows:

First stage equation:

$$\textbf{Labourforce Participation}_{it} = \alpha_0 + \alpha_1\textbf{Married}_{it} + \alpha_2\textbf{NChildren}_{it} + \alpha_3\textbf{Nadults}_{it} + \alpha_4(\textbf{X})_{it} + \varepsilon_{it} \qquad \textbf{(8.4)}$$

Where X = the set of exogenous explanatory variables of the outcome equation indicated in equation 8.5 and e = stochastic error term.

Second stage equation:

$$\boldsymbol{Ln\ Hourly\ Earnings_{it} = \beta_0 + \beta_1 Education_{it} + \beta_2 Experience_{it} + \beta_3 Experience_{it}^2 + \beta_4 Networking\ Intensity_{it} + \beta_5 IMR_{it} + \beta_6 Z_{it} + u_{it}} \quad \textbf{(8.5)}$$

Where IMR = the inverse Mills ratio generated from the first stage selection equation, Z = the set of included instrumental variables from the selection equation (eq 8.4) (Married, Nchildren, and Nadults), u= error term.

Descriptive statistics

Table 8. 9: Descriptive statistics

Variable	Description		Mean	SD	Min	Max	Obs
lnHourly_Earnings	Natural logarithm of hourly earnings observed	overall	3.02	0.82	-0.84	6.92	21413
		between		0.77	-0.84	6.71	15601
		within		0.28	1.10	4.94	1.37
Education	Completed years of education	overall	7.90	4.44	0	16	52134
		between		4.03	0	16	29878
		within		1.63	-0.10	15.90	1.74
Experience	Potential experience : {(age - education - 5) years}	overall	19.51	13.53	0	45	52378
		between		13.62	0	45	29889
		within		2.71	-1.99	41.01	1.75
Experience square	Square of potential experience	overall	563.77	629.60	0	2025	52378
		between		627.80	0	2025	29889

		within		138.98	-446.7	1574.27	1.75
Networking intensity	Number of social or economic committees that household members were associated with	overall	0.71	1.08	0	7	52225
		between		0.90	0	7	29870
		within		0.62	-2.79	4.21	1.75
Labor force participation	Binary variable ('1' if hourly earnings are observed, '0' otherwise)	overall	0.41	0.49	0	1	52378
		between		0.43	0	1	29889
		within		0.25	-0.09	0.91	1.75
Married	Binary variable ('1' if the individual is married, '0' otherwise)	overall	0.62	0.49	0	1	52378
		between		0.46	0	1	29889
		within		0.18	0.12	1.12	1.75
Children	Number of children in the household	overall	1.54	1.62	0	18	52378
		between		1.41	0	15	29889
		within		0.84	-4.96	8.04	1.75
Adults	Number of adults in the household	overall	3.47	1.74	0	18	52378
		between		1.50	0	18	29889
		within		0.90	-4.53	11.47	1.75

Religion	Categories:	overall	1.33	0.69	1	3	52378
	'1' if Hindu,	between		0.70	1	3	29889
	'2' if Other Minorities, and '3' if Muslims	within		0.10	0.33	2.33	1.75

Source: Author's calculations based on IHDS I and II

Table 8.9 provides the descriptive statistics. We found that the (natural log of) average hourly earnings of an individual (overall) was 3.02 with a standard deviation of 0.8. Average education was 7.9 years, with a standard deviation of 4.4 years (overall). The average experience was 19.5 years, with a standard deviation of 13.5 years (overall). Of the total sample, 62% were married, and 39% were female. The average number of children was 1.54, and the average number of adults was 3.47.

8.6.1. Heckman Estimates

Table 8. 10: Estimates from Fixed Effects- Heckman Selection model by Religion (dependent variable: Ln Hourly Earnings).

Main Equation	Hindu	Other Minorities	Muslim
Education	0.0377***	0.0998***	0.0215*
	(0.0049)	(0.0161)	(0.0118)
Experience	0.0137**	0.0364*	0.0250
	(0.0061)	(0.0259)	(0.0177)
Experience square	-0.0000914	-0.000352	-0.0000380
	(0.0001)	(0.0005)	(0.0003)
Networking Intensity	0.0223***	-0.0419	0.0127
	(0.0078)	(0.0455)	(0.0319)
2012	0.215***	0.677***	0.359**
	(0.0622)	(0.1652)	(0.1439)
IMR			
2005	-0.239***	0.224**	-0.188*
	(0.0509)	(0.1024)	(0.1070)
2012	-0.326***	-0.631***	-0.399***
	(0.0751)	(0.1880)	(0.1023)
Constant	1.591***	1.574***	1.630***
	(0.0616)	(0.1877)	(0.1853)

Source: Author's calculations based on IHDS I and II
Note: Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

We observe that the coefficient for Education for Hindus (0.0377), Other Minorities (0.0998), and Muslims (0.0215). The coefficients were positive and significant for Hindus, Other Minorities, and Muslims. We also observe the coefficient for Experience for Hindus (0.0137) and other Minorities (0.0364), and not significant for Muslims. The coefficients were positive and significant for Hindus and Other Minorities.

The coefficient for Networking Intensity for Hindus (0.0223) was positive and significant only for Hindus.

The coefficient for IMR (Inverse Mills Ratio) was significant for all categories, indicating the presence of selection bias. Hence, it justifies the use of Heckman's Two Step Correction model to address selection bias.

We find that the returns to an additional year of education are lowest for Muslims and highest for Other Minorites. We find that the returns to an additional year of experience are lowest for Hindus and highest for Other Minorities. Last, we find that social capital, as measured by networking intensity, positively impacts earnings for Hindus only.

Table 8. 11: Selection equation estimates from Fixed Effects- Heckman Selection model by Religion (dependent variable: Labourforce Participation).

	Hindu	Other Minorities	Muslim
Year Specific Education			
2005	0.0430***	0.119***	0.0108
	(0.0122)	(0.0275)	(0.0206)
2012	-0.0225**	-0.0178	-0.0873***
	(0.0091)	(0.0220)	(0.0236)
Year Specific Experience			
2005	0.0174	0.0910*	-0.0534
	(0.0168)	(0.0498)	(0.0362)
2012	0.0600***	0.00910	0.0415*
	(0.0115)	(0.0377)	(0.0223)
Year Specific Experience_square			
2005	-0.000635**	-0.000940	0.000540
	(0.0003)	(0.0008)	(0.0006)
2012	-0.00101***	-0.000350	-0.00145***
	(0.0002)	(0.0008)	(0.0005)
Year Specific Networking Intensity			
2005	-0.0251*	0.00497	-0.0171
	(0.0142)	(0.0523)	(0.0495)
2012	0.0135	0.0672	0.0104
	(0.0155)	(0.0605)	(0.0420)
Year Specific Married			
2005	-0.779***	-0.879***	-1.153***
	(0.0608)	(0.2155)	(0.1617)
2012	0.841***	0.711***	1.179***
	(0.0580)	(0.1841)	(0.1555)

Year Specific Nchildren			
2005	-0.0762***	-0.00823	0.00759
	(0.0129)	(0.0512)	(0.0292)
2012	0.000249	0.00284	-0.0562*
	(0.0091)	(0.0490)	(0.0311)
Year Specific Nadults			
2005	-0.0371***	0.00482	-0.0506*
	(0.0117)	(0.0581)	(0.0306)
2012	-0.00963	0.0507	-0.0109
	(0.0113)	(0.0468)	(0.0226)
Year-Specific Individual Average Education			
2005	-0.0320**	-0.0816***	0.0334
	(0.0135)	(0.0285)	(0.0237)
2012	0.0163*	0.0239	0.103***
	(0.0091)	(0.0231)	(0.0250)
Year-Specific Individual Average Experience			
2005	0.116***	0.0610	0.203***
	(0.0185)	(0.0549)	(0.0426)
2012	-0.00114	0.0670*	0.0131
	(0.0116)	(0.0402)	(0.0244)
Year Specific Individual Average Experience_square			
2005	-0.00159***	-0.00161*	-0.00289***
	(0.0003)	(0.0009)	(0.0008)
2012	-0.000228	-0.00139*	0.000370
	(0.0002)	(0.0008)	(0.0005)
Year-Specific Individual Average Networking Intensity			
2005	0.0369**	-0.0104	-0.105*
	(0.0149)	(0.0587)	(0.0590)
2012	0.0106	-0.0384	-0.0297
	(0.0189)	(0.0723)	(0.0486)
Year Specific Individual Average Married			
2005	0.724***	0.871***	1.111***
	(0.0664)	(0.2326)	(0.1850)
2012	-1.045***	-0.942***	-1.306***
	(0.0663)	(0.2187)	(0.1800)

Year Specific Individual Average Nchildren			
2005	0.129***	0.0262	-0.0328
	(0.0155)	(0.0551)	(0.0382)
2012	-0.000625	-0.0192	0.0272
	(0.0118)	(0.0525)	(0.0326)
Year Specific Individual Average Nadults			
2005	-0.131***	-0.138**	-0.00175
	(0.0153)	(0.0638)	(0.0383)
2012	-0.123***	-0.162***	-0.0694***
	(0.0121)	(0.0546)	(0.0245)
Year			
2012	1.501***	1.808***	2.092***
	(0.0679)	(0.1705)	(0.2165)
Constant	-1.442***	-1.893***	-2.392***
	(0.0669)	(0.1428)	(0.2082)
N	41509	3751	6729

Source: Author's calculations based on IHDS I and II
Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

The table above summarizes the results of the first stage equation of the Heckman model. In addition to the instruments used to estimate labor force participation, the model also estimates interactions between these instruments and other covariates by year. The instruments used for the first stage probit model; Married, Nchildren, and NAdults were also significant in the first stage equation. These estimates were jointly significant across the model.

8.7. PROJECTED EARNINGS BY RELIGION WITH EDUCATION AND EXPERIENCE AS COVARIATES

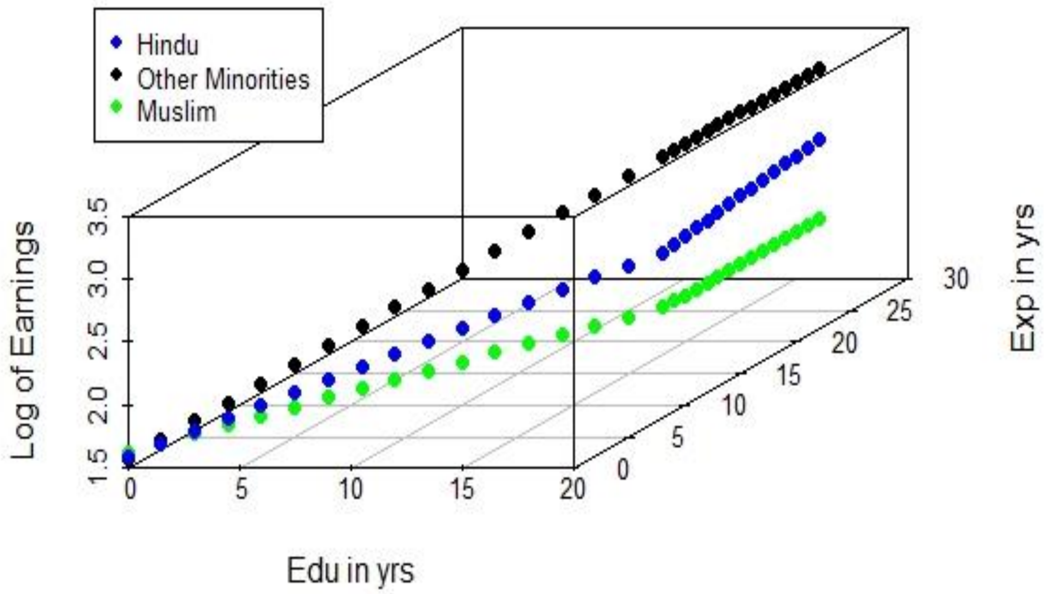

Figure 8. 1: Projected earnings by religion with education and experience as covariates.

Among the religious groups (Figure 8.1), Hindus (3.8%) and Muslims (2.3%) have a lower return than other minorities (9.9%). These projected earnings illustrate evidence of divergence with the gap between Muslims is and other communities is widening, suggesting the need for affirmative action to ensure reduced inequalities (SDG10).

8.8. FINDINGS

IHDS I :

The private returns to an additional year of Education in the OLS model range between the lowest for Muslims (8.07%) and highest for Other Minorities (11.%).The private returns to an additional year of Education in the Heckman model (MLE) range between the lowest for Muslims (7.78%) and highest for Other Minorities (11.6%). The private returns to an additional year of Education in the Heckman model (TSE) range between the lowest for Muslims (7.76%) and highest for Other Minorities (11.6%).

The private returns to an additional year of Experience in the OLS model range between the lowest for Other minorities (3.64%) and highest for Hindus (4.93%).The private returns to an additional year of Experience in the Heckman model (MLE) range between the lowest for Other minorities (4%) and highest for Hindu (5.36%). The private returns to an additional year of Experience in the Heckman model (TSE) range between the lowest for Other minorities (4%) and highest for Hindu (5.14%).

The Networking Intensity in all of the 3 models, OLS, Heckman (MLE) and Heckman (TSE) showed a negative significant impact for Other minorities.

IHDS II :

The private returns to an additional year of Education in the OLS model range between the lowest for Muslims (6.54%) and highest for Other Minorities (9.77.%).The private returns to an additional year of Education in the Heckman model (MLE) range between the lowest for Muslims (6.51%) and highest for Other Minorities (10.4%). The private returns to an additional year of Education in the Heckman model (TSE) range between the lowest for Muslims (6.48%) and highest for Other Minorities (9.87%).

The private returns to an additional year of Experience in the OLS model range between the lowest for Hindu (3.02%) and highest for Muslim (4.69%).The private returns to an additional year of Experience in the Heckman model (MLE) range between the lowest for Hindu (3.24%) and highest for Muslim (4.69%). The private returns to an additional year of Experience in the Heckman model (TSE) range between the lowest for Hindu (3.3%) and highest for Other Minorities (4.7%).

The Networking Intensity in all of the 3 models, OLS, Heckman(MLE) and Heckman(TSE) showed a positive significant impact for Hindu and Muslim.

Panel data IHDS I & II

Education was significant and positive for all three religious' groups. Experience, on the other hand, was significant and positive for Hindus and Other minorities but not for the other groups. Networking intensity was significant and positive for Hindus only.

8.9. CONCLUSION

We conclude that Education, and Experience have a positive and significant impact on earnings for all religions. Among the religious groups Hindus (3.8%) and Muslims (2.3%) have a lower return than other minorities (9.9%) in the panel data. These projected earnings illustrate evidence of divergence with the gap between Muslims and other communities is widening. We believe our findings are along the lines of Chakraborty and Bohara's (2021) study on the role of caste and religion in India and their recommendation for policy intervention for socially backward classes and Muslims to establish equity in the labor market. The government has implemented affirmative action policies in the domain of jobs and education for certain groups based largely on caste (Deshpande, 2013) and, more recently, economic backwardness (Nath, 2019). We find that despite education positively impacting earnings, the less privileged groups would require affirmative action to achieve reduced inequalities (SGD 10) and escape poverty (SGD 1).

The next chapter (Chapter 9) moves away from the discussion on earnings and education to assess the impact of natural disasters on educational attainment in India using a panel data approach. In Chapter 9, we use a new framework (difference in differences, a quasi-experimental approach) to estimate the average treatment effect of natural disasters on educational attainment, disaggregated by gender, class, caste, and religion. We believe the estimate would help better policy decisions in the context of climate change resilience and investment in education.

CHAPTER 9: IMPACT OF NATURAL DISASTERS ON EDUCATIONAL ATTAINMENT IN INDIA: A PANEL DATA ANALYSIS

In the previous four chapters (5-8), we estimated disaggregated estimated of private returns to education by class, gender, and social groups (Caste and Religion). We discussed the implications of lower returns to vulnerable sections within these social constructs.

There is a concern that the increase in frequency and severity of natural disasters caused by climate change could have detrimental effects. In this context, we believe that climate change leading to more frequent and intense natural hazards would have more serious implications for vulnerable sections. This chapter focuses on examining how such natural disasters impact the educational outcomes of students in rural India.

9.1. INTRODUCTION

Studies have found that natural disasters have disruptive effects on welfare in the form of widening inequalities, increased disparities, as well as long-term impacts on education. This impact on education can lead to human capital loss and adversely affect future growth and development. We use data from two of the latest rounds of the Indian Human Development Survey (IHDS) and employ a panel difference in difference regression model with continuous treatment fixed effects at the individual level. This allows us to examine the impact of natural disasters on education outcomes between 2004-05 and 2011-12 at the individual level. Our estimates improve upon all earlier studies on this theme that have relied on cross-section data or pseudo-panel data at the district level.

