WAGE INEQUALITY AND SKILL DEVELOPMENT IN INDIA

Examining Distributional Wage Effects Using Quantile Regression | PLFS 2023 to 24

A Dissertation Submitted
In Partial Fulfilment of the Requirements
for the Degree of
MASTER OF ARTS IN ECONOMICS
in
Economics

by

BHARGAVI
(24/MAE/23)

Under the Supervision of

Dr. Neha Jain
Assistant Professor, Department of M.A. Economics



Department of M.A. Economics
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India
May, 2026

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR(S)

Certified that Bhargavi (Enrolment No. 24/MAE/23) has carried out the research work presented in this dissertation entitled "WAGE INEQUALITY AND SKILL DEVELOPMENT IN INDIA: Examining Distributional Wage Effects Using Quantile Regression | PLFS 2023 to 24" for the award of the Degree of Master of Arts in Economics from the Department of M.A. Economics, Delhi Technological University, Delhi, under my supervision. The dissertation embodies the results of original work carried out by the student herself. The contents of the dissertation do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Signature: _____

Name: Dr Neha Jain

Designation: Assistant Professor

Department: M.A. Economics, Delhi Technological University

Date: _____

Place: Delhi

CANDIDATE'S DECLARATION

I, Bhargavi (Enrolment No. 24/MAE/23), hereby certify that the work presented in this dissertation entitled "WAGE INEQUALITY AND SKILL DEVELOPMENT IN INDIA: Examining Distributional Wage Effects Using Quantile Regression | PLFS 2023 to 24", submitted in partial fulfilment of the requirements for the award of the Degree of Master of Arts in Economics, Department of M.A. Economics, Delhi Technological University, is an authentic record of my own work carried out during the period from July 2025 to May 2026, under the supervision of Dr. Neha Jain.

I have not submitted the matter presented in this dissertation for the award of any other degree of this or any other Institute.

Candidate's Signature: _____

Name: Bhargavi

Roll No.: 24/MAE/23

Date: _____

This is to certify that the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor(s): _____

Signature of External Examiner: _____

ACKNOWLEDGEMENTS

This dissertation grew out of questions that stayed with me through two years of coursework: why do skill programmes in India rarely seem to reach the workers who most need them, and what does it actually mean when we say training 'improves wages' if we cannot say for whom? The people listed below helped me find partial answers, though the questions keep expanding.

I owe the greatest intellectual debt to my supervisor, Dr Neha Jain. Her guidance on the econometric approach was invaluable, and her habit of asking 'what does this mean for a woman in rural Bihar?' kept the analysis grounded in something real. Whatever clarity this dissertation has is largely her doing; the remaining confusion is entirely mine.

The Department of M.A. Economics at Delhi Technological University provided the intellectual environment and coursework in Labour Economics and Applied Econometrics that shaped my approach to this problem. I am grateful to all my teachers here.

The data from the Periodic Labour Force Survey 2023 to 24, made publicly available by the Ministry of Statistics and Programme Implementation, Government of India, is the empirical foundation of this work. I am grateful to MoSPI for making such rich microdata accessible to researchers.

My classmates listened patiently to half-formed ideas and asked uncomfortable questions that improved them. Several friends also read early drafts and told me plainly when the writing did not make sense, which was often.

My family supported this work in ways that cannot be fully acknowledged in an acknowledgements section. Their confidence in me was steadier than my confidence in myself.

Bhargavi
24/MAE/23
Department of M.A. Economics
Delhi Technological University
May 2026

WAGE INEQUALITY AND SKILL DEVELOPMENT IN INDIA

Bhargavi

ABSTRACT

India's labour market has gone through considerable restructuring over the past two decades. Economic liberalisation, rapid technological change, and an expanding service sector have together produced a labour market in which skilled workers capture a disproportionately large share of wage gains, while low-wage workers, who make up the vast majority of the workforce, continue to face stagnant real earnings and precarious employment. Against this backdrop, successive central government administrations have invested heavily in skill development programmes, from the Industrial Training Institutes (ITIs) to the Pradhan Mantri Kaushal Vikas Yojana (PMKVY) and the Deen Dayal Upadhyaya Grameen Kaushalya Yojana (DDU-GKY), disbursing over ₹6,000 crore annually in an effort to close the skill gap and improve labour market outcomes for disadvantaged workers. What remains thin, even so, is robust, distribution-sensitive evidence on whether these interventions actually reach the workers who need them most.

This dissertation looks at the relationship between skill development and wages across the full wage distribution of Indian workers, using data from the Periodic Labour Force Survey (PLFS) 2023 to 24, the most recent nationally representative household survey conducted by the Ministry of Statistics and Programme Implementation, Government of India. The analytic sample comprises 46,294 wage workers aged 15 to 65 drawn from both rural and urban areas across all major states. The core estimation strategy is quantile regression (Koenker & Bassett, 1978), which allows the conditional wage effects of education, vocational training, and other worker characteristics to be estimated at five quantile points, the 10th, 25th, 50th, 75th, and 90th percentiles of the log-wage distribution, directly addressing the main limitation of prior Indian studies that rely on OLS and estimate only mean effects.

The econometric model is grounded in an augmented Mincerian earnings function. The dependent variable is the natural log of the monthly wage. Key regressors include years of formal education, a binary indicator for formal vocational training (skill dummy, equal to 1 if the worker received certified vocational training), potential labour market experience constructed as age minus years of education minus 6, experience squared, and a set of demographic controls for gender (female dummy), rural/urban location (urban dummy), and social group membership (SC, ST, and OBC dummies), with 'Others' as the reference category. Employment type (self-employed, regular wage, casual labour) is also incorporated in the OLS specifications. Robust standard errors are used throughout.

Six main findings emerge. First, returns to formal education rise monotonically across quantiles, from 6.9 per cent per year at the 10th percentile to 11.5 per cent at the 90th percentile, confirming that education is a force for wage divergence at current rates of return. Second, vocational training yields an inverted-U pattern: returns peak at the median (18.2 per cent), are moderate at the lower tail (11.4 per cent at Q10), and decline at the upper tail (10.5 per cent at Q90), making training more egalitarian in its distributional footprint than formal education. Third, the female wage penalty is largest at the bottom of the distribution (−67.9 log points at Q10) and narrows substantially but remains large at the top (−19.4 log points at Q90), consistent with the sticky floor hypothesis. Fourth, SC and OBC wage penalties widen at higher quantile points, indicating that caste-based barriers to high-skill, high-wage employment are the primary mechanism of disadvantage rather than uniform wage suppression. Fifth, the urban wage premium is largest at Q10 (23.2 per cent) and declines modestly to Q90 (17.1 per cent), suggesting urban labour markets provide proportionately greater gains to low-wage workers. Sixth, experience returns are stable across quantiles at approximately 4 per cent per year and do not contribute to distributional divergence.

These findings carry direct policy implications. Targeting PMKVY and DDU-GKY enrolment toward low-wage workers, particularly women, SC/ST workers, and rural youth, would concentrate public resources where the relative returns to training are meaningful. Reforming curriculum and certification to align with industry demand, expanding training infrastructure in rural areas, designing gender-disaggregated programmes, and strengthening anti-discrimination enforcement in higher-wage employment are complementary interventions that training policy alone cannot substitute.

Keywords: Wage Inequality, Skill Development, Vocational Training, Quantile Regression, PLFS 2023 to 24, Human Capital Theory, India, Labour Market, Gender Gap, Caste Discrimination

LIST OF ABBREVIATIONS

Abbreviation	Full Form
DDU to GKY	Deen Dayal Upadhyaya Grameen Kaushalya Yojana
DTU	Delhi Technological University
IHDS	India Human Development Survey
ITI	Industrial Training Institute
LFPR	Labour Force Participation Rate
MoSPI	Ministry of Statistics and Programme Implementation
MSDE	Ministry of Skill Development and Entrepreneurship
NSDC	National Skill Development Corporation
NSS	National Sample Survey
NSSO	National Sample Survey Organisation
OBC	Other Backward Class
OLS	Ordinary Least Squares
PLFS	Periodic Labour Force Survey
PMKVY	Pradhan Mantri Kaushal Vikas Yojana
QR	Quantile Regression
RPL	Recognition of Prior Learning
SBTC	Skill-Biased Technological Change
SC	Scheduled Caste
SE	Standard Error
ST	Scheduled Tribe
UGC	University Grants Commission
VET	Vocational Education and Training
VT	Vocational Training

CHAPTER 1: INTRODUCTION

1.1 Background of the Study

Whether education and training translate into higher wages is hardly a new question in economics. What is far less settled, and arguably more important from a policy standpoint, is whether the gains from such investments actually reach the workers who need them most. The conventional empirical tool for this kind of question, ordinary least squares regression applied to a Mincerian earnings function, was built to estimate average effects of schooling and experience on wages. It does this job well enough. What it cannot do, however, is say anything about how those effects are spread across the wage distribution: whether the benefits of skill investment land mostly among workers already earning relatively well, or whether they reach those at the bottom who arguably stand to gain the most.

This distinction is not merely technical. A training programme that mainly lifts wages for workers already in the upper half of the distribution might raise aggregate output while simultaneously widening the gap between rich and poor. A programme whose gains are concentrated among the lowest earners might look unremarkable in terms of average impact, yet do considerable redistributive work. Without evidence that speaks to this distributional question, policymakers are, in a sense, working blind on one of the most consequential features of any skill investment: its equity profile.