9.2. METHODOLOGY

9.2.1. Framework

We use a natural experimental study approach (Quasi-experimental design), using Natural Disasters as a continuous treatment to estimate the average treatment effect of natural disasters on Educational Attainment in India. A natural experiment is said to be a form of quasi-experimental design if an intervention where the researcher does not control the circumstances of implementation of the intervention itself (Leatherdale, 2019). Such approaches are most preferred in the absence of randomized controls in studies where researchers need to evaluate the effectiveness of programs or interventions (Chen et al., 2020; González et al., 2021). (Tamuly & Mukhopadhyay, 2022) use this approach to estimate the impact of natural disasters on the well-being (measured by consumption expenditure) of households in India. We follow (Tamuly & Mukhopadhyay, 2022) in constructing and using a continuous measure of the frequency of occurrence of natural disasters as *Natural Disaster Intensity (NDI)*, trust in public institutions as *Confidence Intensity*, membership in various socioeconomic groups as *Membership Intensity* (a measure of social capital) and other covariates used by them to estimate their impact on Educational attainment. In addition, we include *School fees*, *School Type (Government, Aided, Private)*, and *Repeated* (if an individual has ever repeated a class) as covariates to estimate their impact on Educational Attainment.

Multiple household and institutional characteristics determine educational attainment. These include household characteristics like household structure (proportion of children in the household) (Blake, 1989; Perkins, 2019), nature of the household indicated by the ownership of household assets (Cooper & Stewart, 2021; Filmer & Pritchett, 1999) and school type (Pianta & Ansari, 2018). However, a substantial variation in educational attainment remains unexplained after accounting for these. We believe these unexplained variations may partly be explained by observing the exposure of the villages to natural disasters. In our model, we incorporate an explanatory variable measuring the cumulative effect of natural disasters on households between 2006 and 2012 (*Natural Disaster Intensity, NDI*). These natural disasters are listed as floods, droughts, earthquakes, hailstorms, tsunamis, and droughts, among others). Past studies have recorded the role of gender (Arnold, 2015), economic status (Blanden, 2004), social groups (Borooah, 2012), religion, and ethnicity in determining educational attainment (Sander, 1992).

The structure of the difference-in-differences regression model can be illustrated as follows (equation 9.1):

$$\boldsymbol{Y_i = \beta_0 + \beta_1 Time\_Period_i + \beta_2 Treated_i + \beta_3 Time\_Period_i * Treated_i + \beta_4 Covariates_i + \varepsilon_i} \qquad \textbf{(9.1)}$$

In equation 9.1, Y_i is the observed value of the response variable, $Time\_Period_i$ is the binary time variable indicating pre or post-treatment, $Treated_i$ is the binary variable indicating the treatment, $Time\_Period_i * Trperiodeated_i$ is the interaction term between the time period and the treatment, $Covariates_i$ are a set of independent variables that possible affect the response variable, and ε_i is the error term.

The functional relationship that we wish to explore is as follows (equation 9.2):

Education = f (Natural Disasters Intensity (treatment), set of controls (Number of Children in Household, Number of Household Assets, Economic cost of education, Institutional type, Membership Intensity, Confidence Intensity, Financial resilience, Repetition of a class) **(9.2)**

In equation 9.2, Education is the dependent variable measured in an individual's years of completed education, which depends on Natural disaster intensity, Household characteristics (*Number of Children in Household, Number of Household Assets)*, Economic cost of Education (School Fees) and Institutional type (Government, Aided and Private), attainment *Membership Intensity* (construct to measure social capital), Confidence in public institutions (construct to measure the level of trust in public institutions), Financial resilience (binary indicators for Life insurance and Health Insurance), and Repetition of class (binary indicator of having repeated a class)

Our objective is to estimate the impact of natural disasters (with differential intensity) on educational attainment in households in rural India. We provide estimates disaggregated by *Gender* (Male and Female), *Consumption quintiles* (Below the poverty line (BPL) and quintiles above the poverty line, from APL1 to APL5), Caste (General, Other backward castes, Scheduled castes, and Scheduled tribes) and *Religion* (Hindu, Other minorities and Muslim). This disaggregation, we believe, would provide more robust estimates to analyze the impact of natural disasters on educational Attainment across sections of the sub-population.

Our analysis is restricted to rural India (because natural disaster data is available at the village level), featuring individuals and households that were in both rounds of IHDS (IHDS I, 2004-05 and IHDS II, 2011-12) and were between 6 to 23 years of age (to restrict the analysis to probable students who were still in the educational system or just exited).

9.2.2. Model

The basic intuition behind the use of a difference-in-differences (DiD) is to evaluate the differences in the causal effect of an event between a treatment group (those affected by the event) and a control group (those unaffected by the event). In such a case, one could infer that if the event had never occurred, then differences between the treatment group and control group would have stayed the same. *Natural Disaster Intensity* serves as a treatment that is similar to 'natural program intervention.' In contrast to conventional DiD models that use a binary treatment variable, we follow the use of a continuous treatment effect. This enables us to capture the slope effect (incremental change due to unit increase in the intensity of the treatment, Natural Disasters) of the treatment in addition to the level effect. In other words, we would be able to capture the marginal change in educational attainment for unit changes in the intensity of natural disasters.

The specification equation is stated as (equation 9.3):

$$\begin{aligned} Education_{it} = {} & \alpha_0 + \alpha_1\, Natural\ Disaster\ Intensity_{it} + \alpha_2\, Number\ of\ children\ in\ Household_{it} + \alpha_3\, Number\ of\ Household\ Assets_{it} + \alpha_4\, School\ Fees_{it} + \alpha_5\, Government\ School_{it} \\ & + \alpha_6\, Aided\ School_{it} + \alpha_7\, Private\ School_{it} + \alpha_8\, Membership\ Intensity_{it} + \alpha_9\, Confidence\ Intensity_{it} + \alpha_{10}\, Life\ Insurance_{it} + \alpha_{11}\, Health\ Insurance_{it} + \alpha_{12}\, Repeated_{it} \\ & + \alpha_{13}\, Year_{it} + \varepsilon_{it} \end{aligned} \tag{9.3}$$

Where,

Education = Years of Education

Natural Disaster Intensity = Ordered rank variable (range lies between 0 to 18)

Number of children in household = Number of children in the household

Number of Household Assets = Number of household assets owned by the household

School Fees = School fees paid by households (in terms of 2011-12 values)

School Type = Type of educational institution (Others (1), Government (2), Aided (3), Private (4))

Membership Intensity = Ordered rank variable for household membership in one or more social or economic membership groups (range lies between 0 to 9)

Confidence Intensity = Ordered rank variable for a household's level of Confidence in public institutions (range lies between 10 to 30)

Life Insurance = Dummy variable for financial resilience (Yes (1), No (0)) of household

Health Insurance = Dummy variable for financial resilience (Yes (1), No (0)) of household

Repeated = Dummy variable for an individual repeating the same class (Yes (1), No (0))

Year = Year indicator

εit = Residual

We then provide the disaggregated estimates obtained from the above model by *Gender* (Male and Female), *Consumption quintiles* (Below the poverty line and quintiles above the poverty line), Caste (General, Other backward castes, Scheduled castes, and Scheduled tribes), and *Religion* (Hindu, Other minorities and Muslim).

9.3. ANALYSIS

9.3.1. Descriptive Statistics

Table 9. 1: Descriptive Statistics

Variable		Mean	Standard Deviation	Minimum	Maximum	Observations
Education	*overall*	5.41	3.85	0.00	16.00	74260
	between		3.35	0.00	16.00	50139
	within		2.12	-2.59	13.41	1.48
Natural Disaster Intensity	*overall*	1.07	2.10	0.00	18.00	66763
	between		1.73	0.00	18.00	47301
	within		1.34	-7.93	10.07	1.41
Number of children in Household	*overall*	2.38	1.86	0.00	18.00	74440
	between		1.65	0.00	18.00	50235
	within		0.94	-5.62	10.38	1.48
Number of Household Assets	*overall*	11.34	5.46	0.00	29.00	74429
	between		5.18	0.00	29.00	50230
	within		2.01	2.34	20.34	1.48
School Fees	*overall*	1465.36	6995.03	0.00	860000.00	48198
	between		4822.34	0.00	433589.00	35906
	within		4520.78	-424945.60	427876.30	1.34
Location	*overall*	0.00	0.00	0.00	0.00	74440
	between		0.00	0.00	0.00	50235
	within		0.00	0.00	0.00	1.48
School Type	*overall*	1.90	0.98	1.00	4.00	74440
	between		0.90	1.00	4.00	50235
	within		0.48	0.40	3.40	1.48
Membership Intensity	*overall*	0.66	1.10	0.00	9.00	74232
	between		1.00	0.00	9.00	50157
	within		0.53	-3.84	5.16	1.48
Confidence Intensity	*overall*	16.60	3.34	10.00	30.00	69772
	between		2.98	10.00	30.00	48749
	within		1.79	7.10	26.10	1.43
Health Insurance	*overall*	0.07	0.25	0.00	1.00	74360
	between		0.22	0.00	1.00	50201
	within		0.14	-0.43	0.57	1.48

Life Insurance	*overall*	0.21	0.40	0.00	1.00	74395
	between		0.37	0.00	1.00	50217
	within		0.20	-0.29	0.71	1.48
Ever Repeated	*overall*	0.13	0.34	0.00	1.00	66778
	between		0.30	0.00	1.00	46008
	within		0.18	-0.37	0.63	1.45
Female	*overall*	0.44	0.50	0.00	1.00	74440
	between		0.50	0.00	1.00	50235
	within		0.04	-0.06	0.94	1.48
Consumption Quintiles	*overall*	1.97	1.65	0.00	5.00	74425
	between		1.50	0.00	5.00	50227
	within		0.77	-0.53	4.47	1.48
Caste Categories	*overall*	2.16	1.07	1.00	4.00	74403
	between		0.92	1.00	4.00	50225
	within		0.63	0.66	3.66	1.48
Religion Groups	*overall*	1.31	0.68	1.00	3.00	74440
	between		0.67	1.00	3.00	50235
	within		0.08	0.31	2.31	1.48
Year	*overall*	2008.57	3.50	2005.00	2012.00	74440
	between		2.52	2005.00	2012.00	50235
	within		2.82	2005.07	2012.07	1.48
Age	*overall*	13.93	5.00	6.00	23.00	74440
	between		4.59	6.00	23.00	50235
	within		2.69	5.43	22.43	1.48

Source: Author's calculation from IHDS I & II

In Table 9.1 (descriptive statistics), we find that an individual's average educational attainment is 5.41 years (overall standard deviation of 3.85). The average number of children in the household was 2 (overall standard deviation of 1.86 years), and the average intensity of natural disasters was 1.07 (overall standard deviation of 2.10). The range for natural disaster intensity lies between 0 and 18; zero is the lowest intensity (absence of natural disasters for the period in consideration), and 18 is the highest intensity. On average, school fees were 1465.36 (overall standard deviation of 6995.03).

9.3.2. Disaggregated estimates of ATET by Gender

Table 9. 2: Disaggregated estimates of ATET by Gender

Variables *Dependent: Education*	Male	Female
ATET		
Natural Disaster Intensity	-0.0136	-0.0221**
	(0.0089)	(0.0095)
Controls		
Number of Children in Household	-0.0768***	-0.0607***
	(0.0175)	(0.0166)
Number of Household Assets	0.0212***	0.0313***
	(0.0064)	(0.0071)
School fees	-0.00000136	0.000000552
	(0.0000)	(0.0000)
School Type		
Government	0.0939	0.0151
	(0.0603)	(0.0590)
Aided	0.236***	0.323***
	(0.0809)	(0.0824)
Private	-0.195***	-0.144**
	(0.0687)	(0.0706)
Membership Intensity	0.00590	-0.0120
	(0.0179)	(0.0178)
Confidence Intensity	0.00527	0.0115**
	(0.0049)	(0.0054)
Health Insurance	0.0429	-0.0195
	(0.0654)	(0.0690)
Life Insurance	0.0851**	-0.0124
	(0.0427)	(0.0444)
Repeated	-0.394***	-0.433***
	(0.0555)	(0.0627)
Year (2012)	5.919***	6.111***
	(0.0531)	(0.0555)
Constant	2.549***	1.592***
	(0.1301)	(0.1381)
Number of Observations	22298	17465

Source: Author's calculation

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We provide disaggregated estimates for the difference in differences estimator with the fixed effects and control covariates by gender (Table 9.2). We first observed that the

Average Treatment Effect on Treated (ATET) was negative and highly significant for females (-0.022). It indicated that females who were exposed to natural disasters experienced lower educational attainment. Secondly, the coefficient indicates that a unit increase in the *Natural Disaster Intensity* would lower the educational attainment of females who were exposed to natural disasters on average by *(0.02 years)*.

We also find that the *Number of children in the household* negatively impacts educational attainment for males (-0.077) and females (-0.06). *The number of household assets* positively impacts educational attainment for males (0.02) and females (0.03). School fees had a positive impact on educational attainment for males and females. However, Aided schools positively affect educational attainment, and Private schools negatively impact educational attainment for both males and females in comparison to the reference category. *Confidence Intensity* (0.012) positively affects female educational attainment, whereas the indicator of *Life Insurance* for households positively influences educational attainment for males.

9.3.3. Disaggregated estimates of ATET by Consumption Quintiles

Table 9. 3: Disaggregated estimates of ATET by Consumption Quintiles

Variables Dependent: Education	BPL	APL1	APL2	APL3	APL4	APL5
ATET						
Natural Disaster Intensity	-0.0591**	-0.000462	0.0692	0.0367	-0.101***	-0.0626*
	(0.0272)	(0.0362)	(0.0446)	(0.0417)	(0.0388)	(0.0344)
Controls						
Number of Children in Household	-0.101***	-0.156**	-0.107	0.0394	0.149	-0.149**
	(0.0329)	(0.0657)	(0.0841)	(0.0962)	(0.1570)	(0.0608)
Number of Household Assets	0.0265*	0.0508**	0.000748	0.0489	0.0258	0.00631
	(0.0142)	(0.0255)	(0.0443)	(0.0338)	(0.0354)	(0.0275)
School fees	0.0000520	0.0000963**	0.00000915	-0.00000980	0.0000125	-0.00000027
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
School Type						
Government	0.0867	0.504**	-1.172**	0.312	0.536**	0.0231
	(0.2019)	(0.2227)	(0.4825)	(0.3020)	(0.2598)	(0.2347)
Aided	0.196	0.411	-0.222	-0.214	0.252	-0.0633
	(0.2556)	(0.2864)	(0.5161)	(0.3547)	(0.4545)	(0.4167)
Private	-0.150	-0.0623	-1.260***	-0.248	0.186	0.107
	(0.2591)	(0.3016)	(0.4401)	(0.3852)	(0.2413)	(0.1798)
Membership Intensity	0.0303	-0.106	0.0560	0.0138	0.0697	-0.00357
	(0.0615)	(0.0694)	(0.0964)	(0.0696)	(0.0677)	(0.0582)
Confidence Intensity	0.0241**	0.00877	0.0296	-0.0193	0.0207	0.00450
	(0.0122)	(0.0169)	(0.0279)	(0.0260)	(0.0235)	(0.0227)
Health Insurance	0.118	-1.124***	0.891***	0.267	-0.463*	-0.192
	(0.1230)	(0.3673)	(0.2609)	(0.2277)	(0.2755)	(0.3250)
Life Insurance	0.160	-0.0426	-0.0345	-0.0111	0.261*	-0.120
	(0.1534)	(0.1940)	(0.2912)	(0.1799)	(0.1579)	(0.1694)
Repeated	-0.464***	-0.0916	-0.266	-0.422	-0.625**	0.0189
	(0.1156)	(0.1824)	(0.2790)	(0.2871)	(0.2500)	(0.4737)
Year (2012)	6.223***	5.749***	5.465***	6.098***	6.517***	5.948***
	(0.1007)	(0.1823)	(0.2780)	(0.2728)	(0.3526)	(0.2172)
Constant	0.813***	2.258***	3.233***	1.835**	1.277***	3.803***
	(0.2971)	(0.4505)	(0.8505)	(0.7172)	(0.4869)	(0.6294)
Number of Observations	9325	7880	6774	6160	5295	4329

Source: Author's calculation

Standard errors in parentheses

* p < 0.1, ** p < 0.05, *** p < 0.01

We provide disaggregated estimates by consumption quintiles (Table 9.3). We find the Average Treatment Effect on Treated (ATET) was negative and highly significant for BPL (-0.05), APL4 (-0.10), and APL5 (-0.06). We are indicating that individuals from these consumption quintiles who were exposed to natural disasters experienced lower educational attainment. Secondly, we find that a unit increase in the *Natural Disaster Intensity* would lower the educational attainment of BPL (0.05 years), APL4 (0.10 years), and APL5 (0.06 years) compared to their respective control groups within consumption quintiles.

The *Number of children in the household* has a negative and significant impact on educational attainment for BPL (-0.10), APL1 (-0.16), and APL5 (-0.15). The *Number of Household Assets has a positive and significant impact on educational attainment for BPL (0.03)* and APL1 (0.05). *School fees* positively and significantly impact educational attainment for APL1. In School Type, *Government school* is positive and significant for APL1 and APL4, except negative for APL2, and *Private school* is negative and significant for APL2. *Confidence Intensity* is positive and significant for BPL. *Health Insurance* is significant and negative for APL1, APL4, and positive for APL2. *Life Insurance* is negative and significant for APL4. *Repeated* is negative and significant for BPL and APL4.

9.3.4. Disaggregated estimates of ATET by Caste

Table 9. 4: Disaggregated estimates of ATET by Caste categories

Variables *Dependent: Education*	General	Other Backward Castes	Scheduled Castes	Scheduled Tribes
ATET				
Natural Disaster Intensity	-0.0161	-0.172*	-0.151**	-0.00598
	(0.0197)	(0.0961)	(0.0712)	(0.0794)
Controls				
Number of Children in Household	-0.0406	-0.156*	0.257*	-0.0733
	(0.0303)	(0.0797)	(0.1552)	(0.2187)
Number of Household Assets	0.0199	0.0463	0.132*	-0.0266
	(0.0155)	(0.0371)	(0.0721)	(0.1034)
School fees	0.00000107	-0.0000212	0.000377**	-0.0000220
	(0.0000)	(0.0000)	(0.0002)	(0.0001)
School Type				
Government	0.143	-0.501	0.900	-0.530
	(0.1209)	(0.4169)	(0.8627)	(0.7248)
Aided	0.147	-1.085	1.637*	-1.424**
	(0.1966)	(0.8591)	(0.9016)	(0.6681)
Private	-0.0861	-0.548	0.188	-0.697
	(0.1284)	(0.4475)	(0.9199)	(0.6287)
Membership Intensity	-0.0323	0.00304	-0.344	-0.789***
	(0.0443)	(0.0997)	(0.2678)	(0.2324)
Confidence Intensity	0.0104	0.0144	0.0381	-0.175**
	(0.0108)	(0.0294)	(0.0741)	(0.0789)
Health Insurance	-0.135	1.435***	0.361	1.493*
	(0.1245)	(0.4309)	(0.4312)	(0.8678)
Life Insurance	-0.0463	-0.162	-1.363**	-0.447
	(0.0848)	(0.3112)	(0.5538)	(1.0419)
Repeated	-0.531***	-0.749**	0.0553	1.106
	(0.1254)	(0.3039)	(0.4897)	(0.9725)
Year (2012)	6.137***	5.889***	6.490***	6.352***
	(0.1072)	(0.2597)	(0.3805)	(0.5139)
Constant	2.765***	2.150***	-2.131	7.316***
	(0.2810)	(0.8158)	(2.0478)	(1.8593)
Number of Observations	14017	12506	6432	6808

Source: Author's calculation

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Next, we provide disaggregated estimates by Caste (Table 9.4). We find the Average Treatment Effect on Treated (ATET) was negative and highly significant for Other

Backward Castes (-0.17) and Scheduled Castes (-0.15). Also, we find that a unit increase in the *Natural Disaster Intensity* would lower the educational attainment of these marginalized castes (Other Backward Castes by 0.17 years and Scheduled Castes by 0.15 years).

The *Number of children in the household* has a significant impact on educational attainment for Other Backward Castes (-0.16), and Scheduled Castes (0.26). The *Number of household assets* has a positive and significant impact on educational attainment for Scheduled Tribes. *School fees* were positive and significant for Scheduled Castes. In-school Types: *Aided schools* are seen to have a positive impact on educational attainment for Scheduled castes and a negative for Scheduled Tribes. *Membership Intensity* is negative and significant for Scheduled Tribes. *Confidence intensity* is negative and significant for Scheduled Tribes. *Health insurance* is positive and significant for Other Backward Castes and Scheduled Tribes, and *Life insurance* is negative for Scheduled castes. *Repeated* is negative and significant for General and Other Backward Castes.

9.3.5. Disaggregated estimates of ATET by Religion

Table 9. 5: Disaggregated estimates of ATET by Religion

Variables *Dependent: Education*	Hindu	Other Minorities	Muslim
ATET			
Natural Disaster Intensity	-0.00697 (0.0068)	-0.0852** (0.0389)	-0.0892*** (0.0253)
Controls			
Number of Children in Household	-0.0507*** (0.0132)	-0.158*** (0.0455)	-0.154*** (0.0429)
Number of Household Assets	0.0257*** (0.0051)	0.0275 (0.0184)	0.0553*** (0.0187)
School fees	0.000000423 (0.0000)	-0.00000125** (0.0000)	0.0000150 (0.0000)
School Type			
Government	0.0741 (0.0459)	0.439*** (0.1577)	-0.122 (0.1859)
Aided	0.278*** (0.0614)	-0.00913 (0.2798)	0.170 (0.3008)
Private	-0.140** (0.0548)	-0.154 (0.1797)	-0.444** (0.1803)
Membership Intensity	-0.00633 (0.0134)	0.0115 (0.0664)	0.00223 (0.0450)
Confidence Intensity	0.00846** (0.0038)	0.0235 (0.0175)	0.00353 (0.0146)
Health Insurance	0.0164 (0.0491)	0.377** (0.1851)	-0.310 (0.2706)
Life Insurance	0.0530 (0.0327)	-0.190 (0.1255)	0.0409 (0.1368)
Repeated	-0.456*** (0.0434)	-0.374** (0.1690)	-0.0228 (0.2080)
Year (2012)	6.064*** (0.0414)	5.888*** (0.1498)	5.432*** (0.1514)
Constant	2.064*** (0.1020)	2.380*** (0.4155)	1.991*** (0.3595)
Number of Observations	32846	2948	3969

Source: Author's calculation

Standard errors in parentheses

* p < 0.1, ** p < 0.05, *** p < 0.01

Finally, we provide disaggregated estimates by Religion (Table 9.5). We find the Average Treatment Effect on Treated (ATET) was negative and highly significant for Other Minorities (-0.085) and Muslims (-0.089). Also, we find that a unit increase in the *Natural Disaster Intensity* would lower the educational attainment of these religious minorities (Other Minorities by 0.085 years and Muslims by 0.089 years).

The *Number of children in the household* has a negative and significant impact on educational attainment for all religions. The *Number of household assets* has a positive and significant impact on educational attainment for Hindus and Muslims. *School fees* negatively and significantly impact educational attainment for Other minorities. In school Types, Government schools are seen to have a positive impact on educational attainment for other minorities, Aided schools are seen to have a positive impact on educational attainment for Hindus, and Private schools are seen to impact educational attainment for Hindus and Muslims negatively. *Confidence intensity* is positive and significant for Hindus. *Health insurance* is positive and significant for other minorities. *Repeated* is negative and significant for Hindus and Other minorities.

9.4. FINDINGS

We find that *Natural Disaster Intensity* (NDI) yields a negative and significant Average Treatment Effect on the Treated (exposed to natural disasters) on *Educational Attainment*. (Zhang & Zhang, 2023) Use a cohort-based DiD approach to estimate that the floods of 1975 in Zhumadian City of China reduced locals' schooling by 0.2 years. Using the DiD approach, (Di Pietro, 2018) found that the probability of dropping out for students affected by the L'Aquila earthquake was 3.4 percent higher, while the probability of on-time graduation was 4.7 percent lower. Similarly, (Thamtanajit, 2020) finds that floods had a significant negative effect on all test scores (a proxy for academic achievement) for grade 6, except for social studies. (Paudel & Ryu, 2018) finds significant loss of human capital in Eastern Nepal post-1988 earthquake. Our findings from disaggregated estimates add to these results by illustrating the severity of these losses to educational attainment for marginalized sections in rural India.

Our results are consistent for Females, three consumption quintiles (BPL, APL4, and APL5), two marginalized caste categories (OBC and SC), and religious minorities (Muslims and Other minorities).

We also find that the number of children in the household negatively and significantly impacts educational attainment for both genders, consumption quintiles (BPL, APL1, and APL5), social groups (OBCs), and all religions. This again confirms the results from the received literature (Blake, 1989).

The *Number of household assets* has a positive and significant impact on educational attainment for both genders, consumption quintiles (BPL and APL1), social groups (SCs), and religion (Hindus and Muslims). This aligns with extensive literature in this area on the links between household assets (financial and non-financial) and educational attainment (Kim & Sherraden, 2011).

We find that *school fees have a positive and significant impact on educational attainment for consumption quintiles (APL1) and social groups (SCs)* and a negative and significant impact on religion (other minorities). Earlier studies have found that higher college fees led to a rise in degree completion (Pascarella et al., 1992) but reduced enrolment (Bietenbeck et al., 2020). However, we did not test for changes in enrolment.

Apart from school fees, we also looked at *school type* since it captures additional indicators beyond school fees, including the learning atmosphere. In *school type,* our reference category was other schools. We find that *Government* and *Aided* schools had a positive and significant impact on educational attainment across gender, consumption quintiles (except APL2, -ve), caste (SCs, but negative for STs), and religion (Hindus). On the other hand, *Private* schools negatively impacted educational attainment for both genders, APL2, Hindus, and Muslims.

Membership Intensity negatively and significantly affects STs. Confidence Intensity positively and significantly affects women, BPL, and Hindus, whereas it negatively affects STs. Financial resilience in the form of Health insurance positively and significantly affects APL3 and other minorities, whereas it negatively affects APL1 and APL4, OBCs, and STs. Financial resilience in the form of Life Insurance positively and significantly affects men, APL4, whereas it negatively affects SCs. Repeating in a class negatively and significantly affects both genders, BPL and APL4, general and OBCs, and Hindus and Other minorities.

9.5. CONCLUSION

Our findings reaffirm that climate change (SDG13) vis a vis natural disasters have a significant negative effect on educational attainment. We additionally conclude that the impact of natural disasters on educational attainment for marginalized sections (women, BPL, OBCs, SCs, minority religions) of the society is much more severe. On the one hand, lower educational attainment (SDG4) in disaster-prone areas could trigger a chain of disruptive welfare effects in the form of lower overall earnings, widening of gender differences (in attainment and earnings) (SDG5), and higher societal inequalities (SDG10). On the other hand, these lower educational attainments reduce the effectiveness and increase the cost of disaster management programs developed for mitigation from natural disasters and to reduce vulnerabilities. Studies have shown that higher education levels lead to better awareness levels and ease of dissemination of information and serve as a measure of development with direct and indirect effects on reducing fatality rates (Parida et al., 2021). The higher vulnerability combined with the increasing frequency of natural disasters in recent times and the inability to migrate leaves the marginalized in danger of intergenerational disruptions in well-being. This possibly creates a spiral of lower education, lower earnings, and lower welfare, inevitably pushing them to poverty. The impacts of climate change on human capital demand effective policy intervention to mitigate human capital losses to marginalized sections. The challenge then is to create a targeted response to more vulnerable sections of the population by 1) minimizing dropout rates post disasters, 2) strengthening schooling mechanisms to resume educational programs rapidly, and 3) reducing the probability of household's decisions to switch from investment in human capital (education) to rebuilding damages to physical assets by providing definitive monetary assistance in rehabilitation.

In the next chapter (chapter 10), we discuss our findings from all chapters and the completion of our objectives. We additionally discuss various policy implications arising from our findings. Chapter 10 is our findings and conclusion chapter.

CHAPTER 10: FINDINGS AND CONCLUSION

In the previous chapter, we discussed the impact of natural disasters on educational attainment with a difference in differences (DiD) method using panel data for India.

In this thesis, we first re-examined the issue of private returns to education in India using two rounds of IHDS (nationally representative data) in a cross-sectional and panel setting. Secondly, we also estimated the impact of natural disasters (with differential intensity) on educational attainment in rural India. We provided estimates disaggregated by Gender (Male and Female), Consumption quintiles (Below the poverty line (BPL) and quintiles above the poverty line, from APL1 to APL5), Caste (General, Other Backward Castes, Scheduled Castes and Scheduled Tribes) and Religion (Hindu, Other minorities and Muslim) for private returns to education and the impact of natural disasters on educational attainment.

10.1. INTRODUCTION

Growth in human capital investment (like quality of institutions) helps accumulate human capital which can lead to reduced income inequalities, this eventually can alter a nation's development path (Dias & Tebaldi, 2012). Guaitoli (2000) ascertains that human capital can link per capita growth rates and inequalities in income distribution. Investing in a nation's human capital is essential for its long-term economic growth and development. Firstly, it involves prioritizing education with targeted policies to influence human capital growth, which can lead to economic progress. Alternatively, human capital in the form of education is the driver of innovation and is essential for technology adoption or creation. Then a nation's human capital can be viewed as a critical factor to generate sustained economic growth and economic development. This implies a more significant investment in education and skills and use of a targeted policy intervention to influence the growth of human capital and, thereby, economic growth and development. Quality education is crucial to human capital investment, and sustainable development goals highlight its importance.

In this context, it becomes important from an individual's perspective to understand the private returns to an additional year in the education system. One hand, there are direct costs of staying enrolled in the education system and on the other the opportunity cost of foregone earnings. The main question is: what kind of impact do private investments in education have on an individual's earnings? Additionally, we ask whether a girl going to school would realise similar returns as a boy? Studies have revealed significant gender-based differences in earnings (Bhaumik & Chakrabarty, 2008; Blau, 2012; Gideon et al., 2022; Mendiratta & Gupt, 2013). Moreover, from a household's perspective of resource allocation and decision making, investing in a child's education then becomes pivotal. Would a child from a poor household experience similar gain as those from the more economically sound? In this regard, we attempt to estimate private returns across economic distributions using household consumption data. Lastly, from a policymaker's perspective these estimates to private returns to education and experience, provide insight for development of effective and targeted intervention programs that could foster the nation's human capital stock.

Furthermore, India has strong social hierarchies, in caste and religion. the influence of these factors on earnings and education are well established in literature (Agte & Bernhardt, 2023; Bhaumik & Chakrabarty, 2006; Borooah, 2012; Dhesi & Singh, 1989; Rani, 2013). For the marginalized educational attainment and higher earnings, serves as a path to overcome social and identity barriers. Investment in education and the returns can then be seen, in this context beyond private returns. Additionally, there are social gain; like lower crime rates, better health among others. It is also true that social capital plays an important role for the marginalized. (Chong et al., 2020; Dinda, 2008; Hunter, 2016).

We examine these issues by estimating private returns to education, and experience, while asserting the impact of social capital (measured by networking intensity). We, estimate private returns disaggregated by gender, class, caste and religion. Such disaggregation would better help understand the impacts of human capital investment across dimensions critical to the Indian context.

Given the social hierarchies and labor market dynamics in India; we then delved into trying to assess, the impact of disruptions to educational attainment in the form of natural disasters. Chari (2010) highlights India's high vulnerability to natural disasters as earthquakes (60 percent), land susceptible flooding (40 million hectares), area prone to droughts (68 percent), and two-thirds of its coastline exposed to cyclones. This combined with climate change and the likely increase in frequency and intensity of natural disasters put India at serious risk. In this context, we examine the impact of natural disasters (with differential intensity) on educational attainment across gender, class, caste and religion using panel difference in differences (DiD) regression model with continuous treatment and fixed effects at the individual level for rural India.

In light of India's commitments to targets for sustainable goals, it was be interesting to study this relationship that links SDG 13 (climate change) to SDG 4 (quality education). Additionally, our disaggregated nature of estimates revealed significant findings pertaining to SDG 5 (gender equality), and SDG 10 (reducing inequalities). Our use of a nationally representative data to cover interrelated topics between earnings, education, and natural disasters; we believe is a significant contribution to literature.

We next, discuss our findings for each objective.

10.2. OBJECTIVES AND CHAPTER-WISE FINDINGS AND CONCLUSIONS

10.2.1. Objective 1: To Estimate The Impact Of Human Capital Investment On Earnings In India. (Chapter 4)

The first Objective was structured to address our first research question, that was to quantify private returns to human capital investment (education), and potential job market experience, in India. We examined the methodology utilized in the derivation of these estimates in an attempt to address the potential problem of selection bias that has been highlighted in literature. We compared estimates of private returns to education in India using OLS method and Heckman's method. We used two specifications of the Heckman model. The first specification uses the maximum likelihood estimator to derive the estimates. Whereas the second specification is a two-step correction estimator that is used in Heckman (1979). We estimated and compared the estimates using both cross-sectional and panel data sets.

Additionally, we tried to assess the impact of social capital on earnings. Our constructed variable to capture social capital was Networking Intensity. We did this by using data on household membership in various social and community groups like NGOS, religious groups, caste groups and cooperative associations among others.

Our findings for this objective are pertinent to 1) estimates of private returns to an additional year of education and experience in India, 2) methodology used to estimate private returns to education and 3)impact of social capital on earnings of individuals.

Through a thorough study of the IHDS I and IHDS II (cross-sectional data set), as well as panel construct of the merged IHDS I & II, we find that that estimate of returns to an additional year of education, if estimated by Ordinary Least Squares (OLS) lead to the problem of selection bias. This, in turn, provides inaccurate estimates and very often overestimates the impact of education on earnings. The way forward is to estimate these returns to an additional year of education using the Heckman selection correction mechanism that addresses selection bias.

The estimates from IHDS I do show evidence of selection bias in the model. However, the coefficient for education, which is the private returns to education for an additional year, is similar in all estimated models (around 10 percent).

The estimates from IHDS II also show evidence of selection bias in the model. However, the coefficient for education, which is the private returns to education for an additional year, is similar in all estimated models (8.3 percent).