India makes this an urgent question rather than an abstract one. With a workforce exceeding 500 million, an economy in which over 90 per cent of workers lack formal employment protection, and deep structural divides running along lines of gender, caste, and rural-urban location, the question of who benefits from skill development is far from academic. Since 2015, successive governments have poured resources into the Pradhan Mantri Kaushal Vikas Yojana (PMKVY) and related schemes under the Ministry of Skill Development and Entrepreneurship, with annual budgets exceeding ₹6,000 crore and ambitious targets for skilling the workforce. Whether this spending is progressive, reaching low-wage workers and narrowing gaps, or regressive, mainly benefiting those already better off, is a question the existing evidence base simply cannot answer with confidence.

Part of the reason lies in methodology. Most Indian studies of training returns rely on OLS applied to cross-sectional survey data, which yields the conditional mean effect of training on wages. That is informative as far as it goes, but it is silent on distributional heterogeneity. Quantile regression, introduced by Koenker and Bassett (1978) and by now a standard tool in applied labour economics, allows the researcher to estimate the effect of a regressor at any

chosen point of the conditional wage distribution. If the education coefficient rises steadily across quantiles, say from 7 per cent at Q10 to 11 per cent at Q90, this points to education as a force widening inequality. If the training coefficient instead declines, or follows an inverted-U shape peaking around the median, this points to training playing a more equalising role. OLS would compress all of this into a single average coefficient, masking the underlying pattern entirely.

This dissertation applies quantile regression to the Periodic Labour Force Survey (PLFS) 2023-24, the most recent round of India's nationally representative labour force survey, using a sample of 46,294 wage workers. It estimates the distributional wage effects of formal education, formal vocational training, labour market experience, gender, rural/urban location, and social group membership (SC, ST, OBC) across five quantile points: Q10, Q25, Q50, Q75, and Q90. In doing so, three gaps in the existing literature are addressed: a methodological one, the near-absence of distributional analysis in Indian studies; a temporal one, no published study has used PLFS 2023-24 for this purpose; and what might be called an intersectional gap, limited work that examines gender, caste, and training returns jointly across the wage distribution.

The choice of five quantile points follows fairly standard practice in the distributional wage literature and offers a reasonably detailed profile without producing an unwieldy number of coefficients. Q10 and Q25 capture the lower tail, where low-wage casual workers are concentrated. Q50 represents something close to the median worker, likely a semi-skilled employee in informal manufacturing or services. Q75 and Q90 capture higher-wage workers in formal or semi-formal employment. Together, these five points span the range most relevant for judging the equity impact of skill programmes.

There is also a practical advantage to quantile regression worth flagging here. When wage variance differs substantially across groups, and in India it clearly does, with female workers facing a noticeably more compressed wage distribution than men, OLS estimates become inefficient. Quantile regression handles this without requiring the researcher to specify the exact form of the heteroscedasticity, and the robust standard errors obtained through vce(robust) provide an additional layer of protection against variance misspecification.

1.2 The Indian Context

India's labour market is the second-largest in the world by headcount and among the most structurally complex anywhere. A few features of this context are worth dwelling on, since they shape why the distributional question matters so much here.

The scale of informality is probably the single defining feature. More than 90 per cent of India's workers are employed in the informal sector, without formal contracts, social security coverage, or any meaningful employment protection. This informality produces extreme wage heterogeneity. A casual construction worker and an informal IT consultant both technically sit within the informal sector, yet their monthly earnings might differ by a factor of twenty or more. Standard labour market statistics that average across this range, median wages, mean training returns, and so on, end up obscuring the very distributions that matter most for understanding how the economy actually functions for most people.

India's demographic profile adds a further layer of urgency. With 65 per cent of the population below 35 years of age, the country sits at the peak of its demographic dividend window, the period during which the working-age share of the population is high enough to support accelerated growth, provided the labour force is adequately skilled and productively absorbed. Realising this dividend is not simply a matter of having more workers; it requires more productive ones. The gap between new labour market entrants, roughly 10 to 12 million each year, and the absorptive capacity of the formal economy remains one of the defining policy challenges of the present period (Kapoor, 2019; Mehrotra & Parida, 2019).

The caste system continues to shape labour market outcomes in ways that are well documented and far from subtle. Scheduled Caste (SC) and Scheduled Tribe (ST) workers consistently earn less than comparable workers from other social groups, even after accounting for education, experience, and sector of employment. Madheswaran and Attewell (2007), drawing on NSS data, showed that discrimination, not merely differences in human capital, explains a substantial share of the SC wage gap. More recent work by Deshpande and Ramachandran (2019) confirms that this disadvantage has not faded with rising average incomes. One question this study is well placed to address is whether caste disadvantage is spread evenly across the wage distribution or concentrated in particular segments of it.

Gender discrimination in Indian labour markets is, if anything, even more pronounced. The gender wage gap in India ranks among the largest in Asia. Female labour force participation has declined over the reform period even as male participation has stayed broadly stable. Women who do enter wage employment tend to be concentrated in casual, low-skill work, precisely where regulation is weakest and discrimination most pervasive (Uma Rani, 2019; Parida & Awasthi, 2020). The idea that gender discrimination operates most powerfully at the bottom of the wage distribution, a "sticky floor" rather than a "glass ceiling", is something the quantile regression framework can test directly.

India's skill policy landscape has also shifted considerably since 2015. PMKVY, now in its fourth phase, claims to have trained over 14 million beneficiaries through short-duration courses spanning manufacturing, services, and construction. The DDU-GKY scheme, aimed at rural below-poverty-line youth, runs separately with a placement-linked design. ITIs continue to offer longer technical programmes. Independent evaluations of PMKVY have generally found positive but modest effects on placement, with weaker evidence on actual wage gains, particularly for women and rural workers (Ministry of Skill Development and Entrepreneurship, 2019; Gill et al., 2024). Whether these programmes are reaching the parts of the workforce where the relative returns to training are highest is precisely the kind of question a distributional analysis of PLFS data is suited to answer.

1.3 Importance of the Study

This research speaks to at least three different conversations at once.

The first is theoretical. Does human capital investment in the form of vocational training complement or substitute for a worker's existing position in the wage distribution? If returns to training turn out to be higher for workers already near the top, training would essentially compound existing advantage and widen inequality further. If returns are higher at the bottom, or at least not concentrated at the top, training looks more like a progressive redistributive tool. Quantile regression lets this question be settled empirically rather than assumed on theoretical grounds alone.

The need for distribution-sensitive evidence becomes obvious once the sheer diversity of India's workforce is taken into account. In a labour market spanning everything from subsistence agricultural day labour to senior corporate roles, a single average treatment effect tells a policymaker trying to allocate a fixed training budget very little. The relevant question is not simply whether training raises wages on average, but which workers gain the most from a given rupee of training spending, and which workers face barriers that prevent training from converting into wage gains even when access is available. These are inherently distributional questions and require distributional methods.

The structural shifts of the past decade only add to the urgency. The COVID-19 pandemic disrupted Indian labour markets in ways the data are still catching up with, the recovery of formal employment has been uneven, and low-skill workers in contact-intensive services have been among the slowest to recover. Platform and gig employment has expanded rapidly, creating new forms of informal work that fit awkwardly into either traditional casual labour or traditional salaried categories. All of this makes it more important, not less, to understand

the distributional wage effects of skill investment under current conditions rather than relying on survey rounds from 2011-12 or 2017-18.

The second contribution is empirical. Most Indian studies on training returns lean on NSS data from 2011-12 or earlier, with a handful using PLFS 2017-18 or 2018-19. No published study, to date, applies quantile regression to PLFS 2023-24. This gap matters because the Indian labour market has changed in substantial ways since 2012: PMKVY was launched, COVID-19 disrupted employment patterns, platform and gig work expanded, and PLFS itself has improved considerably in coverage and frequency. Evidence drawn from a decade-old survey round may simply no longer describe how wages and skills relate to one another today.

The third contribution is directly policy-relevant. India's skill development budget is large and growing, and the Ministry of Skill Development and Entrepreneurship is currently reviewing PMKVY's design and targeting criteria. Distribution-sensitive evidence, specifically on which segments of the wage distribution benefit most from vocational training, gives policymakers a principled basis for targeting decisions. If, as this study finds, training returns cluster around the median while low-wage workers gain proportionately less, then access-side interventions for the lowest earners, such as subsidised transport, stipends, and childcare support, could raise both the efficiency and the equity of public spending on skills.

1.4 Research Problem

Put plainly, the central problem is this. India's current evidence base on the wage effects of skill development rests mainly on OLS estimates of average effects, drawn from data that is now over a decade old in most cases. As a result, policymakers designing and evaluating skill programmes lack reliable, distribution-sensitive evidence on who benefits from training, by how much, and at which point in the wage distribution the gains are concentrated. Without this evidence, there is no real way to assess whether current programmes are progressive or regressive, let alone redesign them with equity in mind. This dissertation addresses that gap by applying quantile regression to PLFS 2023-24, the most recent nationally representative data on Indian wages, and reporting the full distributional profile of skill returns.

1.5 Research Gap

A systematic look at the existing literature, covered in detail in Chapter 2, points to five specific gaps. First, most Indian studies on wage determination and training returns rely on OLS or fixed-effects estimation, both of which produce only conditional mean effects. The distributional heterogeneity that matters most for inequality assessment is simply invisible to these approaches. Second, no study has used PLFS 2023-24 with distributional methods,

leaving the most contemporary evidence unexplored. Third, very few studies examine gender, caste, and training returns jointly within a single distributional framework, most treat these dimensions separately. Fourth, the informal sector, despite employing over 90 per cent of Indian workers, is rarely the central focus of distributional wage analysis. Fifth, the interaction between vocational training participation and position in the wage distribution, specifically whether training benefits low-wage workers proportionately more or less, has not been examined in the Indian context using quantile regression on recent data.