In the Heckman selection model with fixed effects using the xtheckmanfe command, we found that the within estimator (OLS) suffers from selection bias with the coefficient for *IMR* for 2005 (-0.2213) and 2012 (-0.3799) being highly significant.

The coefficients for *Education* were positive and significant in the within estimator (OLS) (0.0666) and Heckman selection model (0.0348). This illustrates that the within estimator (OLS) overestimated the coefficients for private returns to an additional year of education when compared with the Heckman model with fixed effects. The coefficient for *Education* was nearly twice as large in the within estimator (OLS).

The coefficient for *Experience* was positive and significant in both the models as well; however, the estimates were overestimated in the within estimator (OLS) (0.0608) as compared to the Heckman selection model (0.012) with fixed effects. The coefficient for square of experience was negative and significant in the within the estimator (OLS) but not significant in the Heckman model.

The variable *Networking Intensity* was positive and highly significant in the within estimator (OLS) (0.0174) and the Heckman model (0.0164); the coefficient was marginally higher for the OLS model. The selection instruments were highly significant (except Nadults 2012) in the Heckman-type model. The number of observations in the within estimator (OLS) is reported as 21291, and in the Heckman, it is reported as 51989.

We also find that *Education* and *Experience* are highly significant explanatory variables in explaining changes in earnings for individuals. We find that both education and experience positively impact earning in the overall data. We find evidence for an inverted U-shaped relationship between Earnings and Experience. This is consistent in all models since we found a negative sign on the coefficient for the Experience square. This is consistent with the literature. This is true in the context of cross-sectional analysis and panel data analysis (except Heckman model for panel data).

We also explored the possibility of social capital in the form of Networking Intensity in explaining variations in earnings between and within. Using cross-sectional and panel data analysis, we find that social capital (*Networking Intensity*) positively affects earnings in cross-sectional models for IHDS II and in panel data models.

10.2.2. Objective 2: To Study Trends And Differences In Returns To Human Capital Investment Across Classes (Economic Distribution) (Chapter 5)

Using the Heckmann two step model for correction of selection bias, we estimated private returns across class (economic distributions) We find that Education, Experience, and social capital (Networking intensity) are significant across consumption quintiles.

The second objective was formulated to address differences in private returns to human capital investment (education), and potential job market experience across economic distributions, in India. The question of whether a child from a economically disadvantage household is likely to realize private returns to education as those better off is sought to be examined in this objective. We pursued this objective in Chapter 5 of this thesis. We used two specifications of the Heckman selection model, namely

Heckman selection model (maximum likelihood estimator) and Heckman selection model (two step consistent estimator) to estimate private returns across class (economic distributions, consumption quintiles) with cross-sectional and panel data settings. We also examined the impact of social capital measured by Networking Intensity on earnings across consumption quintiles. Moreover, we used our estimates from the Heckman selection approach with fixed effects from the panel data to construct projected earnings across the consumption quintiles. Here, we summarize these findings by nature of data; namely cross-sectional (IHDS I), cross-sectional (IHDS II), and panel data (IHDS I and IHDS II merged).

IHDS I:

The estimates for private returns to an additional year of education, in the OLS model range between the lowest for APL1 (5.3%) and highest for APL5(13%). The private returns to an additional year of experience in the OLS model range between the lowest for BPL (2%) and highest for APL5(6.2%). Estimates for Networking Intensity in the OLS model range between the lowest of APL5 (- 4.6%) and the highest for BPL(4%).

The private returns to an additional year of education in the Heckman model (MLE) range between the lowest BPL (5.2%) and highest for APL5(12.9%). The private returns to an additional year of experience in the Heckman model (MLE) range between the lowest for APL1 (2.6%) and the highest for APL5(6.4%). Estimates for networking intensity in the Heckman model (MLE) range between the lowest of APL5 (- 4.6%) and the highest for BPL(3.3%). The overall models are significant across quintiles, with a significant Chi-squared value. The model also estimates IMR.

The private returns to an additional year of education in the Heckman model (TSE) range between the lowest BPL (5 %) and highest for APL5(13%). The private returns to an additional year of experience in the Heckman model (TSE) range between the lowest for BPL and APL1 (2.6%) and the highest for APL5(6.5%). Estimates for networking intensity in the Heckman model (TSE) range between the lowest of APL5 (- 4.8%) and the highest for BPL(3.6%). The overall models are significant across quintiles, with a significant Chi-squared value. The model also estimates IMR. The coefficient for IMR was significant across all consumption quintiles, indicating the presence of selection bias.

IHDS II:

The estimates for private returns to an additional year of education, in the OLS model range between the lowest for APL1 (3.9%) and highest for APL5 (12%). The private returns to an additional year of experience in the OLS model range between the lowest for BPL (1.9%) and highest for APL5(4.5%). Estimates for Networking Intensity in the OLS model range between the lowest of APL5 (- 2.1%) and the highest for BPL(7.3%). The private returns to an additional year of education in the Heckman model (MLE) range between the lowest APL1 (3.9%) and highest for APL5(12.3%). The private returns to an additional year of experience in the Heckman model (MLE) range between the lowest for APL1 (1.9%) and the highest for APL5(4.6%). Estimates for networking intensity in the Heckman model (MLE) range between the lowest of APL5 (- 2.2%) and the highest for BPL(6.9%). The private returns to an additional year of education in the Heckman model (TSE) range between the lowest APL1 (3.9%) and highest for APL5(12%). The private returns to an additional year of experience in the Heckman model (TSE) range between the lowest for APL1 (1.9%) and the highest for APL5(4.5%). Estimates for networking intensity in the Heckman model (TSE) range between the lowest of APL5 (- 2.1%) and the highest for BPL(7%). The overall models are significant across quintiles, with a significant Chi-squared value. The model also estimates IMR.

Panel Data:

We find that the returns to an additional year of education are lowest for APL1, followed by BPL, and highest for APL5, followed by APL4. We find that the returns to an additional year of experience are lowest for BPL & APL1 and highest for APL5. We also find that an inverted U-shaped relationship exists between earnings and experience for all quintiles. Last, we find that social capital, as measured by networking intensity, bears a positive impact on BPL and APL1 and a negative impact on APL3, APL4, and APL5. We also find evidence of selection bias across quintiles.

We find that Education, Experience, and social capital (Networking intensity) are significant across consumption quintiles. We conclude that private returns to education are lower for those below the poverty line and just above the poverty line. We find a similar trend in private returns to experience. However social capital has significant positive bearing for the lower strata. Additionally, our projected earnings from panel estimates lead us to believe that earnings across consumption quintiles provide evidence of economic disparities though the private returns to education and experience are positive.

10.2.3. Objective 3:To Study Trends And Differences In Returns To Human Capital Investment Across Genders (Chapter 6)

Gender differences in earnings have been studied extensively. This objective of the thesis addresses differences in private returns to education across gender. It seeks to answer questions like, do women realize private returns to education and experience similar to men? Would going to school get a girl child similar returns to boys in India? In this context, we find education was positive and significant for both males and females.

IHDS I:

The private returns to an additional year of *Education* in the OLS model range between the lowest of males (9.29%) and females (10.2%). The private returns to an additional year of *Education* in the Heckman (MLE) model range between the lowest of females (7.8%) and males (9.19%). The private returns to an additional year of *Education* in the Heckman (TSE) model range between the lowest of females (10%) and males (9.3%).

The private returns to an additional year of *Experience* in the OLS model range between the lowest of females (3.95%) and males (5.3%). The private returns to an additional year of *Experience* in the Heckman (MLE) model range between the lowest of females (4.5%) and males (6%). The private returns to an additional year of *Experience* in the Heckman (TSE) model range between the lowest of females (4.4%) and males (5.9%).

The private returns to *Networking Intensity* in the OLS model range between the lowest of females (- 0.8%) (not significant) and males(1.65%). The estimates for *Networking Intensity* in the Heckman (MLE) model range between the lowest of females (- 1.1%) and males(1.4%). The estimates for *Networking Intensity* in the Heckman (TSE) model range between the lowest of females (- 1.2%) and males(1.5%).

IHDS II

The private returns to an additional year of *Education* in the OLS model range between the lowest of males (7.3%) and females (7.9%). The private returns to an additional year of *Education* in the Heckman (MLE) model range between the lowest of males (7.2%) and females (7.9%). The private returns to an additional year of *Education* in the Heckman (TSE) model range between the lowest of males (7.2%) and females (7.9%).

The private returns to an additional year of *Experience* in the OLS model range between the lowest of males (3.4%) and females (3.5%). The private returns to an additional year of *Experience* in the Heckman (MLE) model range between the lowest of males (3.54%) and females (3.55%). The private returns to an additional year of *Experience* in the Heckman (TSE) model are similar for males (3.6%) and females (3.6%).

The estimates for *Networking Intensity* in the OLS model range between the lowest of females (1.34%) (not significant) and males(5.22%). The estimates for *Networking Intensity* in the Heckman (MLE) model range between the lowest of females (1.22%) and males(5.11%). The estimates for *Networking Intensity* in the Heckman (TSE) model range between the lowest of females (1.2%) and males(5%).

Panel data IHDS I & II:

Education was positive and significant for both males and females. The experience was positive and significant for females and jointly significant for males. The square term (Experiencesqr) was negative. This suggests an inverted U-shaped relation between earnings and experience for females. The implication is that while women's earnings increase with experience, the rate of increase in earnings—caused by a rise in experience—occurs at a decreasing rate. Networking intensity was positive and significant for males. The IMR was significant for both groups in 2005 and for males in 2012. Returns to females (5.2%) are much higher than the returns to males (3.7%). This we believe is evidence on possible long-run convergence in earnings.

10.2.4. Objective 4: To Study Trends And Differences In Returns To Human Capital Investment Across Castes (Chapter 7)

Next, we, investigated the private returns to education and experience across castes in India. This objective of the thesis is concerned with evaluating the magnitude of private returns across the caste categories. This we believed would provide meaning insight into the earnings trajectory of these social groups, given social hierarchies in India. We look into questions like; do individuals from marginalized castes realise similar private returns to education and experience in India? Does building social capital help the marginalized? In this context, we found that education has a positive and significant impact on all castes groups.

IHDS I :

The private returns to an additional year of *Education* in the OLS model range between the lowest for OBC (8.5%) and highest for ST (11.5%). The private returns to an additional year of *Education* in the Heckman model (MLE) range between the lowest for OBC (7.6%) and highest for ST (11.4%). The private returns to an additional year of *Education* in the Heckman model (TSE) range between the lowest for OBC (8.4%) and highest for ST (11.4%).

The private returns to an additional year of *Experience* in the OLS model range between the lowest for OBC (4.25%) and highest for ST (5.45%). The private returns to an additional year of *Experience* in the Heckman model (MLE) range between the lowest of OBC (4.6%) and highest for ST (5.7%). The private returns to an additional year of *Experience* in the Heckman model (TSE) range between the lowest of General and forward (4.7%) and highest for ST (5.7%).

The estimate for *Networking Intensity* in the OLS model range between the lowest for ST (-5.5%) and the highest for SC (4.7%). The estimate for *Networking Intensity* in the Heckman model (MLE) range between the lowest for ST (- 5.7%) and highest for SC (3.2%). The estimate for *Networking Intensity* in the Heckman model (TSE) range between the lowest for ST (- 5.8%) and highest for SC (4.6%).

IHDS II :

The private returns to an additional year of *Education* in the OLS model range between the lowest for OBC (6.4%) and highest for ST (8.99%). The private returns to an additional year of *Education* in the Heckman model (MLE) range between the lowest of OBC (6.3%) and highest for ST (8.9 %). The private returns to an additional year of *Education* in the Heckman model (TSE) range between the lowest of OBC (6.38%) and highest for ST (8.9 %).

The private returns to an additional year of *Experience* in the OLS model range between the lowest for OBC (2.35%) and the highest for SC (3.95%). The private returns to an additional year of *Experience* in the Heckman model (MLE) range between the lowest of OBC (2.63%) and highest for SC (4.2%). The private returns to an additional year of *Experience* in the Heckman model (TSE) range between the lowest of OBC (2.7%) and highest for SC (4.2%).

The estimate for *Networking Intensity* in the OLS model range between the lowest of ST (2.85%) and the highest for General (3.82%). The estimate for *Networking Intensity* in the Heckman model (MLE) range between the lowest for ST (2.8%) and highest for OBC (3.8%). The estimate for *Networking Intensity* in the Heckman model (TSE) range between the lowest for ST (2.8%) and highest for OBC (3.8%).

Panel data IHDS I & II:

We found that *Education* has a positive and significant impact on all castes. The *Experience* was positive and significant for general castes, and SC. *Experience*sqr was negative and significant for general. This suggests a U-shaped relation between earnings and *Experience* for (general). The IMR was significant for most of the values, indicating the presence of a selection bias.

Education, and Experience have a positive and significant impact on earnings for all castes. Panel data results show that SCs, realise highest private returns to education followed by General. Estimates fom cross-sectional data and panel data lead us to conclude that affirmative action has helped to some extentent in reducing social disparities and generating opportunities. However, the results are not uniform across social groups. Our earnings projection using panel data estimates lead us to conclude that while some social groups may have benefitted (like SCs) , others like OBCs and STs seem to need futher targeted intervention. Education (SDG 4), then becomes crucial in context to breaking social barriers and reduce inequalities (SDG10).

10.2.5. Objective 5: To Study Trends And Differences In Returns To Human Capital Investment Across Religions (Chapter 8)

Another major social groups in India are religious affiliations. In this objective we examine the magnitude of private returns to education across religious affiliations. We looked at how these private returns might differ across religions. We addressed questions like, do individual from minority communities find similar private returns to education and experience in India? In this context; we found that education has a positive and significant impact on earnings across religions, but vary in magnitude. The experience was positive and significant for all religions. Networking Intensity was not significant for Other minorities in any form of analysis.

IHDS I :

The private returns to an additional year of Education in the OLS model range between the lowest for Muslims (8.07%) and highest for Other Minorities (11.%).The private returns to an additional year of Education in the Heckman model (MLE) range between the lowest for Muslims (7.78%) and highest for Other Minorities (11.6%). The private returns to an additional year of Education in the Heckman model (TSE) range between the lowest for Muslims (7.76%) and highest for Other Minorities (11.6%).

The private returns to an additional year of Experience in the OLS model range between the lowest for Other minorities (3.64%) and highest for Hindus (4.93%).The private returns to an additional year of Experience in the Heckman model (MLE) range between the lowest for Other minorities (4%) and highest for Hindu (5.36%). The private returns to an additional year of Experience in the Heckman model (TSE) range between the lowest for Other minorities (4%) and highest for Hindu (5.14%).

The Networking Intensity in all of the 3 models, OLS, Heckman (MLE) and Heckman (TSE) showed a negative significant impact for Other minorities.

IHDS II :

The private returns to an additional year of Education in the OLS model range between the lowest for Muslims (6.54%) and highest for Other Minorities (9.77.%).The private returns to an additional year of Education in the Heckman model (MLE) range between the lowest for Muslims (6.51%) and highest for Other Minorities (10.4%). The private returns to an additional year of Education in the Heckman model (TSE) range between the lowest for Muslims (6.48%) and highest for Other Minorities (9.87%).

The private returns to an additional year of Experience in the OLS model range between the lowest for Hindu (3.02%) and highest for Muslim (4.69%).The private returns to an additional year of Experience in the Heckman model (MLE) range between the lowest for Hindu (3.24%) and highest for Muslim (4.69%). The private returns to an additional year of Experience in the Heckman model (TSE) range between the lowest for Hindu (3.3%) and highest for Other Minorities (4.7%).

The Networking Intensity in all of the 3 models, OLS, Heckman(MLE) and Heckman(TSE) showed a positive significant impact for Hindu and Muslim.

Panel data IHDS I & II

Education was significant and positive for all three religious' groups. Experience, on the other hand, was significant and positive for Hindus and Other minorities but not for Muslims. Networking intensity was significant and positive for Hindus only.

10.2.6. Objective 6: To Estimate The Impact Of Natural Disasters (With Differential Intensity) On Educational Outcomes In Rural Villages In India (Chapter 9)

We provide first estimates of the impact of natural disasters on educational attainment disaggregated by social (including caste, religion, and gender) and economic groups (consumption quintiles).

Panel data : We find that *Natural Disaster Intensity* (NDI) yields a negative and significant Average Treatment Effect on the Treated (exposed to natural disasters) on *Educational Attainment*. Our findings from disaggregated estimates add to previous studies by illustrating the severity of these losses to educational attainment for marginalized sections in rural India.

Our results are consistent for Females, three consumption quintiles (BPL, APL4, and APL5), two marginalized caste categories (OBC and SC), and religious minorities (Muslims and Other minorities).

We also find that the number of children in the household negatively and significantly impacts educational attainment for both genders, consumption quintiles (BPL, APL1, and APL5), social groups (OBCs), and all religions.

The *Number of household assets* has a positive and significant impact on educational attainment for both genders, consumption quintiles (BPL and APL1), social groups

(SCs), and religion (Hindus and Muslims). This aligns with extensive literature in this area on the links between household assets (financial and non-financial) and educational attainment (Kim & Sherraden, 2011).

We find that *school fees have a positive and significant impact on educational attainment for consumption quintiles (APL1) and social groups (SCs)* and a negative and significant impact on religion (other minorities). However, we did not test for changes in enrolment.

Apart from school fees, we also looked at *school type* since it captures additional indicators beyond school fees, including the learning atmosphere. In *school type,* our reference category was other schools. We find that *Government* and *Aided* schools had a positive and significant impact on educational attainment across gender, consumption quintiles (except APL2, -ve), caste (SCs, but negative for STs), and religion (Hindus). On the other hand, *Private* schools negatively impacted educational attainment for both genders, APL2, Hindus, and Muslims.

Membership Intensity negatively and significantly affects STs. Confidence Intensity positively and significantly affects women, BPL, and Hindus, whereas it negatively affects STs. Financial resilience in the form of Health insurance positively and significantly affects APL3 and other minorities, whereas it negatively affects APL1 and APL4, OBCs, and STs. Financial resilience in the form of Life Insurance positively and significantly affects men, APL4, whereas it negatively affects SCs. Repeating in a class negatively and significantly affects both genders, BPL and APL4, general and OBCs, and Hindus and Other minorities.

10.3. POLICY IMPLICATIONS

Our results demonstrate that education is significant and positive for all categories of individuals—economic strata, gender, caste, and religion. However, there is a wide range of returns from education to each of these subgroups.

The received literature on gender differentials in earnings has universally found that women earn less than men for a given level of education and experience. We did not find evidence contrary to this. However, we did find that there is a tendency towards convergence in earnings between males and females for given levels of education and experience. This has been noted by other authors as well, and it can be attributed in part to structural and technological change (Cortes et al., 2020). As the workforce has moved from manufacturing to service-oriented employment, women are less likely to be disadvantaged in such work situations. Moreover, their soft skills at the workplace may be more sought after than those of men. Our findings provide evidence of convergence in the long run in the Indian context for gender We found that the initial value of earnings (intercept value) for males is much higher than that of females. However, the marginal returns to an additional year of education are much higher for women than men.

Among the caste groups, the SCs show the highest returns to education (5.9 %), followed by the general castes (4.6%). However, general castes have a higher intercept, which elevates their earnings above other castes for almost the entire range of years. Among the religious groups Hindus (3.8%) and Muslims (2.3%) have a lower return than other minorities (9.9%). These results suggest that investment in education does pay dividends for social development in India, in alignment with the literature (Tilak, 2021).

There is, therefore, a need for a targeted approach to ensure that these groups do not fall out of the education system for lack of ability to pay (Mishra & Ramakrishna, 2023). The Right of Children to Free and Compulsory Education (RTE) Act of 2009 makes it obligatory for the government to ensure that every child in the age group of 6 to 14 years is provided free and compulsory education (GoI, 2009). This is a step towards inclusive development (Das, 2020), although there is evidence of unintended outcomes (Chatterjee et al., 2020). Educational affirmative interventions based on caste and gender have been found to have positive impacts in India (Bagde et al., 2016). Our findings are along the lines of Chakraborty and Bohara's (2021) study on the role of caste and religion in India and their recommendation for policy intervention for socially backward classes and Muslims to establish equity in the labor market.

There is a constitutionally mandated reservation of jobs in the public sector for less privileged castes (Bhaskar, 2021). This has helped overcome selection bias against the lower castes in the public sector to some extent. However, reservations for women have been limited to the lowest tier of government despite the longstanding proposal to bring in 33% reservations at all tiers of government and in other areas of employment (Marwah, 2019). The presence of women in government has had positive outcomes in local assets and employment creation (Deininger et al., 2020).

Our findings have important policy implications and connect to multiple SDGs, such as poverty (SDG 1), quality education (SDG 4), gender equality (SDG 5), and reduced inequalities (SDG 10). The government of India has been committed to the fulfillment of the SDGs and has undertaken multiple social interventions. In the sphere of education, as well, policies have been developed to enhance access and improve quality education. About a decade and a half ago, the Right of Children to Free and Compulsory

Education Act, or Right to Education Act, 2009, for free and compulsory schooling (GoI, 2009), was enacted to reduce the dropout rate in schools and ensure compulsory education up to 14 years. The New Education Policy announced in 2020 also attempts to upgrade quality education. Our findings provide evidence that education (SDG 4) helps in achieving the goals of gender equality (SDG 5) and social justice for marginalized (caste and religious) groups (SDG 10).

The government has also implemented affirmative action policies in the domain of jobs and education for certain groups based largely on caste (Deshpande, 2013) and, more recently, economic backwardness (Nath, 2019). We find that despite education positively impacting earnings, the less privileged groups would require affirmative action to achieve reduced inequalities (SGD 10) and escape poverty (SGD 1).

Our study addresses the issues of economic distribution, gender, caste, and religion in the context of larger sustainable developmental goals (Dhar, 2018; Pandey, 2019). Our findings of higher marginal returns to education among females justify the need for greater investment in education for women. The lower returns on education among STs and Muslims indicate the need for affirmative action for these groups. Similarly, the positive and significant coefficient for all consumption quintiles suggests that anti-poverty programs, in combination with educational opportunities for the less privileged (BPL), would meet the goals of social justice. Development policies that combine multiple objectives would, therefore, have a more effective welfare outcome.

Our study also looked at the role of social capital by examining the impact of networking intensity on different groups. We find that this variable had a positive impact on the earnings of males (among gender groups). Our results are in line with the findings of a study based in China (Liu, 2017), where individual-level social capital was found to have a greater impact on men than women. Similar effects were reported (Smith, 2000) in the context of the United States. This could be anticipated in India and might explain the wage gaps for women.

In context to the Impact of Natural disasters on educational attainment, our findings reaffirm that climate change (SDG13) vis a vis natural disasters have a significant negative effect on educational attainment. We additionally conclude that the impact of natural disasters on educational attainment for marginalized sections (women, BPL, OBCs, SCs, minority religions) of the society is much more severe. On the one hand, lower educational attainment (SDG4) in disaster-prone areas could trigger a chain of disruptive welfare effects in the form of lower overall earnings, widening of gender differences (in attainment and earnings) (SDG5), and higher societal inequalities (SDG10). On the other hand, these lower educational attainments reduce the effectiveness and increase the cost of disaster management programs developed for mitigation from natural disasters and to reduce vulnerabilities. Studies have shown that higher education levels lead to better awareness levels and ease of dissemination of information and serve as a measure of development with direct and indirect effects on reducing fatality rates (Parida et al., 2021). The higher vulnerability combined with the increasing frequency of natural disasters in recent times and the inability to migrate leaves the marginalized in danger of intergenerational disruptions in well-being. This possibly creates a spiral of lower education, lower earnings, and lower welfare, inevitably pushing them to poverty. The impacts of climate change on human capital demand effective policy intervention to mitigate human capital losses to marginalized sections. The challenge then is to create a targeted response to more vulnerable sections of the population by 1) minimizing dropout rates disasters, 2) strengthening schooling mechanisms to resume educational programs rapidly, and 3) reducing the probability of household's decisions to switch from investment in human capital (education) to rebuilding damages to physical assets by providing definitive monetary assistance in rehabilitation.

10.4. CONTRIBUTION OF THIS THESIS

We re-examine the issue of private returns to education earnings in India and the impact of natural disasters on educational attainment in India. We use data from two rounds of IHDS to construct a panel data set. To our understanding, ours is the first attempt using a panel data model with individual-level fixed effects. We follow the Heckman selection model and use a new and improved command (xtheckmanfe) in Stata16.1 to control for selection bias. The thesis adds to the literature on the methodology applied, robust estimation, and coverage of interrelated topics between earnings, education, and natural disasters. This study contributes in multiple ways:

1. Firstly, we add to the literature by providing panel estimates using a Heckman-type selection model to the private returns to an additional year of education and experience.
2. Secondly, we provide first estimates on the private returns to an additional year of education and experience disaggregated by consumption quintiles, caste, gender, and religion using a nationally representative dataset (Hussain & Mukhopadhyay, 2023). Our analysis provides greater disaggregation, which can help devise more effective and targeted policies for affirmative action.
3. Previously, studies in India have employed a cross-section data set (Agrawal, 2014; Bairagya, 2020; Geetha Rani, 2014; Kingdon & Theopold, 2008; Mitra, 2019) or a quasi-panel (Bhattacharya & Sato, 2017; Mendiratta & Gupt, 2013) to study earnings in India. Our panel data approach with individual fixed effects improves upon estimates of all earlier studies on this subject.

4. Thirdly, the thesis explores the impact of social capital in the form of social networking, from membership of households to various economic and social associations.
5. Lastly, we account for the impact on education based on the frequency of natural disasters in a large countries like India. We estimate the impact of natural disasters (with differential intensity) on educational outcomes in rural villages. Our study bridges the research gap by incorporating the severity of natural disasters in a longitudinal dataset covering all states and union territories in India. We provide estimates disaggregated by Gender (Male and Female), Consumption quintiles (Below the poverty line (BPL) and quintiles above the poverty line, from APL1 to APL5), Caste (General, Other backward castes, Scheduled castes and Scheduled tribes) and Religion (Hindu, Other minorities and Muslim).

10.5. LIMITATIONS

While several innovations have been implemented in our study, we acknowledge some limitations.

1. We were unable to separate the impact of ability and inherent skills from the years of education in wage determination (ability bias).
2. We did not account for levels of schooling.
3. The reporting of caste was non-uniform in the two rounds of IHDS. This could be due to the re-classification of castes, which was an ongoing process at the time of the survey. This could be further resolved when data from new ongoing survey rounds becomes available.
4. Our study did not stratify the labor market by state and sector. Incorporating these classifications could yield new insights in a country like India, with large geographic and economic heterogeneity.

REFERENCES

1) Aakvik, A., Salvanes, K. G., & Vaage, K. (2010). Measuring heterogeneity in the returns to education using an education reform. European Economic Review, 54(4), 483–500.

2) Abdullah, A., Doucouliagos, H., & Manning, E. (2015). Does Education Reduce Income Inequality? A Meta-Regression Analysis. Journal of Economic Surveys, 29(2), 301–316. https://doi.org/10.1111/joes.12056

3) Abraham, K. G., & Mallatt, J. (2022). Measuring Human Capital. Journal of Economic Perspectives, 36(3), 103–130. https://doi.org/10.1257/jep.36.3.103

4) Abraham, R., & Shrivastava, A. (2022). How comparable are India's labour market surveys? The Indian Journal of Labour Economics, 65(2), 321–346. doi: https://doi.org/10.1007/s41027-022-00381-x

5) Acemoglu, D., & Autor, D. (2012). What does human capital do? A review of Goldin and Katz's: The race between education and technology. Journal of Economic Literature, 50(2), 426–463. doi: https://doi.org/10.1257/jel.50.2.426

6) Agrawal, T. (2012). Returns to Education in India: Some Recent Evidence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1989460

7) Agrawal, T. (2014). Gender and caste-based wage discrimination in India: Some recent evidence. Journal for Labour Market Research, 47(4), 329–340. doi: https://doi.org/10.1007/s12651-013-0152-z

8) Agte, P., & Bernhardt, A. (2023). The Economics of Caste Norms: Purity, Status, and Women's Work in India. Job Market Paper.

9) Ahmed, T., & Chattopadhyay, R. (2016). Return to general education and vocational education&training in Indian context: Policy implications.

International Journal of Educational Management, 30(3), 370–385. https://doi.org/10.1108/IJEM-10-2014-0135

10) Akter, S., & Chindarkar, N. (2020). An empirical examination of sustainability of women's empowerment using panel data from India. The Journal of Development Studies, 56(5), 890–906. doi: https://doi.org/10.1080/00220388.2019.1605054

11) Alba-Ramirez, A., & San Segundo, M. J. (1995). The returns to education in Spain. Economics of Education Review, 14(2), 155–166.

12) Andrews, K., Bosch, L., Teixeira, J., Meidina, I., & Nagpal, S. (2023). Seizing Opportunities of a Lifetime: The Timor-Leste Human Capital Review. https://openknowledge.worldbank.org/handle/10986/40623

13) Anees, M. M., Shukla, R., Punia, M., & Joshi, P. K. (2020). Assessment and visualisation of inherent vulnerability of urban population in India to natural disasters. Climate and Development, 12(6), 532–546. https://doi.org/10.1080/17565529.2019.1646629

14) Angrist, J. D., & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics, 106(4), 979–1014. doi: https://doi.org/10.2307/293795

15) Angrist, J. D., & Krueger, A. B. (1999). Empirical strategies in labor economics. In Handbook of labor economics (Vol. 3, pp. 1277–1366). Elsevier. https://www.sciencedirect.com/science/article/pii/S1573446399030047

16) Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: from supply and demand to natural experiments. Journal of Economic Perspectives, 15(4), 69–85. doi: https://doi.org/10.1257/jep.15.4.69

17) Angrist, J., & Krueger, A. B. (1992). Estimating the payoff to schooling using the Vietnam-era draft lottery. National bureau of economic research Cambridge, Mass., USA.

18) Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. Nature, 592, 403–408. https://doi.org/10.1038/s41586-021-03323-7

19) Arabsheibani, G., Gupta, P., Mishra, T., & Parhi, M. (2018). Wage differential between caste groups: Are younger and older cohorts different? Economic Modelling, 74, 10–23. doi: https://doi.org/10.1016/j.econmod.2018.04.019

20) Asadullah, M. N., & Xiao, S. (2020). The changing pattern of wage returns to education in post-reform china. Structural Change and Economic Dynamics, 53, 137–148.

21) Ashraf, Q. H., Weil, D. N., & Wilde, J. (2013). The effect of fertility reduction on economic growth. Population and Development Review, 39(1), 97–130. doi: https://doi.org/10.1111/j.1728-4457.2013.00575.x

22) Azam, M. (2018). Does social health insurance reduce financial burden? Panel data evidence from India. World Development, 102, 1–17. doi: https://doi.org/10.1016/j.worlddev.2017.09.007

23) Azam, M., Chin, A., & Prakash, N. (2013). The returns to English-language skills in India. Economic Development and Cultural Change, 61(2), 335–367. doi: https://doi.org/10.1086/668277

24) Baez, J. E., Kronick, D., & Mason, A. D. (2013). Rural Households in a Changing Climate. The World Bank Research Observer, 28(2), 267–289. https://doi.org/10.1093/wbro/lks008

25) Baez, J., de la Fuente, A., & Santos, I. V. (2010). Do Natural Disasters Affect Human Capital? An Assessment Based on Existing Empirical Evidence (Discussion Paper IZA Discussion Papers, No. 5164). IZA–Institute of Labor Economics.

26) Bagde, S., Epple, D., & Taylor, L. (2016). Does affirmative action work? Caste, gender, college quality, and academic success in India. The American Economic Review, 106(6), 1495–1521. doi: https://www.jstor.org/stable/43861129

27) Bairagya, I. (2020). Returns to education in self-employment in India: A comparison across different selection models. WIDER Working Paper 2020/5. Helsinki: UNU-WIDER. doi: https://doi.org/10.35188/UNU-WIDER/2020/762-0

28) Baltagi, B. H., & Khanti-Akom, S. (1990). On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. Journal of Applied Econometrics, 5(4), 401–406. doi: https://doi.org/10.1002/jae.3950050408

29) Barrett, A., Callan, T., & Nolan, B. (1999). Rising Wage Inequality, Returns to Education and Labour Market Institutions: Evidence from Ireland. British Journal of Industrial Relations, 37(1), 77–100. https://doi.org/10.1111/1467-8543.00119

30) Barro, R. J. (1999). Human capital and growth in cross-country regressions. Swedish Economic Policy Review, 6(2), 237–277.

31) Barro, R. J., & Lee, J. W. (1996). International Measures of Schooling Years and Schooling Quality. The American Economic Review, 86(2), 218–223.

32) Barro, R. J., & Lee, J.-W. (1993). International comparisons of educational attainment. Journal of Monetary Economics, 32(3), 363–394.

33) Barro, R. J., & McCleary, R. M. (2003). Religion and economic growth across countries. American Sociological Review, 68(5), 760–781. doi: https://doi.org/10.2307/1519761

34) Bartik, T., & Hershbein, B. (2018). Degrees of Poverty: The Relationship between Family Income Background and the Returns to Education (SSRN Scholarly Paper 3141213). https://doi.org/10.2139/ssrn.3141213

35) Beam, E. A., Hyman, J., & Theoharides, C. (2020). The relative returns to education, experience, and attractiveness for young workers. Economic Development and Cultural Change, 68(2), 391–428. doi: https://doi.org/10.1086/701232

36) Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of Political Economy, 70(5, Part 2), 9–49. doi: https://doi.org/10.1086/258724

37) Becker, G. S. (1975). Investment in Human Capital: Effects on Earnings. In Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition (pp. 13–44). NBER. https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-second-edition/investment-human-capital-effects-earnings

38) Becker, G. S. (1993). Human capital: A theoretical and empirical analysis, with special reference to education (3rd ed). The University of Chicago Press.

39) Becker, G. S. (1994). Human capital: A theoretical and empirical analysis, with special reference to education (3rd ed.). Chicago: University of Chicago Press.