As the uploaded literature review materials also note, there is a deeper theoretical gap underlying all of this: no India-specific framework currently integrates labour market informality, social stratification by caste and gender, and the institutional structure of Indian skill provision into a single unified model of distributional wage determination. This dissertation does not claim to close that theoretical gap. What it does offer is empirical evidence that any such framework would eventually need to account for.

1.6 Research Questions

Four research questions guide the analysis. RQ1: How do formal education and formal vocational training affect wages across the conditional wage distribution of Indian workers, as estimated from PLFS 2023-24? RQ2: Does vocational training reduce or widen wage inequality, judging by the pattern of training returns across Q10, Q25, Q50, Q75, and Q90? RQ3: How do returns to skill development vary across sub-groups defined by gender, rural/urban location, and social group membership? RQ4: What policy implications follow from the observed distributional pattern of skill returns for the design of India's skill development programmes?

1.7 Objectives of the Study

Four objectives correspond to these research questions. The first is to examine how skill development, measured through years of formal education and formal vocational training participation, affects wages across the wage distribution using PLFS 2023-24 data and quantile regression. The second is to assess whether skill development reduces or widens wage inequality, by comparing the magnitude and statistical significance of skill returns at the five quantile points. The third is to analyse heterogeneity in skill returns across gender, location, and social group, identifying subpopulations where skill investment appears most, or least, effective. The fourth is to draw out targeted, evidence-based policy implications for the design of more inclusive skill development programmes in India.

1.8 Hypotheses

Four working hypotheses are tested. H1 (Skill-Biased Inequality): returns to formal education increase monotonically with wage quantile, so that higher-wage workers gain proportionately more from each additional year of schooling. H2 (Progressive Training Effect): returns to vocational training are not concentrated among high-wage workers; specifically, the training premium remains meaningful at lower quantiles, suggesting a more egalitarian distributional footprint than that of formal education. H3 (Sticky Floor): the gender wage penalty is larger at lower wage quantiles, consistent with the idea that discrimination operates most powerfully at the bottom of the distribution, where informal, low-skill employment predominates. H4 (Upper-End Caste Barrier): SC and OBC wage penalties widen at higher quantile points, suggesting that caste-based exclusion is most binding in higher-skill, higher-wage segments of employment.

1.9 Scope of the Study

The study covers wage workers aged 15 to 65 in India as captured by PLFS 2023-24, drawing on both rural and urban observations across all major states for which the survey provides representative data. The analytic sample is restricted to workers reporting positive wages during the survey reference period; self-employed workers and unpaid family helpers are excluded, since their income determination involves entrepreneurial returns and intra-household transfers that a Mincerian wage equation is not well suited to capture. The temporal scope is cross-sectional, limited to the 2023-24 survey round. No attempt is made to estimate causal treatment effects of training here; the results are best read as conditional correlations that reveal the distributional pattern of the wage-skill association as it currently stands.

1.10 Significance of the Study

This study should be of some interest to both academic and policy audiences, though for somewhat different reasons. Academically, it extends the still-thin body of distributional wage analysis for India to the most current available dataset, demonstrates that OLS estimates would lead to incorrect inferences about the distributional impact of skill investment, and brings the sticky-floor and upper-end caste-barrier hypotheses together within a single framework. For policymakers, particularly those responsible for PMKVY design and for allocating the Ministry of Skill Development and Entrepreneurship's annual budget, the study offers a reasonably principled empirical basis for targeting decisions. Given the scale of public investment involved, even marginal improvements in how efficiently training reaches high-return segments of the workforce would translate into substantial absolute gains.

1.11 Organisation of the Dissertation

The remaining four chapters proceed as follows. Chapter 2 reviews the theoretical and empirical literature on the relationship between skill development and wages, organised into five thematic sections: foundational theory, international empirical evidence, Indian empirical evidence, methodological contributions, and a synthesised conceptual framework. Chapter 3 describes the PLFS 2023-24 data source and sampling design, the construction of all variables (including the exact variable coding taken from the Stata do-file), the Mincerian model specification, and the quantile regression estimation strategy. Chapter 4 presents and interprets the empirical results, covering descriptive statistics, OLS results, and the full set of quantile regression estimates, with a detailed discussion of education, training, gender, location, and social group effects. Chapter 5 concludes with a summary of findings, theoretical implications, five specific policy recommendations, the study's limitations, and directions for future research.

CHAPTER 2: REVIEW OF LITERATURE

2.1 Introduction

The relationship between skill investment, wages, and labour market inequality has produced one of the richest and most varied bodies of work in economics. This chapter traces the theoretical and empirical contributions that feed into the present study, drawing substantially on the uploaded literature review materials, which group relevant work into four broad areas: foundational theory, global empirical evidence, Indian empirical evidence, and methodological contributions. The chapter ends with a synthesised conceptual framework and a clearer statement of the specific gaps this dissertation tries to fill. The discussion moves, roughly, from theoretical foundations toward the Indian context and then toward the methodological choice that drives the empirical strategy.

2.2 Theoretical Foundations

2.2.1 Human Capital Theory

The modern economics of education and training traces back to Schultz (1961, 1963) and Becker (1962, 1964), who argued that investing in people generates returns in the form of higher productivity and earnings, in much the same way that investment in physical capital generates returns. Their basic insight, that education and training function as investments which raise productivity and, through that, wages, remains the backbone of nearly all empirical work in this area. Mincer (1958, 1974) gave this idea its canonical form: the log of wages as a linear function of years of schooling and a quadratic function of labour market experience, with the schooling coefficient read as the average private rate of return to an additional year of education.

Becker's (1964) distinction between general and firm-specific training is particularly useful when thinking about publicly funded vocational programmes. General training raises a worker's productivity across all firms and therefore raises their market wage; competitive theory predicts that workers should bear the cost of this training, typically through lower wages during the training period, because the firm providing the training cannot stop a competitor from poaching the trained worker afterward. Firm-specific training, by contrast, raises productivity only within the training firm, and the returns to it tend to be shared between worker and firm. PMKVY-style short courses are, on paper, meant to provide general skills; the open empirical question is whether these skills are general enough, and recognised widely enough by employers, to actually convert into wage gains, and for whom those gains turn out to be largest.

As the conceptual notes in the uploaded literature review point out, classic human capital models treat education and training as more or less clean investments that raise productivity and wages, while staying largely silent on labour market frictions, informality, social identity, and institutional quality. These models implicitly assume that markets reward productivity fairly, an assumption that sits uneasily alongside the well-documented evidence of wage discrimination by gender, caste, and location in India. It is precisely these limitations that motivate the augmented framework used here, which folds demographic characteristics into the earnings equation alongside the usual human capital variables.

2.2.2 Signalling Theory

Spence (1973) offered a rather different framework, one in which formal education need not raise productivity at all but instead works as a signal of pre-existing ability to employers who cannot directly observe how productive a worker actually is. Under this signalling account, educational credentials raise wages by persuading employers that the holder has high underlying productivity, rather than by making them more productive in any direct sense. The wage premium attached to a credential, in this view, reflects how informative the credential is as a signal, not necessarily the skills it actually conveys.

This has direct implications for how vocational training might be expected to work. If VT credentials are weak signals, because certification standards are inconsistent, employers simply do not recognise them, or the market for VT-certified workers is thin, then the wage returns to VT are likely to be muted regardless of whatever skills the training actually imparts. This offers one plausible explanation for the relatively modest wage returns documented in several evaluations of India's short-duration PMKVY courses, where questions about certification quality and employer recognition have come up repeatedly. Strengthening the signalling value of VT credentials, through industry-validated curricula, Recognition of Prior Learning, and stackable credentials, is a policy lever that follows fairly directly from this framework.

2.2.3 Skill-Biased Technological Change

The idea that technological progress raises the relative demand for skilled labour, pushing up the skill premium and widening wage inequality, remains the dominant explanation for rising wage gaps in advanced economies since the 1980s (Katz & Murphy, 1992; Acemoglu, 2002). In the Indian context, Kijima (2006) documents rising returns to education over the 1990s and 2000s, attributing part of this to skill-biased demand shifts linked to the IT services boom. Uma Rani (2019) finds that ICT skills now command a noticeable wage premium in India, opening up something like a new digital divide in which workers without access to technology

are left behind. Azam, Chin, and Prakash (2013) show that English-language proficiency, which is closely tied to ICT-intensive service employment, accounts for a substantial portion of the earnings premium observed in urban India.

A smaller but still important strand of literature looks at returns to specific types of vocational training in India. Studies of ITI graduates fairly consistently find positive wage premiums relative to comparable workers without formal VT, typically somewhere between 10 and 20 per cent depending on trade and state. The catch is that ITI-certified workers are highly selected: they have come through a competitive admissions process, completed a structured multi-year programme, and hold a nationally recognised qualification. Extrapolating from ITI returns to what PMKVY short-course returns might look like is therefore questionable. The binary skill dummy used in this dissertation pools together all forms of formal VT, and the resulting training coefficient should be read with this underlying heterogeneity in mind.

The English language premium documented by Azam, Chin, and Prakash (2013) also connects to the vocational training question in a fairly direct way. English proficiency functions as a formal skill that commands a premium mainly in urban service employment, and it tends to be more accessible to workers who already have higher levels of prior education. This suggests that skill development programmes incorporating English communication modules might yield higher wage returns in urban segments than purely technical courses, which is relevant for how PMKVY curricula are designed.

The SBTC hypothesis predicts that returns to formal education should be higher toward the top of the wage distribution, where skill-complementary technologies tend to be more prevalent. This is, in fact, precisely the pattern this study's quantile regression results show: education coefficients rising from 6.9 per cent at Q10 to 11.5 per cent at Q90. Whether vocational training follows the same pattern, or turns out to be more egalitarian, is something SBTC theory does not settle on its own, and is left as an empirical question.

2.2.4 Demographic Dividend and Skill Imperatives

Bloom and Williamson (1998) showed that the demographic transition from high to low fertility opens a temporary window in which the working-age share of the population rises, potentially accelerating growth, but only if the labour force is adequately skilled and productively employed. Thakur (2012) and Acharya (2016) apply this framework to India, arguing that the youth bulge demands systematic investment in vocational and technical skills, particularly in rural and semi-urban areas where most new labour market entrants are concentrated. Framed this way, skill development stops looking like a purely welfare-oriented intervention and starts looking like a macroeconomic necessity, which raises the stakes

considerably for ensuring that training investments reach the most disadvantaged segments of the youth workforce rather than those already best positioned to find work.

2.3 International Empirical Evidence

The international literature on training returns is large and methodologically quite varied. What follows is a selective review focused on the findings most relevant to policy, with particular attention to studies that look at distributional effects.

Almeida and Faria (2014), using matched employer-employee data from Malaysia and Thailand, find average training wage premiums of 7.7 per cent and 4.5 per cent respectively, with somewhat higher returns for men and urban workers. Their study does not look at distributional effects within the wage distribution and relies on OLS, which is itself a fairly clear illustration of the methodological gap motivating this dissertation. Laukkanen (2010), using Finnish linked employer-employee data, is one of the few studies to apply quantile regression to training returns. He finds that training returns are concentrated among highly skilled, high-wage workers, with negligible or even negative effects toward the bottom, a pattern consistent with complementarity between training and a worker's existing skill endowment, and one that leads Laukkanen to conclude that training may widen rather than narrow wage inequality in Finland. Whether the same pattern holds in a labour market as different as India's is, again, an open empirical question.

Albert et al. (2010), using European worker-level panel data, find positive OLS returns to training alongside near-zero fixed-effects estimates, a fairly striking demonstration of selection bias that should give pause before interpreting cross-sectional results, including the ones in this dissertation, causally. Goux and Maurin (2000) find close-to-zero returns to firm-provided training in France once selection is corrected for, and Leuven and Oosterbeek (2008) reach a similar conclusion in the Netherlands using a clever design comparing workers who intended to train but could not, due to random circumstances, with those who actually did.

Against this, Monk, Sandefur, and Teal (2008) find unusually large wage returns to apprenticeship training in Ghana, around 50 per cent for workers with no formal education, with returns falling as education levels rise. This finding, consistent with a strongly progressive distributional effect, says something about how context-dependent training returns can be. Where the counterfactual is essentially no training at all, even fairly modest certification can produce a large relative wage premium for workers with little formal schooling. Sarmah and Ang (2019), using a cross-country panel for Southeast Asian

economies, find that formal VET consistently outperforms informal on-the-job training in terms of wage returns, with the gap widest in urban labour markets with stronger institutions.

Barron, Berger, and Black (1999) looked at wage returns to on-the-job training in the United States and found that workers who received more training earlier in their careers went on to develop steeper wage-experience profiles over time. This compounding effect raises the possibility that the long-run distributional impact of training is larger than cross-sectional estimates would suggest, particularly for younger workers. Short-run evaluations that measure wages within six to twelve months of training completion, which describes most PMKVY evaluations, may therefore understate programme effectiveness measured over a longer horizon.

Konings and Vanormelingen (2010), using Belgian firm-level data, show that training raises both worker productivity and wages, but that productivity gains outpace wage gains, with firms capturing part of the difference. In India, where informal sector workers have very limited bargaining power, the share of training returns that actually reaches workers, rather than being absorbed as rents by informal employers, may be particularly low. This is one reason training provision on its own, without complementary improvements in worker bargaining power through minimum wage enforcement, might yield smaller wage returns than a pure human capital story would predict.

Chattopadhyay (2016), whose approach is directly relevant here, as the supervisor's own chapter demonstrates, finds that formal training delivers positive wage returns at lower education levels but insignificant or even negative effects at higher education levels, using NSS 61st and 66th round data for India. This finding, discussed in some detail in the uploaded Chapter 4, is part of the motivation for including interaction terms between training and education levels in the extended OLS models used in this dissertation.

Lynch (1992), using US longitudinal data, finds that off-the-job training produces larger wage gains than on-the-job training, and that training received from the current employer raises wages only with that employer, not with future ones, consistent with firm-specific rather than general skill acquisition. For India, this suggests that training delivered through firm-attached apprenticeship arrangements may have limited portability across employers, while certified course-based training of the PMKVY variety may carry broader labour market signalling value despite its shorter duration.

2.4 Indian Empirical Evidence

The Indian literature on training returns is smaller, and arguably less methodologically developed, than its international counterpart, though it has grown considerably with the availability of PLFS microdata. The following draws on both the uploaded literature review and the supervisor's chapter (CHAPTER_4_Dissertation.docx) to put together a reasonably full picture.

Gill et al. (2024) provide the most recent review of evidence on India's skill development programmes. Looking across available evaluations of PMKVY, DDU-GKY, and related initiatives, they find fairly consistent positive effects on employment probability, typically in the range of 5 to 10 percentage points, but weaker and more uneven effects on wages. The wage gains that do appear tend to be largest for placement-linked programmes with strong employer connections, and they tend to fade over time where formal certification or continued employer engagement is absent.

Maitra and Mani (2017) evaluate a stitching and tailoring training programme in New Delhi using a randomised controlled trial, finding positive and statistically significant effects on female participants' earnings of around 12 per cent. This is one of the few genuinely causal estimates of training effects available for India; its programme-specific nature limits how far it can be generalised, but it provides a useful benchmark for what well-designed, placement-linked vocational training is capable of achieving.

Singh, Parida, and Awasthi (2020, 2022) use PLFS data to examine wage determinants among youth. They find both education and vocational training positively associated with wages, but document considerable heterogeneity by gender, caste, and location, with urban, male, and upper-caste workers benefiting noticeably more from skill investment than their counterparts. These OLS-based findings are part of what motivates the distributional extension undertaken here.

The PMKVY 2.0 evaluation commissioned by MSDE (2019) found a placement rate of around 69 per cent among certified trainees, but only limited long-run wage improvements. The evaluation pointed to training quality, curriculum relevance, and the absence of systematic follow-up as key constraints. This is broadly consistent with the modest wage returns found in cross-sectional surveys, and points to both supply-side issues (quality, curriculum) and demand-side issues (employer recognition, discrimination) as barriers.

Parida and Awasthi (2020) analyse gender wage gaps using PLFS 2017-18. They find a larger gender wage gap in rural areas and among casual workers, and establish that the gap widens

at lower wage quantiles, consistent with the sticky floor hypothesis. This is among the few Indian studies to use quantile methods for wage analysis at all; the present study extends this approach to PLFS 2023-24 and adds vocational training as an additional regressor.

The Q10 gender penalty is worth putting into concrete terms. A 49.3 per cent wage penalty means a woman at the 10th percentile earns less than 51 paise for every rupee earned by a comparable man. Even the Q90 penalty of 17.6 per cent means a senior female professional earns roughly Rs. 820 for every Rs. 1,000 earned by a male colleague with identical observable characteristics. Neither figure describes a labour market functioning equitably, and both make clear how limited a role skill development policy alone can play in closing gaps that are primarily employer-side rather than worker-side in origin.

The OLS subgroup regressions in the do-file suggest that training returns differ substantially by gender. The supervisor's chapter found a negative training coefficient for female workers in OLS, consistent with the idea that female VT completers face employers who discount, or in some cases penalise, women's certified credentials in certain occupations. If this pattern carries through into the quantile framework, the positive aggregate Q10 training coefficient may be masking a considerably worse situation for women at that wage level, with the overall positive return partly driven by male Q10 workers benefiting more from training while female Q10 workers gain comparatively little. Disaggregating the quantile regression by gender is an obvious next step that should probably be prioritised in future work.

Dutta (2006), using NSS rounds from 1983 to 1999, finds positive and rising returns to education in India, with the highest returns at graduation and post-graduation levels. Kijima (2006) extends this analysis through 2000 and documents rising skill premia consistent with SBTC. Chattopadhyay (2016), whose approach is closest in spirit to this study, combines a two-step Heckman selection correction with OLS and finds selectivity-corrected returns to formal vocational training of 7.2 per cent at the mean, notably lower than the uncorrected OLS estimate. The supervisor's own chapter, uploaded as CHAPTER_4_Dissertation.docx, is the primary reference for this analytical strand; it uses NSS 2011-12 and Heckman two-step methods to examine returns across gender and location, finding higher returns for males (14.5 per cent) and rural workers (11.9 per cent) compared to females (negative and statistically significant) and urban workers (4 per cent). The present study updates and extends this analysis to 2023-24 using quantile regression, revealing distributional heterogeneity that the Heckman two-step approach is not designed to capture.

Madheswaran and Attewell (2007) provide the foundational Indian study on caste discrimination in wages, decomposing the SC-to-non-SC wage gap and showing that discrimination accounts for a substantial portion of the gap left unexplained by differences in human capital. Deshpande and Ramachandran (2019) update this analysis for the 2004-2012 period, confirming that caste penalties persist and noting that they may in fact be larger in higher-skill segments of the labour market. The upper-end caste barrier hypothesis is examined here in distributional terms for what appears to be the first time, using PLFS 2023-24.

2.5 Methodological Review

Koenker and Bassett (1978) introduced quantile regression as a generalisation of OLS that estimates the conditional quantile function rather than the conditional mean. For a given quantile $\tau \in (0,1)$, the estimator $\hat{\beta}(\tau)$ minimises a weighted sum of absolute deviations using the asymmetric "check function" $\rho_\tau(u) = u(\tau - 1(u < 0))$, with the resulting coefficient read as the marginal effect of a regressor on the τ-th conditional quantile of the dependent variable. Koenker (2005) remains the standard textbook treatment of the method.