40) Becker, G. S., Murphy, K. M., & Tamura, R. (1990). Human Capital, Fertility, and Economic Growth. Journal of Political Economy, 98(5, Part 2), S12–S37. https://doi.org/10.1086/261723

41) Bedi, A. S., & Edwards, J. H. Y. (2002). The impact of school quality on earnings and educational returns—Evidence from a low-income country. Journal of Development Economics, 68(1), 157–185. https://doi.org/10.1016/S0304-3878(02)00010-X

42) Behrman, J. R., & Rosenzweig, M. R. (1999). "Ability" biases in schooling returns and twins: A test and new estimates. Economics of Education Review, 18(2), 159–167.

43) Behtoui, A., & Neergaard, A. (2010). Social capital and wage disadvantages among immigrant workers. Work, Employment and Society, 24(4), 761–779. doi: https://doi.org/10.1177/0950017010380640

44) Belzil, C., & Hansen, J. (2002). Unobserved ability and the return to schooling. Econometrica, 70(5), 2075–2091.

45) Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. Journal of Political Economy, 75(4, Part 1), 352–365. https://doi.org/10.1086/259291

46) Benson, C., & Clay, E. (2004). Understanding the Economic and Financial Impacts of Natural Disasters. The World Bank. https://doi.org/10.1596/0-8213-5685-2

47) Bhambhani, C., & Inbanathan, A. (2018). Not a mother, yet a woman: Exploring experiences of women opting out of motherhood in India. Asian Journal of Women's Studies, 24(2), 159–182. doi: https://doi.org/10.1080/12259276.2018.1462932

48) Bhandari, L., & Bordoloi, M. (2006). Income differentials and returns to education. Economic and Political Weekly, 3893–3900.

49) Bhaskar, A. (2021). Reservations, efficiency, and the making of Indian constitution. Economic and Political Weekly, 56(19), 42-49. https://www.epw.in/journal/2021/19/special-articles/reservations-efficiency-and-making-indian.html

50) Bhattacharya, P., & Sato, T. (2017). Estimating regional returns to education in India: A fresh look with pseudo-panel data. Progress in Development Studies, 17(4), 282–290. doi: https://doi.org/10.1177/1464993417716357

51) Bhaumik, S. K., & Chakrabarty, M. (2006). Earnings Inequality in India: Has the Rise of Caste and Religion based Politics in India had an impact?

52) Bhaumik, S. K., & Chakrabarty, M. (2008). Does move to market have an impact on earnings gap across gender? Some evidence from India. Applied Economics Letters, 15(8), 601–605. doi: https://doi.org/10.1080/13504850600706651

53) Bhaumik, S. K., & Chakrabarty, M. (2010). Earnings inequality in India: Has the rise of caste and religion-based politics in India had an impact? In Rakesh Basant & Abusaleh Shariff (Ed.). Handbook of Muslims in India Empirical and Policy Perspectives (pp. 235–253), New Delhi: Oxford University Press.

54) Bietenbeck, J., Marcus, J., & Weinhardt, F. (2020). Tuition Fees and Educational Attainment (Discussion Paper IZA Discussion Papers, No. 13709;

p. 66). Institute of Labor Economics (IZA). https://www.econstor.eu/bitstream/10419/227236/1/dp13709.pdf

55) Black, S. E., & Lynch, L. M. (1996). Human-capital investments and productivity. The American Economic Review, 86(2), 263–267.

56) Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). Why the Apple Doesn’t Fall Far: Understanding Intergenerational Transmission of Human Capital. American Economic Review, 95(1), 437–449. https://doi.org/10.1257/0002828053828635

57) Blake, J. (1989). Number of Siblings and Educational Attainment. Science, 245(4913), 32–36. https://doi.org/10.1126/science.2740913

58) Blanden, J. (2004). Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain. Oxford Review of Economic Policy, 20(2), 245–263. https://doi.org/10.1093/oxrep/grh014

59) Blanden, J. (2020). Chapter 10—Education and inequality. In S. Bradley & C. Green (Eds.), The Economics of Education (Second Edition) (pp. 119–131). Academic Press. https://doi.org/10.1016/B978-0-12-815391-8.00010-0

60) Blau, F. D., Gielen, A. C., & Zimmermann, K. F. (2012). Gender, inequality, and wages (1st ed). Oxford: Oxford University Press. doi: https://doi.org/10.1093/acprof:oso/9780199665853.001.0001

61) Blundell, R., Dearden, L., & Sianesi, B. (2001). Estimating the Returns to Education: Models, Methods and Results.

62) BMTPC. (2006). Vulnerability Atlas of India. Building Materials and Technology Promotion Council, Ministry of Housing and Urban Poverty Alleviation, Government of India.

https://bmtpc.org/admin/PublisherAttachement/An%20Introduction%20to%20the%20Vulnerability%20Atlas%20of%20IndiaAtt.pdf

63) Borooah, V. K. (2012). Social Identity and Educational Attainment: The Role of Caste and Religion in Explaining Differences between Children in India. The Journal of Development Studies, 48(7), 887–903. https://doi.org/10.1080/00220388.2011.621945

64) Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association, 90(430), 443–450. https://doi.org/10.1080/01621459.1995.10476536

65) Bourque, L. B., Siegel, J. M., Kano, M., & Wood, M. M. (2007). Morbidity and Mortality Associated with Disasters. In H. Rodríguez, E. L. Quarantelli, & R. R. Dynes, Handbook of Disaster Research (pp. 97–112). Springer, New York. https://doi.org/10.1007/978-0-387-32353-4_6

66) Briggs, D. C. (2004). Causal inference and the Heckman model. Journal of Educational and Behavioral Statistics, 29(4), 397–420. doi: https://doi.org/10.3102/10769986029004397

67) Butz, W. P., Lutz, W., & Sendzimir, J. (2014). Education and differential vulnerability to natural disasters. Ecology and Society.

68) Byron, R. P., & Manaloto, E. Q. (1990). Returns to education in China. Economic Development and Cultural Change, 38(4), 783–796.

69) Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. National Bureau of Economic Research Cambridge, Mass., USA.

70) Card, D. (1999). The Causal Effect of Education on Earnings. In Handbook of Labor Economics (Vol. 3, pp. 1801–1863). Elsevier. https://doi.org/10.1016/S1573-4463(99)03011-4

71) Card, D., & Krueger, A. B. (1992a). Does school quality matter? Returns to education and the characteristics of public schools in the United States. Journal of Political Economy, 100(1), 1–40.

72) Carneiro, P., Heckman, J. J., & Vytlacil, E. (2011). Estimating marginal returns to education. American Economic Review, 101 (6): 2754-81.doi: https://doi.org/10.1257/aer.101.6.2754

73) Caruso, G. D. (2017). The legacy of natural disasters: The intergenerational impact of 100 years of disasters in Latin America. Journal of Development Economics, 127, 209–233. https://doi.org/10.1016/j.jdeveco.2017.03.007

74) Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic Natural Disasters and Economic Growth. The Review of Economics and Statistics, 95(5), 1549–1561. https://doi.org/10.1162/REST_a_00413

75) Cawley, J., Heckman, J. J., & Vytlacil, E. J. (1998). Understanding the role of cognitive ability in accounting for the recent rise in the economic return to education. National Bureau of Economic Research Cambridge, Mass., USA. https://www.nber.org/papers/w6388

76) Chakraborty, S., & Bohara, A. K. (2021). The cost of being ‘backward’ in India: Socio-religious discrimination in the labour market. Indian Journal of Human Development, 15(2), 252–274. doi: https://doi.org/10.1177/09737030211029634

77) Chakraborty, S., Bohara, A. K., Binder, M., & Santos, R. (2020). Is the Indian labor market biased against women? Yes! In B. Jivetti & Md. N. Hoque

(Eds.), Population Change and Public Policy (pp. 229–242). Cham: Springer International Publishing. doi: https://doi.org/10.1007/978-3-030-57069-9_11

78) Chamarbagwala, R. (2008). Regional returns to education, child labour and schooling in India. The Journal of Development Studies, 44(2), 233–257.

79) Chatterjee, C., Hanushek, E. A., & Mahendiran, S. (2020). Can greater access to education be inequitable? New evidence from India's Right to Education Act. Working Paper No. 27377. Cambridge: National Bureau of Economic Research. doi: https://doi.org/10.3386/w27377

80) Chatterjee, E., & Sennott, C. (2021). Fertility intentions and child health in India: Women's use of health services, breastfeeding, and official birth documentation following an unwanted birth. PLOS ONE, 16(11), e0259311. doi: https://doi.org/10.1371/journal.pone.0259311

81) Chen, Y., Jiang, S., & Zhou, L.-A. (2020). Estimating returns to education in urban China: Evidence from a natural experiment in schooling reform. Journal of Comparative Economics, 48(1), 218–233.

82) Chen, Y., Jiang, S., & Zhou, L.-A. (2020). Estimating returns to education in urban China: Evidence from a natural experiment in schooling reform. Journal of Comparative Economics, 48(1), 218–233. https://doi.org/10.1016/j.jce.2019.09.004

83) Chong, S. T., Koh, D., Nazri, Y. F. M., Ibrahim, F., & Rahim, S. A. (2020). Social capital and well-being among marginalized young Malaysians. Kasetsart Journal of Social Sciences, 41(1), 176–180.

84) Chun, H., & Oh, J. (2002). An instrumental variable estimate of the effect of fertility on the labour force participation of married women. Applied

Economics Letters, 9(10), 631–634. doi:
https://doi.org/10.1080/13504850110117850

85) Coady, D., & Dizioli, A. (2018). Income inequality and education revisited: Persistence, endogeneity and heterogeneity. Applied Economics, 50(25), 2747–2761.

86) Coelli, M., & Tabasso, D. (2019). Where are the returns to lifelong learning? Empirical Economics, 57, 205–237.

87) Cohn, E., Kiker, B. F., & De Oliveira, M. M. (1987). Further evidence on the screening hypothesis. Economics Letters, 25(3), 289–294. https://doi.org/10.1016/0165-1765(87)90230-8

88) Collin, M., & Weil, D. N. (2020). The effect of increasing human capital investment on economic growth and poverty: A simulation exercise. Journal of Human Capital, 14(1), 43–83. doi: https://doi.org/10.1086/708195

89) Comola, M., & de Mello, L. (2013). Salaried employment and earnings in Indonesia: New evidence on the selection bias. Applied Economics, 45(19), 2808–2816. doi: https://doi.org/10.1080/00036846.2012.667551

90) Cooper, K., & Stewart, K. (2021). Does Household Income Affect children's Outcomes? A Systematic Review of the Evidence. Child Indicators Research, 14(3), 981–1005. https://doi.org/10.1007/s12187-020-09782-0

91) Cortes, G. M., Oliveira, A., & Salomons, A. (2020). Do technological advances reduce the gender wage gap? Oxford Review of Economic Policy, 36(4), 903–924. doi: https://doi.org/10.1093/oxrep/graa051

92) Cruz, L. M., & Moreira, M. J. (2005). On the validity of econometric techniques with weak instruments inference on returns to education using

compulsory school attendance laws. Journal of Human Resources, 40(2), 393–410. doi: https://doi.org/10.3368/jhr.XL.2.393

93) Cuaresma, J. C. (2010). Natural Disasters and Human Capital Accumulation. The World Bank Economic Review, 24(2), 280–302. https://doi.org/10.1093/wber/lhq008

94) Cuaresma, J. C., Havettová, M., & Lábaj, M. (2013). Income convergence prospects in Europe: Assessing the role of human capital dynamics. Economic Systems, 37(4), 493–507.

95) Das, A. (2020). Understanding equity through Section 12(1)(c) of the Right to Education Act in India. Cambridge Educational Research e-Journal, 7, 53–69. doi: https://doi.org/10.17863/CAM.58327

96) Dasgupta, P. (2005). Economics of social capital. Economic Record, 81(s1), S2–S21. doi: https://doi.org/10.1111/j.1475-4932.2005.00245.x

97) Dasgupta, S., Wheeler, D., Bandyopadhyay, S., Ghosh, S., & Roy, U. (2022). Coastal dilemma: Climate change, public assistance and population displacement. World Development, 150, 105707. https://doi.org/10.1016/j.worlddev.2021.105707

98) de Baldini Rocha, M. S., & Ponczek, V. (2011). The effects of adult literacy on earnings and employment. Economics of Education Review, 30(4), 755–764. https://doi.org/10.1016/j.econedurev.2011.03.005

99) De la Fuente, A., & Doménech, R. (2006). Human capital in growth regressions: How much difference does data quality make? Journal of the European Economic Association, 4(1), 1–36.

100) De Vreyer, P., Guilbert, N., & Mesple-Somps, S. (2015). Impact of Natural Disasters on Education Outcomes: Evidence from the 1987-89 Locust

Plague in Mali. Journal of African Economies, 24(1), 57–100. https://doi.org/10.1093/jae/eju018

101) Deaton, A. (1985). Panel data from time series of cross-sections. Journal of Econometrics, 30(1–2), 109–126. doi: https://doi.org/10.1016/0304-4076(85)90134-4

102) Deininger, K., Nagarajan, H. K., & Singh, S. K. (2020). Women's political leadership and economic empowerment: Evidence from public works in India. Journal of Comparative Economics, 48(2), 277–291. doi: https://doi.org/10.1016/j.jce.2019.12.003

103) Deming, D. J. (2022). Four facts about human capital. Journal of Economic Perspectives, 36(3), 75–102.

104) Denison, E. F. (1962). Education, Economic Growth, and Gaps in Information. Journal of Political Economy, 70(5, Part 2), 124–128. https://doi.org/10.1086/258729

105) Desai, S., & Vanneman, R. (2008). India Human Development Survey (IHDS), 2005: Version 12 (Version v12) [Data set]. ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR22626.V12

106) i, S., & Vanneman, R. (2015). India Human Development Survey-II (IHDS-II), 2011-12: Version 6 [Data set]. Inter-University Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR36151.V6

107) Desai, S., Vanneman, R. & National Council of Applied Economic Research (2005). India Human Development Survey (IHDS), 2005.Inter-University Consortium for Political and Social Research. doi: https://doi.org/10.3886/ICPSR22626.v12

108) Desai, S., Vanneman, R. & National Council of Applied Economic Research (2015). India Human Development Survey-II (IHDS-II), 2011-12. Inter-University Consortium for Political and Social Research. doi: https://doi.org/10.3886/ICPSR36151.v6

109) Deshpande, A. (2011). The grammar of caste: Economic discrimination in contemporary India. Delhi: Oxford University Press.

110) Deshpande, A. (2013). Social justice through affirmative action in India: An assessment. In J. Wicks-Lim, & R. Pollin (Eds.), Capitalism on trial. Cheltenham: Edward Elgar Publishing. doi: https://doi.org/10.4337/9781782540854.00029

111) Deshpande, A., & Khanna, S. (2021). Can weak ties create social capital? Evidence from self-help groups in rural India. World Development, 146, 105534. doi: https://doi.org/10.1016/j.worlddev.2021.105534

112) Dhar, S. (2018). Gender and sustainable development goals (SDGs). Indian Journal of Gender Studies, 25(1), 47–78. doi: https://doi.org/10.1177/0971521517738451

113) Dhesi, A. S., & Singh, H. (1989). Education, labour market distortions and relative earnings of different religion-caste categories in India (A case study of Delhi). Canadian Journal of Development Studies, 10(1), 75–89. doi: https://doi.org/10.1080/02255189.1989.9669352

114) Di Pietro, G. (2018). The academic impact of natural disasters: Evidence from L'Aquila earthquake. Education Economics, 26(1), 62–77. https://doi.org/10.1080/09645292.2017.1394984

115) Dias, J., & Tebaldi, E. (2012). Institutions, human capital, and growth: The institutional mechanism. Structural Change and Economic Dynamics, 23(3), 300–312.

116) Dinda, S. (2008). Social capital in the creation of human capital and economic growth: A productive consumption approach. The Journal of Socio-Economics, 37(5), 2020–2033. https://doi.org/10.1016/j.socec.2007.06.014

117) Donald, S. G., & Lang, K. (2007). Inference with Difference-in-Differences and Other Panel Data. Review of Economics and Statistics, 89(2), 221–233. https://doi.org/10.1162/rest.89.2.221

118) Drzewiecki, D. M., Wavering, H. M., Milbrath, G. R., Freeman, V. L., & Lin, J. Y. (2020). The association between educational attainment and resilience to natural hazard-induced disasters in the West Indies: St. Kitts & Nevis. International Journal of Disaster Risk Reduction, 47, 101637. https://doi.org/10.1016/j.ijdrr.2020.101637

119) Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. American Economic Review, 91(4), 795–813. doi: https://doi.org/10.1257/aer.91.4.795

120) Duncan, G. J., & Hoffman, S. (1979). On-the-job training and earnings differences by race and sex. The Review of Economics and Statistics, 594–603.

121) Duraisamy, P. (2002). Changes in returns to education in India, 1983–94: By gender, age-cohort and location. Economics of Education Review, 21(6), 609–622.