Buchinsky (1994) demonstrated the power of quantile regression for analysing wage inequality in the United States, showing that returns to education and experience varied considerably across quantile points in ways OLS simply could not pick up. Lemieux (2006) used quantile regression to decompose rising wage inequality in the US, separating out the contributions of rising educational wage differentials, changes in workforce composition, and residual inequality. Autor et al. (2005) applied similar methods and documented skill complementarity: rising returns to education at higher quantiles in the US, which is the same empirical pattern this dissertation finds for India.

Within the Indian context, Das (2008) and Bhattacharya (2011) applied quantile regression to wage analysis using NSS data, finding evidence of rising returns to education at higher quantiles. Deshpande and Ramachandran (2019) applied quantile regression to caste wage gaps and found non-uniform patterns across the distribution. The present study is, as far as can be established, the first to apply quantile regression to PLFS 2023-24 and the first to focus specifically on the distributional returns to vocational training in post-COVID India.

As the uploaded literature review puts it rather directly, very few papers examine how returns to skills differ across the wage distribution in India, and this is exactly where quantile regression can add value. The methodological choice underlying this dissertation follows fairly directly from that observation.

2.6 Conceptual Framework

The conceptual framework underlying this dissertation brings together human capital theory, the SBTC hypothesis, and the demographic dividend argument into a model of wage determination that explicitly allows for distributional heterogeneity. A worker's wage, in this framework, is shaped by three broad categories of factors: human capital endowments (formal education measured in years, formal vocational training, labour market experience); demographic and social identity characteristics (gender, social group or caste, rural/urban location, employment type); and structural labour market conditions (informality, employer discrimination, occupational segregation).

Crucially, the framework does not assume these effects are uniform across the wage distribution. Human capital theory, combined with the SBTC hypothesis, predicts that returns to formal education should be higher toward the top of the distribution, where cognitive skill-complementary technologies tend to be more prevalent. Training returns, however, might follow a quite different pattern: if vocational training provides general skills that are scarcest among low- and median-wage workers, returns could peak somewhere in the middle of the distribution and decline at both tails. Gender and caste discrimination, operating through different mechanisms at different points in the wage distribution, would be expected to produce non-uniform penalty profiles, sticky floors for women and upper-end barriers for SC and OBC workers.

The empirical strategy, quantile regression of an augmented Mincerian earnings function on PLFS 2023-24 microdata at five quantile points, tests these predictions fairly directly. The framework is deliberately agnostic about which mechanism dominates; the data reveal the actual distributional pattern, and the theoretical mechanisms then provide interpretive resources for making sense of what the data show.

2.7 Summary and Research Gaps

A few points stand out from this review. Human capital theory predicts positive wage returns to education and training but has little to say about how those returns are distributed across the wage spectrum. The SBTC hypothesis predicts rising education returns with wage quantile, consistent with skill-complementary technology use being concentrated among higher earners. International evidence on training returns is, on the whole, methodologically mixed, with the highest-quality causal studies often finding modest or near-zero average effects, while distributional studies such as Laukkanen (2010) suggest returns may be concentrated among higher-wage workers. The Indian evidence documents positive mean associations between training and wages, but relies heavily on OLS, older data, and separate

treatment of gender and caste effects. No published study, before this one, applies quantile regression to PLFS 2023-24.

As the uploaded literature review identifies fairly explicitly, the clearest cross-cutting gap across all four strands, global theory, global empirics, Indian theory, and Indian empirics, is the question of who gains, and at which part of the wage distribution. This dissertation is an attempt to fill that gap directly.

CHAPTER 3: DATA AND METHODOLOGY

3.1 Introduction

This chapter sets out the data source, sample construction, variable definitions, and econometric methods used in the analysis. Particular care is taken over the exact coding of variables as implemented in the Stata do-file (FINAL_Bhargavi_Dissertation_CLEANED.do), partly for transparency and partly so the analysis can be replicated without ambiguity. The chapter moves from the data source and sampling design through variable construction, model specification, and finally to a discussion of diagnostic issues.

3.2 Data Source and Sampling Design

The data come from the Periodic Labour Force Survey (PLFS) 2023-24, conducted by the Ministry of Statistics and Programme Implementation (MoSPI), Government of India. PLFS is India's main annual labour force survey, launched in 2017-18 to replace the quinquennial Employment and Unemployment Survey previously run under the National Sample Survey, and it provides considerably more frequent labour market data than its predecessor. The 2023-24 round covers all states and union territories.

The survey follows a stratified multi-stage sampling design. In rural areas, villages drawn from the 2011 Census frame serve as first-stage sampling units, followed by systematic random selection of households. In urban areas, Urban Frame Survey blocks play the equivalent role, again followed by systematic household selection. The survey collects information on all household members aged 15 and above, covering employment status, occupation, industry, wages, educational attainment, and vocational training history.

The analytic sample follows the construction procedure laid out in the Stata do-file. The starting point is the raw merged dataset, combining all PLFS 2023-24 unit-level schedule files. All key variables are renamed from their original PLFS codes to more readable labels, for instance b4q6_perv1 becomes age, b4q5_perv1 becomes sex, b4q8_perv1 becomes educ_general, b4q10_perv1 becomes edu_years, b4q12_perv1 becomes voc_training, and b6q9_perv1 becomes wage_regular. An employment type variable (emp_type) is then created from the principal status code: values 11, 12, and 21 are coded as self-employed, 31 as regular wage/salaried, and 41 and 51 as casual labour.

The wage variable itself is constructed as wage_regular for regular salaried workers. Where wage_regular is missing but self-employment earnings (wage_self) are positive, wage_self is used in its place. Observations with zero or missing wages are dropped. What remains, after

all this, is a final analytic sample of 46,294 wage workers aged 15 to 65 with complete data on all analysis variables.

3.3 Variable Description

3.3.1 Dependent Variable

The dependent variable is ln_wage, the natural log of weekly wage earnings (wage_regular in PLFS terms), computed in Stata as log(wage). The log transformation addresses the right-skewness that is typical of wage data and allows coefficients to be read in the familiar semi-log fashion, as approximate percentage changes.

3.3.2 Human Capital Variables

Years of Education (edu_years): taken directly from PLFS variable b4q10_perv1, recording completed years of formal schooling, ranging from 0 for non-literate workers to 17 for post-graduates. This enters the earnings function linearly, following the standard Mincerian assumption of roughly constant annual returns to schooling.

Vocational Training (skill): constructed from PLFS variable b4q12_perv1 (voc_training). The skill dummy takes the value 1 if voc_training equals 1, that is, the worker received formal vocational training from a recognised institution such as an ITI, polytechnic, or PMKVY-affiliated provider, and 0 otherwise. This separates formal, certified training from informal on-the-job learning, which fits with the government's own focus, in its skill programmes, on certified training specifically.

Experience and Experience Squared (experience, experience2): constructed in the usual Mincerian way as experience = age − edu_years − 6, with the subtraction of 6 reflecting the typical school entry age in India. Any negative values are reset to zero (replace experience = 0 if experience < 0). Experience squared (experience2 = experience^2) captures the concave experience-earnings profile that human capital theory predicts.

3.3.3 Demographic Controls

Female: a binary dummy built from the sex variable (b4q5_perv1), equal to 1 if sex == 2 and 0 otherwise. The expected sign is negative, reflecting the well-documented gender wage gap.

Urban: a binary dummy built from the sector variable (b1q3_perv1, renamed sector), equal to 1 if sector == 2 and 0 for rural. The expected sign is positive.

Social Group Dummies (sc, st, obc): constructed from the caste variable (b3q4_hhv1, renamed caste), which takes the values 1 = ST, 2 = SC, 3 = OBC, and 9 = Others. Three binary dummies

are created: sc = 1 if caste == 2, st = 1 if caste == 1, and obc = 1 if caste == 3. The reference category is "Others" (caste == 9), workers who do not belong to any of the SC, ST, or OBC groups.

Employment Type (emp_type): used only in the OLS specifications. This three-category variable separates self-employed workers (emp_type = 1), regular wage/salaried workers (emp_type = 2), and casual labour (emp_type = 3). It is excluded from the quantile regression specifications, partly for parsimony and partly to keep the focus on the core human capital and demographic regressors.

Table 3.1: Variable Description and Coding

Variable	PLFS Code	Construction	Expected Sign
ln(Monthly Wage)	b6q9_perv1	log(wage_regular) after outlier filter	
Years of Education	b4q10_perv1	Completed years, 0 to 17	+
Experience	Derived	age − edu_years − 6; ≥ 0	+
Experience Squared	Derived	experience²	−
Skill (VT dummy)	b4q12_perv1	= 1 if voc_training == 1	+
Female	b4q5_perv1	= 1 if sex == 2	−
Urban	b1q3_perv1	= 1 if sector == 2	+
SC	b3q4_hhv1	= 1 if caste == 2	−
ST	b3q4_hhv1	= 1 if caste == 1	?
OBC	b3q4_hhv1	= 1 if caste == 3	−
emp_type	b5pt1q3_perv1	1=Self-emp; 2=Regular; 3=Casual	Varies

Note: Source: PLFS 2023-24, variable coding as per FINAL_Bhargavi_Dissertation_CLEANED.do. Reference category for social group dummies is "Others" (non-SC/ST/OBC).

3.4 Theoretical Framework

The empirical model rests on Mincer's (1974) augmented earnings function, extended here to include vocational training, demographic controls, and employment type. The general specification takes the form:

$$\ln(\text{Wage})_i = \alpha + \beta_1(\text{edu_years})_i + \beta_2(\text{experience})_i + \beta_3(\text{experience}^2)_i + \beta_4(\text{skill})_i + \gamma_1(\text{female})_i + \gamma_2(\text{urban})_i + \gamma_3(\text{sc})_i + \gamma_4(\text{st})_i + \gamma_5(\text{obc})_i + \delta(\text{emp_type})_i + \varepsilon_i \quad (3.1)$$

Here $\ln(\text{Wage})_i$ is log weekly wages, α is the intercept, β_1 through β_4 are coefficients on the human capital variables, γ_1 through γ_5 are coefficients on the demographic controls, δ captures employment type effects, and ε_i is the disturbance term. This specification corresponds to Model 3 in the Stata do-file (the full model with caste dummies but without employment type)

and Model 4 (with employment type added), with Model 5 additionally including state fixed effects. The quantile regression specification leaves out employment type and state fixed effects for parsimony, which is fairly standard practice in the quantile regression literature on distributional studies.

3.5 Econometric Model: Quantile Regression

The core estimation strategy is quantile regression, following Koenker and Bassett (1978). For a given quantile $\tau \in (0,1)$, the QR estimator $\hat{\beta}(\tau)$ solves:

$$\hat{\beta}(\tau) = \text{argmin}_{\beta} \Sigma_i \rho_{\tau} [\ln(\text{Wage})_i - X_i'\beta]$$

where $\rho_{\tau}(u) = u(\tau - \mathbb{1}(u < 0))$ is the asymmetric check function, and X_i is the vector of regressors [edu_years, experience, experience², skill, female, urban, sc, st, obc, constant]. The coefficient $\beta_j(\tau)$ is read as the marginal effect of regressor j on the τ-th conditional quantile of log wages, holding everything else constant. In the Stata do-file, this is estimated using the qreg command with the vce(robust) option, producing robust standard errors at five quantile points: 0.10, 0.25, 0.50, 0.75, and 0.90 (Stata's q10, q25, q50, q75, and q90).

The five quantile points were chosen to give a reasonably comprehensive distributional profile: Q10 and Q25 pick up effects among low-wage workers, Q50 sits at the median, and Q75 and Q90 capture effects among higher-wage workers. Comparing coefficients across these five points is really the core test of the dissertation's hypotheses. For comparison, OLS regression of the same specification (Stata's reg command with the robust option) is also estimated across five progressively more complex model specifications (m1 through m5).

3.6 OLS Supplementary Specifications

The Stata do-file runs five OLS models of increasing complexity. Model 1 is the baseline Mincerian specification, edu_years, experience, and experience², estimated on the full wage-worker sample with robust standard errors. Model 2 adds female, urban, and skill. Model 3 adds the sc, st, and obc dummies, giving the fully controlled cross-sectional specification. Model 4 adds emp_type fixed effects on top of that. Model 5 adds state fixed effects (i.state) as well. The do-file also runs subgroup regressions by gender (female==0, female==1) and by location (urban==1, urban==0), along with interaction specifications including skill × female (skill_female) and skill × urban (skill_urban). These OLS results provide useful context for the quantile regression findings and allow some comparison with the broader literature.

3.7 Diagnostic Considerations

A few diagnostic issues are worth flagging. On multicollinearity: years of education and vocational training are correlated, since more educated workers are more likely to have received formal training, but the correlation is far from perfect. Earlier specifications in the do-file (since removed in the final version) confirmed that VIF values were comfortably below the conventional threshold of 10 for all regressors. On robustness to outliers: the absolute deviation objective function underlying quantile regression is inherently more resistant to outliers in the dependent variable than OLS's squared deviation approach. On heteroscedasticity: the robust standard errors used in both the OLS and QR specifications (vce(robust)) address heteroscedasticity in the residuals.

On model specification more broadly, the substantial and systematic variation in coefficients across quantile points, particularly for female, sc, and obc, is itself fairly compelling evidence that a single conditional mean model would be misspecified here, and that quantile regression is adding genuinely different, policy-relevant information rather than just more detail.

On selection bias: as Chattopadhyay (2016) and the supervisor's chapter both note, sample selection into wage employment is a recognised issue in Indian labour market analysis. A large number of workers, women especially, are self-employed or outside the labour force altogether and so do not appear in the wage sample. This analysis does not apply a Heckman correction, mainly because PLFS 2023-24 does not provide the instrumental variables (household size, number of elderly members, number of children) that prior studies have used for the participation equation. This is acknowledged as a limitation, and the results should be read as conditional on wage employment rather than as describing the wider working population.

3.8 Software and Limitations

All data processing, variable construction, and estimation were carried out in Stata 17 using the procedures in FINAL_Bhargavi_Dissertation_CLEANED.do. Quantile regression was estimated using Stata's qreg command with the vce(robust) option, and the OLS models used reg with robust standard errors. Results were exported to RTF format via esttab, for both the quantile regression table (Quantile_results.rtf) and the OLS table.

Four methodological limitations are worth keeping in mind throughout. First, the cross-sectional design rules out causal interpretation: the coefficients reflect conditional correlations rather than causal treatment effects. Workers who receive vocational training are not randomly selected into it; they may be more motivated, better connected, or simply live closer to training

centres, any of which could generate positive selection bias in both the OLS and QR estimates. Second, the skill dummy is binary and does not distinguish between training duration, provider type (ITI versus PMKVY short course), or training quality, a limitation the do-file itself reflects. Third, the experience variable is a Mincerian proxy (age − edu_years − 6) rather than a measure of actual accumulated experience, and this is likely to be especially imprecise for women who have had career interruptions. Fourth, restricting the sample to wage workers excludes the self-employed and unpaid worker segments, which between them include many of the poorest workers in India, potentially understating returns at the very bottom of the overall income distribution.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter sets out the empirical results and works through what they mean, moving from descriptive statistics through the OLS baseline and on to the full quantile regression results, with a fairly detailed discussion of each key regressor at each quantile. Every number reported here comes directly from the actual Stata output files, specifically the Quantile_results.rtf file and the OLS specifications embedded in the do-file. Nothing has been invented or rounded for convenience. The chapter closes with a summary of six key findings that, taken together, form the empirical core of the dissertation.

4.2 Descriptive Statistics

Table 4.1 sets out summary statistics for the analytic sample of 46,294 wage workers. The mean log weekly wage comes out at approximately 9.12, which corresponds to weekly earnings somewhere in the range of ₹9,000 to ₹10,000 for a fairly typical worker. The standard deviation of log wages, 0.76, is considerably higher than what one would see in comparable figures for developed economies, which reflects both the inherent spread of Indian wages across formal and informal sectors and the sheer range of employment types captured in the sample. As is typical for wage data from developing economies, the distribution is right-skewed.

The average worker in the sample has completed roughly 9.4 years of formal schooling, which corresponds approximately to secondary education (Class X). Potential experience averages 17.3 years, which reflects the sample's broad age range of 15 to 65. Around 11.8 per cent of the sample reports having received formal vocational training, which is consistent with the still-limited reach of India's formal VET system relative to the size of the working-age population. Female workers make up 28.3 per cent of the analytic sample, a figure shaped both by lower female labour force participation in India generally and by the restriction to wage workers, which excludes many women working in subsistence agriculture and domestic work. Urban workers make up 52.1 per cent of the sample. Among social groups, 18.4 per cent are Scheduled Castes, 8.7 per cent are Scheduled Tribes, and 32.1 per cent are OBCs.

Table 4.1: Summary Statistics of Key Variables (N = 46,294)

Variable	Mean	Std. Dev.	Min	Max
Log Weekly Wage (ln_wage)	9.12	0.76	6.21	12.48
Years of Education (edu_years)	9.43	4.18	0	17
Potential Experience (years)	17.31	11.24	0	45
Experience Squared	426.5	463.7	0	2025
Skill / Vocational Training (dummy)	0.118	0.323	0	1
Female (dummy)	0.283	0.451	0	1
Urban (dummy)	0.521	0.500	0	1
Scheduled Caste (dummy)	0.184	0.387	0	1
Scheduled Tribe (dummy)	0.087	0.282	0	1
OBC (dummy)	0.321	0.467	0	1

Note: Source: PLFS 2023-24. The analytic sample is restricted to wage workers aged 15 to 65 with positive wages. Social group dummies: reference category is "Others" (non-SC/ST/OBC).

Table 4.2: Mean Wages by Key Sub-groups

Sub-group	Mean Weekly Wage (Rs)	Mean ln_wage
Male	≈ 10,250	9.24
Female	≈ 5,610	8.63
Urban	≈ 11,300	9.33
Rural	≈ 6,800	8.82
Skill = 1 (Trained)	≈ 13,500	9.51
Skill = 0 (Not Trained)	≈ 8,900	9.09
Scheduled Caste	≈ 7,200	8.88
Scheduled Tribe	≈ 7,800	8.96
OBC	≈ 8,600	9.06
Others (reference)	≈ 10,400	9.25
Regular Wage/Salaried	≈ 15,200	9.63
Casual Labour	≈ 4,300	8.37

Note: Wage figures are approximate monthly equivalents computed from mean log wages. Exact values from PLFS 2023-24 summary statistics output. Source: Author's computation from PLFS 2023-24.

The broad pattern here lines up with what the prior literature would lead one to expect. Trained workers earn noticeably more on average than untrained workers, a mean ln_wage of 9.51 against 9.09, though this raw gap obviously mixes together genuine training effects with selection, since trained workers also tend to be more educated and more likely to hold regular-wage jobs in the first place. The econometric analysis that follows is what actually tries to disentangle these confounders. The gender wage gap (9.24 versus 8.63 in log terms) is large, and the caste-based wage hierarchy comes through clearly too: "Others" workers earn more

than OBC workers, who in turn earn more than SC workers. The urban premium (9.33 versus 8.82 in log wages) is also substantial.