122) Dutta, P. V. (2006). Returns to education: New evidence for India, 1983–1999. Education Economics, 14(4), 431–451. doi: https://doi.org/10.1080/09645290600854128

123) Eisnecker, P. S., & Adriaans, J. (2023). How do my earnings compare? Pay referents and just earnings. European Sociological Review, jcad002. doi: https://doi.org/10.1093/esr/jcad002

124) Ellenberg, J. H. (1994). Selection bias in observational and experimental studies. Statistics in Medicine, 13(5–7), 557–567. https://doi.org/10.1002/sim.4780130518

125) England, P., & Folbre, N. (2023). Reconceptualizing human capital. In M. Tåhlin (Ed.), A Research agenda for skills and inequality (pp. 177–195). Cheltenham: Edward Elgar Publishing. doi: https://doi.org/10.4337/9781800378469.00017

126) Ezaki, N. (2018). Impact of the 2015 Nepal earthquakes on children's schooling: Focusing on individual children's enrolment flow. Education 3-13, 46(7), 867–878. https://doi.org/10.1080/03004279.2017.1383502

127) Faggian, A., Modrego, F., & McCann, P. (2019). Human capital and regional development. In R. Capello, & P. Nijkamp (Eds.), Handbook of regional growth and development theories (pp. 149–171). Cheltenham: Edward Elgar Publishing. doi: https://doi.org/10.4337/9781788970020.00015

128) Fan, W., Ma, Y., & Wang, L. (2015). Do We Need More Public Investment in Higher Education? Estimating the External Returns to Higher Education in China*. Asian Economic Papers, 14(3), 88–104. https://doi.org/10.1162/ASEP_a_00380

129) Ferreira, S., Hamilton, K., & Vincent, J. R. (2013). Does development reduce fatalities from natural disasters? New evidence for floods. Environment and Development Economics, 18(6), 649–679. https://doi.org/10.1017/S1355770X13000387

130) Field, C., Barros, V., Stocker, T., & Dahe, Q. (Eds.). (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. Cambridge university press.

131) Filmer, D., & Pritchett, L. (1999). The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries. Population and Development Review, 25(1), 85–120. https://doi.org/10.1111/j.1728-4457.1999.00085.x

132) Filmer, D., Rogers, H., Angrist, N., & Sabarwal, S. (2020). Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education. Economics of Education Review, 77, 101971. https://doi.org/10.1016/j.econedurev.2020.101971

133) Flabbi, L., & Gatti, R. (2018). A primer on human capital. World Bank Policy Research Working Paper, 8309. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3105769

134) Fuchs, S., & Thaler, T. (Eds.). (2018). Vulnerability and Resilience to Natural Hazards (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316651148

135) Fulford, S. (2014). Returns to education in India. World Development, 59(C), 434–450. doi: https://doi.org/10.1016/j.worlddev.2014.02.005

136) Gao, W., & Smyth, R. (2015). Education expansion and returns to schooling in urban China, 2001–2010: Evidence from three waves of the

China Urban Labor Survey. Journal of the Asia Pacific Economy, 20(2), 178–201. https://doi.org/10.1080/13547860.2014.970607

137) Garbero, A., & Muttarak, R. (2013). Impacts of the 2010 Droughts and Floods on Community Welfare in Rural Thailand: Differential Effects of Village Educational Attainment. Ecology and Society, 18(4), art27. https://doi.org/10.5751/ES-05871-180427

138) Geetha Rani, P. (2014). Disparities in earnings and education in India. Cogent Economics & Finance, 2(1), 941510. doi: https://doi.org/10.1080/23322039.2014.941510

139) Ghosh, S., & Roy, S. (2022). Climate Change, Ecological Stress and Livelihood Choices in Indian Sundarban. In A. K. Enamul Haque, P. Mukhopadhyay, M. Nepal, & M. R. Shammin (Eds.), Climate Change and Community Resilience (pp. 399–413). Springer Singapore. https://doi.org/10.1007/978-981-16-0680-9_26

140) Gideon, M., Heggeness, M. L., Murray-Close, M., & Myers, S. L. (2022). The Returns to Education and the Determinants of Race and Gender Earnings Gaps Using Self-Reported versus Employer-Reported Earnings Data. Journal of Income Distribution®.

141) GoI. (2006). Social, economic and educational status of the Muslim community of India: A report. New Delhi: Prime Minister's High Level Committee, Cabinet Secretariat, Government of India. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/sachar_comm.pdf

142) GoI. (2009). The Right of Children to Free and Compulsory Education (RTE) Act, 2009. New Delhi: Government of India. Retrieved from

https://www.education.gov.in/sites/upload_files/mhrd/files/upload_document/rte.pdf

143) GoI. (2016). National commission for scheduled castes handbook-2016. New Delhi: National Commission for Scheduled Castes, Government of India.

144) Goldin, C. (2016). Human capital. In C. Diebolt, & M. Haupert (Eds.), Handbook of cliometrics (pp. 55–86). Berlin, Heidelberg: Springer. doi: https://doi.org/10.1007/978-3-642-40406-1_23

145) Goldin, C., & Katz, L. F. (2007). The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005 (Working Paper 12984). National Bureau of Economic Research. https://doi.org/10.3386/w12984

146) Goldin, C., & Katz, L. F. (2020). The Incubator of Human Capital: The NBER and the Rise of the Human Capital Paradigm. National Bureau of Economic Research. https://www.nber.org/papers/w26909

147) González, F. A. I., Santos, M. E., & London, S. (2021). Persistent effects of natural disasters on human development: Quasi-experimental evidence for Argentina. Environment, Development and Sustainability, 23(7), 10432–10454. https://doi.org/10.1007/s10668-020-01064-7

148) Green, D. A., & Craig Riddell, W. (2003). Literacy and earnings: An investigation of the interaction of cognitive and unobserved skills in earnings generation. Labour Economics, 10(2), 165–184. https://doi.org/10.1016/S0927-5371(03)00008-3

149) Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. Econometrica: Journal of the Econometric Society, 1–22.

150) Guaitoli, D. (2000). Human capital distribution, growth and convergence. Research in Economics, 54(4), 331–350.

151) Gunderson, M., & Oreopolous, P. (2020). Returns to education in developed countries. In The economics of education (pp. 39–51). Elsevier.

152) Han, S., & Mulligan, C. B. (2001). Human Capital, Heterogeneity and Estimated Degrees of Intergenerational Mobility. The Economic Journal, 111(470), 207–243. https://doi.org/10.1111/1468-0297.00606

153) Harmon, C., Oosterbeek, H., & Walker, I. (2003). The returns to education: Microeconomics. Journal of Economic Surveys, 17(2), 115–156.

154) Harmon, C., Walker, I., & Westergård-Nielsen, N. C. (Eds.). (2001). Education and earnings in Europe: A cross country analysis of the returns to education. Cheltenham: Edward Elgar Publishing.

155) Hartog, J., & Gerritsen, S. (2016). Mincer earnings functions for the Netherlands 1962–2012. De Economist, 164(3), 235–253. doi: https://doi.org/10.1007/s10645-016-9279-y

156) Harwell, M. R. (1988). Choosing Between Parametric and Nonparametric Tests. Journal of Counseling & Development, 67(1), 35–38. https://doi.org/10.1002/j.1556-6676.1988.tb02007.x

157) Heckman, J. (1990). Varieties of selection bias. The American Economic Review, 80(2), 313–318.

158) Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. Econometrica, 42(4), 679–694. doi: https://doi.org/10.2307/1913937

159) Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In Annals of economic and social measurement, volume 5, number 4 (pp. 475–492). NBER.

160) Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica, 47(1), 153–161. doi: https://doi.org/10.2307/1912352

161) Heckman, J. J., Humphries, J. E., & Veramendi, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. Journal of Political Economy, 126(S1), S197–S246.

162) Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In E. Hanushek, & F. Welch (Eds.), Handbook of the economics of education (Vol. 1, pp. 307–458). Amsterdam, Holland: Elsevier. doi: https://doi.org/10.1016/S1574-0692(06)01007-5

163) Heckman, J. J., Lochner, L., & Todd, P. E. (2003). Fifty years of Mincer earnings regressions. National Bureau of Economic Research Cambridge, Mass., USA.

164) Heckman, J., & Polachek, S. (1974). Empirical evidence on the functional form of the earnings-schooling relationship. Journal of the American Statistical Association, 69(346), 350–354. doi:https://doi.org/10.1080/01621459.1974.10482952

165) Hicks, D., & Duan, H. (2023). Education as opportunity? The causal effect of education on labor market outcomes in Jordan. Oxford Development Studies, 0(0), 1–19. doi: https://doi.org/10.1080/13600818.2023.2177264

166) Hoffmann, R., & Kassouf, A. L. (2005). Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure. Applied Economics, 37(11), 1303–1311. doi: https://doi.org/10.1080/00036840500118614

167) Hsiao, C. (2007). Panel data analysis—advantages and challenges. TEST, 16(1), 1–22. doi: https://doi.org/10.1007/s11749-007-0046-x

168) Hu, A., & Hibel, J. (2015). Increasing heterogeneity in the economic returns to higher education in urban China. The Social Science Journal, 52(3), 322–330. https://doi.org/10.1016/j.soscij.2013.09.002

169) Huang, J. (2013). Intergenerational transmission of educational attainment: The role of household assets. Economics of Education Review, 33, 112–123. https://doi.org/10.1016/j.econedurev.2012.09.013

170) Hunt, J. C., & Kau, J. B. (1985). Migration and Wage Growth: A Human Capital Approach. Southern Economic Journal, 51(3), 697–710. https://doi.org/10.2307/1057873

171) Hunter, B. A. (2016). Social Capital: Models and Efforts to Build and Restore among Marginalized Individuals and Communities. In A. G. Greenberg, T. P. Gullotta, & M. Bloom (Eds.), Social Capital and Community Well-Being: The Serve Here Initiative (pp. 199–212). Springer International Publishing. https://doi.org/10.1007/978-3-319-33264-2_12

172) Hussain, Y. R., & Mukhopadhyay, P. (2023). How Much do Education, Experience, and Social Networks Impact Earnings in India? A Panel Data Analysis Disaggregated by Class, Gender, Caste and Religion. SAGE Open, 13(4), 21582440231220942. https://doi.org/10.1177/21582440231220942

173) Jacob, J. F. (2018). Higher education in India from 1983 to 2014: Participation, access and labour market outcomes across socio-religious groups. Indian Journal of Human Development, 12(1), 74–92. doi: https://doi.org/10.1177/0973703018777180

174) Jana, S. K. (2020). Education in India: Goals and Achievements. In S. Hazra & A. Bhukta (Eds.), Sustainable Development Goals (pp. 57–77). Springer International Publishing. https://doi.org/10.1007/978-3-030-42488-6_4

175) Jaumotte, F. (2003). Female labour force participation: Past trends and main determinants in OECD countries. OECD Working Paper No. 376. doi: http://dx.doi.org/10.1787/082872464507

176) Johnson, G. E., & Stafford, F. P. (1973). Social returns to quantity and quality of schooling. Journal of Human Resources, 139–155.

177) Joshi, K. (2019). The impact of drought on human capital in rural India. Environment and Development Economics, 24(04), 413–436. https://doi.org/10.1017/S1355770X19000123

178) Kamhöfer, D. A., & Schmitz, H. (2016). Reanalyzing zero returns to education in Germany. Journal of Applied Econometrics, 31(5), 912–919. doi: https://doi.org/10.1002/jae.2461

179) Kanjilal-Bhaduri, S., & Pastore, F. (2018). Returns to Education and Female Participation Nexus: Evidence from India. The Indian Journal of Labour Economics, 61(3), 515–536. https://doi.org/10.1007/s41027-018-0143-2

180) Keller, G. (2018). Statistics for management and economics. Cengage Learning. https://thuvienso.hoasen.edu.vn/handle/123456789/11514

181) Khan, A., Chenggang, Y., Khan, G., & Muhammad, F. (2020). The dilemma of natural disasters: Impact on economy, fiscal position, and foreign direct investment alongside Belt and Road Initiative countries. Science of The Total Environment, 743, 140578. https://doi.org/10.1016/j.scitotenv.2020.140578

182) Khanna, G. (2023). Large-scale education reform in general equilibrium: Regression discontinuity evidence from India. Journal of Political Economy. doi: https://doi.org/10.1086/721619

183) Kikuchi, N. (2017). Marginal Returns to Schooling and Education Policy Change in Japan (SSRN Scholarly Paper 2943908). https://doi.org/10.2139/ssrn.2943908

184) Kim, Y., & Sherraden, M. (2011). Do parental assets matter for children's educational attainment?: Evidence from mediation tests. Children and Youth Services Review, 33(6), 969–979. https://doi.org/10.1016/j.childyouth.2011.01.003

185) Kingdon, G. G., & Theopold, N. (2008). Do returns to education matter to schooling participation? Evidence from India. Education Economics, 16(4), 329–350. doi: https://doi.org/10.1080/09645290802312453

186) Kingdon, G. G., & Unni, J. (2010). Education and Women's Labour Market Outcomes in India. Education Economics, 9(2), 173–195. https://doi.org/10.1080/09645290110056994

187) Klasen, S., Le, T. T. N., Pieters, J., & Silva, M. S. (2020). What drives female labour force participation? Comparable micro-level evidence from eight developing and emerging economies. The Journal of Development Studies, 57(3), 417–442. doi: https://doi.org/10.1080/00220388.2020.1790533

188) Klees, S. J. (2016). Human Capital and Rates of Return: Brilliant Ideas or Ideological Dead Ends? Comparative Education Review, 60(4), 644–672. https://doi.org/10.1086/688063

189) Korpi, M., & Clark, W. A. V. (2015). Internal migration and human capital theory: To what extent is it selective? Economics Letters, 136, 31–34. https://doi.org/10.1016/j.econlet.2015.08.016

190) Korwatanasakul, U. (2023). Returns to schooling in Thailand: Evidence from the 1978 Compulsory Schooling Law. The Developing Economies, 61(1), 3–35. doi: https://doi.org/10.1111/deve.12341

191) Kugler, A. D., & Kumar, S. (2017). Preference for boys, family size, and educational attainment in India. Demography, 54(3), 835–859. doi: https://doi.org/10.1007/s13524-017-0575-1

192) Kumar, A., & Hashmi, N. I. (2020). Labour market discrimination in India. The Indian Journal of Labour Economics, 63(1), 177–188. doi: https://doi.org/10.1007/s41027-019-00203-7

193) Leatherdale, S. T. (2019). Natural experiment methodology for research: A review of how different methods can support real-world research. International Journal of Social Research Methodology, 22(1), 19–35. https://doi.org/10.1080/13645579.2018.1488449

194) Leigh, A. (2008). Returns to education in Australia. Economic Papers, 27(3), 233–249. doi: https://doi.org/10.1111/j.1759-3441.2008.tb01040.x

195) Lemieux, T. (2006). The “Mincer Equation” Thirty Years After Schooling, Experience, and Earnings. In S. Grossbard (Ed.), Jacob Mincer A Pioneer of Modern Labor Economics (pp. 127–145). Kluwer Academic Publishers. https://doi.org/10.1007/0-387-29175-X_11

196) Lemieux, T. (2014). Occupations, fields of study and returns to education. Canadian Journal of Economics/Revue Canadienne d'économique, 47(4), 1047–1077. https://doi.org/10.1111/caje.12116

197) Li, H. (2003). Economic transition and returns to education in China. Economics of Education Review, 22(3), 317–328. doi: https://doi.org/10.1016/S0272-7757(02)00056-0

198) Liu, Y. (2017). Role of individual social capital in wage determination: Evidence from China. Asian Economic Journal, 31(3), 239–252. doi: https://doi.org/10.1111/asej.12122

199) Lin, N. (2017). Building a network theory of social capital. Social Capital, 3–28.**https://doi.org/10.1111/1467-6419.00127**

200) Lolayekar, A. P., & Mukhopadhyay, P. (2020). Understanding growth convergence in India (1981–2010): Looking beyond the usual suspects. PLOS ONE, 15(6), e0233549. doi: https://doi.org/10.1371/journal.pone.0233549

201) Lucas Jr, R. E. (1988). On the mechanics of economic development. Journal of Monetary Economics, 22(1), 3–42.

202) Lyons, E. (2020). The impact of job training on temporary worker performance: Field experimental evidence from insurance sales agents. Journal of Economics & Management Strategy, 29(1), 122–146. https://doi.org/10.1111/jems.12333

203) Madheswaran, S., & Singhari, S. (2016). Social exclusion and caste discrimination in public and private sectors in India: A decomposition analysis. The Indian Journal of Labour Economics, 59(2), 175–201. doi: https://doi.org/10.1007/s41027-017-0053-8

204) Makwana, N. (2019). Disaster and its impact on mental health: A narrative review. Journal of Family Medicine and Primary Care, 8(10), 3090. https://doi.org/10.4103/jfmpc.jfmpc_893_19

205) Mamgain, R. P. (2021). Understanding labour market disruptions and job losses amidst COVID-19. Journal of Social and Economic Development, 23(2), 301–319. doi: https://doi.org/10.1007/s40847-020-00125-x

206) Marginson, S. (2019). Limitations of human capital theory. Studies in Higher Education, 44(2), 287–301.

207) Marwah, V. (2019). Gender, caste and Indian feminism the case of the Women's Reservation Bill. In A. Aneja (Ed.), Women's and gender studies in India: Crossings (p. 13). London: Routledge India.