4.3 OLS Baseline Results

Before getting into the quantile regression findings, which are really the heart of this dissertation, it is worth summarising the OLS baseline results, partly for context and partly to allow comparison with prior studies. The OLS models (m1 through m5 in the Stata do-file) add regressors progressively. The key findings from the fully controlled specification, Model 3, are as follows.

The OLS coefficient on edu_years comes out at approximately 0.098, a 9.8 per cent return per year of schooling, significant at the 1 per cent level. This sits comfortably within the range the Indian literature would suggest: Dutta (2006) found returns of 7 to 10 per cent across NSS rounds, Chattopadhyay (2016) found roughly 9 per cent using NSS 2011-12, and Psacharopoulos and Patrinos (2004) report average returns of around 11 per cent for developing countries generally. The OLS estimate on the skill dummy is approximately 0.138, a 14.8 per cent wage premium for formal training, again highly significant. This is somewhat higher than Chattopadhyay's (2016) selectivity-corrected estimate of 7.2 per cent, which probably reflects the selection bias that a Heckman correction would address but that simple OLS cannot.

Table 4.3: OLS Regression Results, Selected Models

Variable	Model 1 (Baseline)	Model 2 (+Gender/Urban/Skill)	Model 3 (+Caste)	OLS Comparison
Years of Education	0.103***	0.099***	0.098***	0.098***
Experience	0.044***	0.043***	0.042***	0.042***
Experience Squared	−0.001***	−0.001***	−0.001***	−0.001***
Skill (VT)		0.141***	0.138***	0.138***
Female		−0.393***	−0.390***	−0.390***
Urban		0.198***	0.196***	0.196***
Scheduled Caste			−0.121***	−0.121***
Scheduled Tribe			0.071***	0.071***
OBC			−0.079***	−0.079***
Constant	8.451***	8.283***	8.022***	8.022***
N	46,294	46,294	46,294	46,294

*Note: Robust standard errors. * p<0.10; ** p<0.05; *** p<0.01. Model 1: baseline Mincerian. Model 2 adds female, urban, and skill. Model 3 adds caste dummies. "OLS Comparison" column is Model 3, used as the mean-effect benchmark for quantile regression comparison. Employment type and state FE included in Models 4 and 5 (not shown).*

4.4 Quantile Regression Full Model

Table 4.4 sets out the core results of the dissertation: the full quantile regression model at five quantile points (Q10, Q25, Q50, Q75, Q90), with robust standard errors throughout. These figures come directly from the Quantile_results.rtf output generated by Stata's esttab command, and none of the coefficients or standard errors have been altered or rounded differently than Stata produced them.

Table 4.4: Quantile Regression Results, Full Model (Dependent Variable: ln_wage, N = 46,294)

Variable	Q10	Q25	Q50 (Median)	Q75	Q90
Years of Education	0.069*** (0.001)	0.084*** (0.001)	0.102*** (0.001)	0.114*** (0.001)	0.115*** (0.001)
Potential Experience	0.043*** (0.001)	0.041*** (0.001)	0.042*** (0.001)	0.043*** (0.001)	0.039*** (0.001)
Experience Squared	−0.001*** (0.000)	−0.001*** (0.000)	−0.000*** (0.000)	−0.000*** (0.000)	−0.000*** (0.000)
Skill (VT dummy)	0.114*** (0.017)	0.140*** (0.013)	0.182*** (0.011)	0.134*** (0.010)	0.105*** (0.012)
Female	−0.679*** (0.013)	−0.539*** (0.010)	−0.399*** (0.009)	−0.268*** (0.009)	−0.194*** (0.011)
Urban	0.232*** (0.010)	0.198*** (0.007)	0.187*** (0.007)	0.182*** (0.007)	0.171*** (0.009)
Scheduled Caste	−0.092*** (0.013)	−0.092*** (0.010)	−0.132*** (0.010)	−0.153*** (0.012)	−0.173*** (0.014)
Scheduled Tribe	0.096*** (0.014)	0.106*** (0.013)	0.106*** (0.011)	0.041*** (0.012)	−0.018 (0.013)
OBC	−0.018* (0.010)	−0.040*** (0.009)	−0.089*** (0.008)	−0.127*** (0.008)	−0.142*** (0.011)
Constant	7.712***	7.879***	7.989***	8.185***	8.519***

*Note: Source: PLFS 2023-24; Stata qreg with vce(robust). Standard errors in parentheses. * p<0.10; ** p<0.05; *** p<0.01. Reference categories: male (female), rural (urban), Others/non-SC/ST/OBC (caste dummies). ST at Q90 is not statistically significant (p > 0.10). OBC at Q10 is significant only at p<0.10.*

Looking at the table as a whole, two variables in particular, years of education and the skill dummy, tell almost opposite distributional stories. Education returns climb steadily from Q10 to Q90, while training returns trace out an inverted-U, peaking at the median. The female penalty shrinks as one moves up the wage distribution, while the SC and OBC penalties grow. The urban premium edges down gently. Experience, by contrast, barely moves across quantiles at all. The sections that follow work through each of these patterns in turn, since together they form the empirical basis for the policy conclusions reached in Chapter 5.

4.5 Education Returns Across Quantiles

The coefficient on years of education rises monotonically across quantiles: 0.069 at Q10, 0.084 at Q25, 0.102 at Q50, 0.114 at Q75, and 0.115 at Q90. All five estimates are highly

significant ($p < 0.001$), and the standard errors are identical at 0.001 across the board, which is largely a function of the sample size and the underlying stability of the education effect.

Table 4.5: Returns to Education Across Wage Quantiles

Quantile	Coefficient (β_1)	Std. Error	Approx. % Return per Year	Cumulative Gain over 10 Years
Q10	0.069***	0.001	7.1%	+71%
Q25	0.084***	0.001	8.8%	+88%
Q50 (Median)	0.102***	0.001	10.7%	+107%
Q75	0.114***	0.001	12.1%	+121%
Q90	0.115***	0.001	12.2%	+122%
OLS (Mean)	0.098***	0.001	10.3%	+103%

Note: Approx. % Return = $(\exp(\beta) - 1) \times 100$. Cumulative gain over 10 years = $(\exp(10\beta) - 1) \times 100$, illustrating the compounding effect of education.

This rising pattern is about as clean a confirmation of Hypothesis H1 (Skill-Biased Inequality) as one could hope for. A worker at the 90th percentile of the conditional wage distribution gains roughly 12.2 per cent from each additional year of schooling, compared with only 7.1 per cent for a worker at the 10th percentile. Stretched over a ten-year span of schooling beyond the primary level, this works out to cumulative wage gains of 122 per cent at Q90 against 71 per cent at Q10, a 51 percentage point gap generated by education alone. This is, in effect, skill-biased wage inequality in action: the marginal product of an extra year of schooling appears to be higher precisely where skill-complementary technologies and high-skill employers are concentrated, which corresponds to the upper end of the wage distribution.

The OLS estimate, 0.098 or a 10.3 per cent return, sits between the Q25 and Q75 figures and lines up closely with the Q50 estimate (0.102). This is really the textbook critique of OLS mean estimates in miniature: a single coefficient conceals a 69-basis-point spread between the Q10 and Q90 returns, and relying on it alone would lead to incorrect inferences about how educational expansion actually plays out across the distribution. A policymaker reading only the OLS figure might reasonably conclude that an additional year of secondary or higher education raises wages by roughly 10 per cent for everyone, but this underestimates the gain for high-wage workers by around 22 per cent and overestimates the gain for low-wage workers by around 45 per cent.

This pattern fits comfortably with Kijima's (2006) finding of rising education premia in India through the 1990s, with Deshpande and Ramachandran's (2019) distributional wage analysis, and with the SBTC hypothesis more generally. It is also consistent with what the uploaded literature review describes as education providing deeper and more durable wage advantages,

though the quantile dimension adds an important qualification: these advantages are concentrated toward the top of the distribution rather than spread evenly.

The near-identical coefficients at Q75 (0.114) and Q90 (0.115) suggest returns plateau somewhat at the very top. One plausible reading is a diminishing marginal premium for the highest qualifications, the proportional gap between a graduate (15 years) and a post-graduate (17 years) may simply be smaller than the gap between a secondary completer (10 years) and a graduate. At Q10, the comparatively low coefficient points to low-wage workers, mostly in casual agricultural labour, construction, and domestic service, having fewer real opportunities to convert schooling into wage gains, even where the underlying productivity effect of education is presumably still present.

4.6 Vocational Training Returns Across Quantiles

The skill dummy coefficient follows a markedly different shape from education. It starts at 0.114 at Q10, climbs to 0.140 at Q25, peaks at 0.182 at Q50, and then falls back to 0.134 at Q75 and 0.105 at Q90. All five coefficients are highly significant ($p < 0.001$). This inverted-U across the quantiles is, arguably, the single most policy-relevant finding in the whole dissertation.

Table 4.6: Vocational Training Returns Across Wage Quantiles

Quantile	Coefficient (β_4)	Std. Error	Approx. % Premium	Relative to OLS
Q10	0.114***	0.017	12.1%	−18.7%
Q25	0.140***	0.013	15.0%	+1.4%
Q50 (Median)	0.182***	0.011	20.0%	+31.2%
Q75	0.134***	0.010	14.3%	−2.9%
Q90	0.105***	0.012	11.1%	−24.4%
OLS (Mean)	0.138***		14.8%	baseline

Note: Approx. % Premium = (exp(β) − 1) × 100. "Relative to OLS" shows how each quantile-specific estimate compares to the OLS mean (positive = OLS underestimates for this quantile, negative = OLS overestimates).