208) Mehrotra, S. (2019). Informal employment trends in the Indian economy: Persistent informality, but growing positive development. Working Paper No. 254. Geneva: International Labour Organization. https://www.ilo.org/employment/Whatwedo/Publications/working-papers/WCMS_734503/lang--en/index.htm

209) Mendiratta, P., & Gupt, Y. (2013). Private returns to education in India by gender and location: A Pseudo-Panel Approach. Arthaniti-Journal of Economic Theory and Practice, 12(1–2), 48–69. doi: https://doi.org/10.1177/0976747920130103

210) Mincer, J. (1958). Investment in human capital and personal income distribution. Journal of Political Economy, 66(4), 281–302. doi: https://doi.org/10.1086/258055

211) Mincer, J. (1962). On-the-Job Training: Costs, Returns, and Some Implications. Journal of Political Economy, 70(5, Part 2), 50–79. https://doi.org/10.1086/258725

212) Mincer, J. A. (1974). The human capital earnings function. In Schooling, experience, and earnings (pp. 83–96). NBER. https://www.nber.org/system/files/chapters/c1767/c1767.pdf

213) Mishra, M., & Ramakrishna, P. (Eds.). (2023). Education of socio-economic disadvantaged groups: From marginalisation to inclusion. London: Routledge India.

214) Mitra, A. (2019). Returns to education in India: Capturing the heterogeneity. Asia & the Pacific Policy Studies, 6(2), 151–169. doi: https://doi.org/10.1002/app5.271

215) Mohanty, S. (2021). A distributional analysis of the gender wage gap among technical degree and diploma holders in urban India. International Journal of Educational Development, 80, 102322. doi: https://doi.org/10.1016/j.ijedudev.2020.102322

216) Montenegro, C. E., & Patrinos, H. A. (2021). A data set of comparable estimates of the private rate of return to schooling in the world, 1970–2014. International Journal of Manpower, 44(6), 1248–1268. https://doi.org/10.1108/IJM-03-2021-0184

217) Moretti, E. (2005). Social returns to human capital. The National Bureau of Economic. https://www.nber.org/reporter/spring-2005/social-returns-human-capital

218) Mosse, D. (2018). Caste and development: Contemporary perspectives on a structure of discrimination and advantage. World Development, 110, 422–436. doi: https://doi.org/10.1016/j.worlddev.2018.06.003

219) Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Econometrica, 55(4), 765–799. doi: https://doi.org/10.2307/1911029

220) Mundlak, Y. (1978). On the pooling of time series and cross section data. Econometrica, 46(1), 69–85. doi: https://doi.org/10.2307/1913646

221) Mutch, C. (2018). The role of schools in helping communities copes with earthquake disasters: The case of the 2010–2011 New Zealand earthquakes. Environmental Hazards, 17(4), 331–351. https://doi.org/10.1080/17477891.2018.1485547

222) Muttarak, R., & Lutz, W. (2014). Is Education a Key to Reducing Vulnerability to Natural Disasters and hence Unavoidable Climate Change? Ecology and Society, 19(1), art42. https://doi.org/10.5751/ES-06476-190142

223) Nath, P. (2019). Employment scenario and the reservation policy. Economic and Political Weekly, 54(19), 56–61. https://www.epw.in/journal/2019/19/notes/employment-scenario-and-reservation-policy.html

224) NSO. (2021). Annual Report Periodic Labour Force Survey (PLFS) (July 2019 - June 2020), New Delhi: National Statistical Office, Ministry of Statistics and Programme Implementation, Government of India. Retrieved from https://www.mospi.gov.in/sites/default/files/publication_reports/Annual_Report_PLFS_2019_20F1.pdf

225) Paik, S. (2018). Impact of Natural Disasters on Schools in India: Current Scenario in Terms of Policies and Practices. In I. Pal & R. Shaw (Eds.), Disaster Risk Governance in India and Cross Cutting Issues (pp. 295–312). Springer Singapore. https://doi.org/10.1007/978-981-10-3310-0_15

226) Paldam, M. (2000). Social Capital: One or Many? Definition and Measurement. Journal of Economic Surveys, 14(5), 629–653. https://doi.org/10.1111/1467-6419.00127

227) Pandey, B. (2019). Ensure quality education for all in India: Prerequisite for achieving SDG 4. In S. Chaturvedi, T. C. James, S. Saha, & P. Shaw (Eds.), 2030 Agenda and India: Moving from Quantity to Quality (pp. 165–196). Singapore: Springer. doi: https://doi.org/10.1007/978-981-32-9091-4_8

228) Parida, Y., Agarwal Goel, P., Roy Chowdhury, J., Sahoo, P. K., & Nayak, T. (2021). Do economic development and disaster adaptation measures reduce the impact of natural disasters? A district-level analysis, Odisha, India. Environment, Development and Sustainability, 23(3), 3487–3519. https://doi.org/10.1007/s10668-020-00728-8

229) Pascapurnama, D. N., Murakami, A., Chagan-Yasutan, H., Hattori, T., Sasaki, H., & Egawa, S. (2018). Integrated health education in disaster risk reduction: Lesson learned from disease outbreak following natural disasters in Indonesia. International Journal of Disaster Risk Reduction, 29, 94–102. https://doi.org/10.1016/j.ijdrr.2017.07.013

230) Pascarella, E. T., Smart, J. C., & Smylie, M. A. (1992). College tuition costs and early career socioeconomic achievement: Do you get what you pay for? Higher Education, 24(3), 275–290. https://doi.org/10.1007/BF00128447

231) Patrinos, H. (2016). Estimating the return to schooling using the Mincer equation. IZA World of Labor, 278, 1–11. doi: https://doi.org/10.15185/izawol.278

232) Patrinos, H. A., & Psacharopoulos, G. (2020). Returns to education in developing countries. In S. Bradley, & C. Green (Eds.), The Economics of Education (2nd ed., pp. 53–64). Cambridge: Academic Press. doi: https://doi.org/10.1016/B978-0-12-815391-8.00004-5

233) Paudel, J., & Ryu, H. (2018). Natural disasters and human capital: The case of Nepal's earthquake. World Development, 111, 1–12. https://doi.org/10.1016/j.worlddev.2018.06.019

234) Peet, E. D., Fink, G., & Fawzi, W. (2015). Returns to education in developing countries: Evidence from the living standards and measurement study surveys. Economics of Education Review, 49, 69–90. doi: https://doi.org/10.1016/j.econedurev.2015.08.002

235) Perkins, K. L. (2019). Changes in Household Composition and Children's Educational Attainment. Demography, 56(2), 525–548. https://doi.org/10.1007/s13524-018-0757-5

236) Pianta, R. C., & Ansari, A. (2018). Does Attendance in Private Schools Predict Student Outcomes at Age 15? Evidence From a Longitudinal Study. Educational Researcher, 47(7), 419–434. https://doi.org/10.3102/0013189X18785632

237) Polachek, S. W. (2008). Earnings over the life cycle: The Mincer earnings function and its applications. Foundations and Trends in Microeconomics, 4(3), 165–272. doi: https://doi.org/10.1561/0700000018

238) Psacharopoulos, G. (1985). Returns to education: A further international update and implications. Journal of Human Resources, 583–604.

239) Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. Education Economics, 26:5, 445-458, doi: 10.1080/09645292.2018.1484426

240) Rani, P. G. (2013). Exploring earnings and education disparities in India across region, caste, religion and English language ability. Artha Vijnana, 55(4), 402–420. doi: https://doi.org/10.21648/arthavij%2F2013%2Fv55%2Fi4%2F111243

241) Ray, R., & Katzenstein, M. F. (2005). Social movements in India: Poverty, power, and politics. Oxford: Rowman & Littlefield.

242) Riley, J. G. (1979). Testing the Educational Screening Hypothesis. Journal of Political Economy, 87(5, Part 2), S227–S252. https://doi.org/10.1086/260830

243) Rios-Avila, F. (2021). XTHECKMANFE: Stata module to fit panel data models in the presence of endogeneity and selection. Statistical Software Components S458770, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s458770.html

244) Romer, P. M. (1989). Human capital and growth: Theory and evidence. National Bureau of Economic Research Cambridge, Mass., USA.

245) Rossi, P. E. (2014). Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. Marketing Science, 33(5), 655–672. doi: https://doi.org/10.1287/mksc.2014.0860

246) Roy, D., Saroj, S., & Pradhan, M. (2022). Nature of employment and outcomes for urban labor: Evidence from the latest labor force surveys in

India. Indian Economic Review, 57(1), 165–221. doi: https://doi.org/10.1007/s41775-022-00131-2

247) Sander, W. (1992). The effects of ethnicity and religion on educational attainment. Economics of Education Review, 11(2), 119–135. https://doi.org/10.1016/0272-7757(92)90003-L

248) Sarkar, S., Sahoo, S., & Klasen, S. (2019). Employment transitions of women in India: A panel analysis. World Development, 115, 291–309. doi: https://doi.org/10.1016/j.worlddev.2018.12.003

249) Schmitt, J., & Wadsworth, J. (2006). Is there an impact of household computer ownership on children's educational attainment in Britain? Economics of Education Review, 25(6), 659–673. https://doi.org/10.1016/j.econedurev.2005.06.001

250) Schultz, T. P. (2003). Human capital, schooling and health. Economics & Human Biology, 1(2), 207–221.

251) Schultz, T. W. (1961). Investment in human capital. American Economic Review, 51(1), 1–17. https://www.jstor.org/stable/1818907

252) Schündeln, M., & Playforth, J. (2014). Private versus social returns to human capital: Education and economic growth in India. European Economic Review, 66, 266–283. https://doi.org/10.1016/j.euroecorev.2013.08.011

253) Schwiebert, J. (2012). Revisiting the composition of the female workforce: A Heckman selection model with endogeneity. Hannover Economic Papers (HEP) DP-502. Hannover: University of Leibniz.

254) Semykina, A., & Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. Journal of Econometrics, 157(2), 375–380. doi: https://doi.org/10.1016/j.jeconom.2010.03.039

255) Sen, A. (2000). Development as freedom. New York: Anchor Books.

256) Sen, A. (2001). Economic development and capability expansion in historical perspective. Pacific Economic Review, 6(2), 179–191. doi: https://doi.org/10.1111/1468-0106.00126

257) Sengupta, A. (2023). Effect of COVID-19 pandemic on employment and earning in urban India during the first three months of pandemic period: An analysis with unit-level data of periodic labour force survey. The Indian Journal of Labour Economics, 66(1), 283–298. doi: https://doi.org/10.1007/s41027-023-00428-7

258) Sengupta, A., & Das, P. (2014). Gender wage discrimination across social and religious groups in India: Estimates with unit level data. Economic and Political Weekly, 49(21), 71–76. https://www.jstor.org/stable/24479554

259) Sheskin, D. J. (2003). Handbook of parametric and nonparametric statistical procedures. Chapman and hall/CRC. https://www.taylorfrancis.com/books/mono/10.1201/9781420036268/handbook-parametric-nonparametric-statistical-procedures-david-sheskin

260) Sianesi, B., & Reenen, J. V. (2003). The returns to education: Macroeconomics. Journal of Economic Surveys, 17(2), 157–200. https://doi.org/10.1111/1467-6419.00192

261) Singha Roy, N. (2020). Wage rate: Is this return to education or return to physical capability? Evidence from rural India. The Indian Journal of Labour Economics, 63(1), 99–117. doi: https://doi.org/10.1007/s41027-020-00205-w

262) Skidmore, M., & Toya, H. (2002). Do Natural disasters promote long-run growth ? Economic Inquiry, 40(4), 664–687. https://doi.org/10.1093/ei/40.4.664

263) Smith, S. S. (2000). Mobilizing social resources: Race, ethnic, and gender differences in social capital and persisting wage inequalities. The Sociological Quarterly, 41(4), 509–537. doi: https://doi.org/10.1111/j.1533-8525.2000.tb00071.x

264) Sobel, I. (1978). The Human Capital Revolution in Economic Development: Its Current History and Status. Comparative Education Review, 22(2), 278–308.

265) Solecki, W., Leichenko, R., & O'Brien, K. (2011). Climate change adaptation strategies and disaster risk reduction in cities: Connections, contentions, and synergies. Current Opinion in Environmental Sustainability, 3(3), 135–141. https://doi.org/10.1016/j.cosust.2011.03.001

266) Soukiazis, E., & Cravo, T. (2008). Human Capital and the Convergence Process Among Countries. Review of Development Economics, 12(1), 124–142. https://doi.org/10.1111/j.1467-9361.2008.00438.x

267) Suar, D., & Kar, S. (2005). Social and Behavioural Consequences of the Orissa Supercyclone. Journal of Health Management, 7(2), 263–275. https://doi.org/10.1177/097206340500700207

268) Takasaki, Y. (2017). Do Natural Disasters Decrease the Gender Gap in Schooling? World Development, 94, 75–89. https://doi.org/10.1016/j.worlddev.2016.12.041

269) Tamuly, R., & Mukhopadhyay, P. (2022). Natural disasters and well-being in India: A household-level panel data analysis. International Journal of

Disaster Risk Reduction, 79, 103158.
https://doi.org/10.1016/j.ijdrr.2022.103158

270) Tan, E. (2014). Human Capital Theory: A Holistic Criticism. Review of Educational Research, 84(3), 411–445.
https://doi.org/10.3102/0034654314532696

271) Teixeira, P. N. (2014). Gary Becker’s early work on human capital – collaborations and distinctiveness. IZA Journal of Labor Economics, 3(1), 12.
https://doi.org/10.1186/s40172-014-0012-2

272) Thamtanajit, K. (2020). The Impacts Of Natural Disaster On Student Achievement: Evidence From Severe Floods in Thailand. The Journal of Developing Areas, 54(4). https://doi.org/10.1353/jda.2020.0042

273) The World Bank. (n.d.). World development indicators. Retrieved from The World Bank Data Catalog:
https://datacatalog.worldbank.org/dataset/world-development-indicators

274) Tian, B., Lin, C., Zhang, W., & Feng, C. (2022). Tax Incentives, On-the-job Training, and Human Capital Accumulation: Evidence from China. China Economic Review, 75, 101850.
https://doi.org/10.1016/j.chieco.2022.101850

275) Tilak, J. B. G. (Ed.). (2021). Education in India: Policy and practice. New Delhi: Sage Publications India Pvt Ltd.

276) Toya, H., & Skidmore, M. (2007). Economic development and the impacts of natural disasters. Economics Letters, 94(1), 20–25.
https://doi.org/10.1016/j.econlet.2006.06.020

277) Tran, D. B. (2023). Returns to education revisited: Evidence from rural Vietnam. Cogent Education, 10(1), 2184019.

278) Truong, T. H., Ogawa, K., & Sanfo, J.-B. M. B. (2021). Educational expansion and the economic value of education in Vietnam: An instrument-free analysis. International Journal of Educational Research Open, 2, 100025. doi: https://doi.org/10.1016/j.ijedro.2020.100025

279) Tsai, S.-L., & Xie, Y. (2011). Heterogeneity in returns to college education: Selection bias in contemporary Taiwan. Social Science Research, 40(3), 796–810. https://doi.org/10.1016/j.ssresearch.2010.12.008

280) Tuijnman, A., Chinapah, V., & Fägerlind, I. (1988). Adult education and earnings: A 45-year longitudinal study of 834 Swedish men. Economics of Education Review, 7(4), 423–437. https://doi.org/10.1016/0272-7757(88)90034-9

281) Vasudeva Dutta, P. (2006). Returns to education: New evidence for India, 1983–1999. Education Economics, 14(4), 431–451.

282) Vatta, K., Sato, T., & Taneja, G. (2016). Indian Labour Markets and Returns to Education. Millennial Asia, 7(2), 107–130. https://doi.org/10.1177/0976399616655001

283) Weisbrod, B. A. (1962). Education and Investment in Human Capital. Journal of Political Economy, 70(5, Part 2), 106–123. https://doi.org/10.1086/258728

284) Wincenciak, L., Grotkowska, G., & Gajderowicz, T. (2022). Returns to education in Central and Eastern European transition economies: The role of macroeconomic context. Research in Comparative and International Education, 17(4), 655–676.

285) Wooldridge, J. M. (2018). Introductory econometrics: A modern approach (7th ed.). Mason: Cengage Learning.

286) Yezer, A. M. J., & Thurston, L. (1976). Migration Patterns and Income Change: Implications for the Human Capital Approach to Migration. Southern Economic Journal, 42(4), 693–702. https://doi.org/10.2307/1056262

287) Yubilianto. (2020). Return to education and financial value of investment in higher education in Indonesia. Journal of Economic Structures, 9(1), 17. https://doi.org/10.1186/s40008-020-00193-6

288) Zhang, Z., & Zhang, P. (2023). The long-term impact of natural disasters on human capital: Evidence from the 1975 Zhumadian flood. International Journal of Disaster Risk Reduction, 91, 103671.

289) Zhao, Y. (1999). Labor Migration and Earnings Differences: The Case of Rural China. Economic Development and Cultural Change, 47(4), 767–782. https://doi.org/10.1086/452431