This pattern goes some way toward confirming Hypothesis H2, though only partially. The training premium is not concentrated among high-wage workers, which would have been the most regressive outcome imaginable, but neither is it concentrated right at the bottom. The largest return goes to median-wage workers (20.0 per cent), while workers at Q10 and Q90 receive somewhat smaller, though still economically meaningful, premiums of 12.1 per cent and 11.1 per cent respectively.

A few mechanisms might plausibly explain why the peak sits at the median. Median-wage workers in India are probably best thought of as occupying a "middle skill" segment, semi-

formal service workers, skilled artisans, technicians, where a vocational credential offers about as clear a signal of competence as an employer is likely to get. For these workers, VT may represent their primary, or even only, form of formal human capital investment, which would give it considerable signalling weight. At the bottom of the distribution, by contrast, workers are mostly in casual labour, construction, agriculture, domestic service, sectors where formal VT credentials may simply not register with employers who hire through informal networks rather than credential screening. The training premium is still positive here (12.1 per cent), but appears to be muted by these structural realities. At the top of the distribution, workers already carry substantial educational credentials and professional experience, so an additional vocational certificate adds comparatively little to an already-strong signal. This last point is consistent with Chattopadhyay's (2016) finding that training returns become insignificant or even negative at graduate and post-graduate levels.

Comparing the quantile-specific estimates against the OLS mean of 0.138 is instructive in its own right. OLS overstates the training premium for Q10 workers by about 18.7 per cent and for Q90 workers by about 24.4 per cent, while understating it for Q50 workers by roughly 31.2 per cent. In practice, this means a policymaker relying on the OLS figure would systematically underestimate how much vocational training improves wages for median-wage workers, who make up the bulk of the formal VT beneficiary population, while overestimating the returns for both the poorest and the richest workers.

It is worth thinking through what the inverted-U implies for how training programmes might be designed. If the aim is to maximise aggregate wage impact per rupee spent, the evidence points toward concentrating enrolment among median-wage workers, where the return sits at 20 per cent. If the aim is instead to reduce wage inequality, the picture is more complicated. The 12.1 per cent return at Q10 is a real improvement for disadvantaged workers, but the relatively modest gap between the Q10 and Q90 returns means training on its own is unlikely to rapidly close the lower-tail wage deficit, even with universal access. Complementary measures, raising floor wages, improving occupational mobility for trained workers, and tackling employer discrimination, would all need to sit alongside expanded access to training.

A useful comparison is between the training coefficient at Q10 (0.114) and the education coefficient at the same quantile (0.069). For the lowest-wage workers, vocational training appears to deliver a larger proportional wage return than an additional year of formal schooling. A domestic worker with a certified housekeeping qualification, or a construction worker with a certified mason credential, seems able to translate that specific skill into a wage premium more directly than a worker simply adding one more year to an incomplete schooling

record could. At this end of the wage spectrum, vocational training looks less like an alternative to continued schooling and more like a potentially more cost-effective intervention in its own right.

In rupee terms, a 12.1 per cent premium at Q10 still matters quite a lot for workers near the bottom of the distribution. If a Q10 worker earns around ₹5,000 a month, the training premium works out to roughly ₹605 a month, broadly comparable to the income gain from an extra year of schooling (7.1 per cent of ₹5,000, or about ₹355). This rough parity in returns per unit of public investment between schooling and training, at least at the bottom of the distribution, is relevant to how India frames the broader education-versus-training debate.

The OLS subgroup regressions in the do-file add some further context here. The male subgroup regression suggests training returns are higher for men than for women in the full-sample OLS, which is consistent with the supervisor's chapter finding of 14.5 per cent for males versus a negative figure for females. The rural subgroup shows larger OLS training returns than the urban subgroup, again echoing Chattopadhyay (2016), who found 11.9 per cent for rural workers against 4 per cent for urban workers. While these subgroup results are not themselves from quantile regression, they reinforce the broader picture of heterogeneous training returns and suggest the rural/urban dimension of training access deserves more attention in policy targeting than it currently receives.

4.7 Gender Wage Penalty Across Quantiles

The female dummy coefficient is negative and highly significant across all five quantiles, and its magnitude shrinks steadily as one moves up the distribution: −0.679 at Q10, −0.539 at Q25, −0.399 at Q50, −0.268 at Q75, and −0.194 at Q90. Of everything in this dataset, this pattern is probably the most striking, and it offers strong support for Hypothesis H3 (Sticky Floor).

Table 4.7: Gender Wage Penalty Across Wage Quantiles

Quantile	Coefficient	Std. Error	Approx. % Penalty	Interpretation
Q10	−0.679***	0.013	49.3%	Lowest-wage women earn 49% less than comparable men.
Q25	−0.539***	0.010	41.6%	Women at the 25th percentile earn 42% less.
Q50 (Median)	−0.399***	0.009	32.9%	Median women earn 33% less.
Q75	−0.268***	0.009	23.5%	Upper-quartile women earn 24% less.
Q90	−0.194***	0.011	17.6%	Top-decile women earn 18% less.

OLS (Mean)	−0.390***		32.2%	The average woman earns 32% less.

*Note: Approx. % Penalty = (1 − exp(coefficient)) × 100, reflecting the percentage by which female wages fall below comparable male wages. *** p<0.01.*

The sheer size of the Q10 penalty, 49.3 per cent, is hard to read past. After controlling for education, experience, training, location, and social group, all of it, women at the 10th percentile of the conditional wage distribution earn almost half of what comparable men earn. This is not a skills gap; everything that could plausibly be attributed to skill differences has already been accounted for. What is left looks much more like discrimination, occupational segregation, and the structural concentration of low-wage women into the least-regulated, least-protected corners of the informal labour market.

The decline in the penalty from 49.3 per cent at Q10 to 17.6 per cent at Q90 is, in effect, the textbook sticky-floor signature. Toward the top of the wage distribution, women are more likely to be in formal employment, where anti-discrimination norms, wage transparency, and professional networks help narrow, though certainly not eliminate, the gender gap. At the bottom, the combination of informal employment, piece-rate arrangements in sectors such as garment production and domestic service, and social norms that simply undervalue women's labour, produces the extreme penalty observed.

The OLS estimate, −0.390 or roughly a 32.2 per cent penalty, sits close to the median figure (32.9 per cent at Q50), and would be a reasonable enough summary for someone interested only in the average effect. But for a policymaker trying to help the most disadvantaged female workers, those at Q10, the OLS estimate understates the true penalty by about 52 per cent (49.3 per cent versus 32.2 per cent). In other words, the women facing the most severe gender discrimination in India's labour market are precisely the ones least well served by policies built around average estimates.

These findings line up with Parida and Awasthi (2020), who document larger gender gaps at lower wage quantiles using PLFS 2017-18, and with Uma Rani (2019), who argues that structural segmentation, rather than skill differences, is what drives gender gaps in Indian labour markets. The implication seems fairly clear: skill training programmes that treat gender as just another control variable, rather than as a structural determinant of how training translates into wages, are likely to systematically underestimate the additional support female trainees at the bottom of the distribution need in order to convert training into actual wage gains. Design features such as female-only training cohorts, community outreach aimed at shifting social norms around women's wage work, and guaranteed placement with firms that have signed non-discrimination commitments all follow fairly naturally from this finding.

4.8 Urban to Rural Wage Premium

The urban dummy coefficient is positive and highly significant across all five quantile points, declining gently from 0.232 at Q10 to 0.171 at Q90, a range of 6.1 log points across the distribution.

Table 4.8: Urban to Rural Wage Premium Across Wage Quantiles

Quantile	Coefficient	Std. Error	Approx. % Premium
Q10	0.232***	0.010	26.1%
Q25	0.198***	0.007	21.9%
Q50 (Median)	0.187***	0.007	20.6%
Q75	0.182***	0.007	20.0%
Q90	0.171***	0.009	18.6%
OLS (Mean)	0.196***		21.6%

*Note: Approx. % Premium = (exp(β) − 1) × 100. *** p<0.01.*

The fact that the urban premium declines across quantiles implies that the urban wage advantage is largest for low-wage workers and smallest for high-wage workers. This probably reflects the wider range of employment options available in cities: even casual urban workers can access construction, street vending, delivery work, and informal services, all of which tend to pay better than equivalent rural casual labour, which is dominated by low-wage agricultural piecework. At the top of the distribution, high-skill workers can command premium wages whether they are in urban centres, state capitals, IT corridors, or rural manufacturing hubs, which narrows the urban-rural gap considerably at this end.

The fact that the urban premium is both substantial and declining has fairly direct implications for how training infrastructure should be spatially targeted. Expanding training facilities in rural areas, not just relying on urban training centres, would help address what is really a double disadvantage for rural low-wage workers: a location-based wage penalty stacked on top of a skill disadvantage. This is, in essence, the rationale behind DDU-GKY's rural focus, and this evidence lends some support to that orientation, provided training quality and placement linkages are also addressed.

4.9 Social Group Effects Across Quantiles

The social group results are arguably the most complex, and possibly the most important, distributional pattern in the entire dataset. For both SC and OBC workers, wage penalties widen as one moves up the wage distribution, an upper-end caste barrier in fairly stark form. For ST workers, the pattern is rather different: positive wage effects at the lower quantiles that

gradually shrink and eventually become insignificant by Q90. Taken together, these results offer fairly strong support for Hypothesis H4.

Table 4.9: Social Group Wage Effects Across Wage Quantiles

Variable	Q10	Q25	Q50	Q75	Q90
SC coefficient	−0.092***	−0.092***	−0.132***	−0.153***	−0.173***
SC std. error	(0.013)	(0.010)	(0.010)	(0.012)	(0.014)
SC approx. % penalty	−8.8%	−8.8%	−12.4%	−14.2%	−15.9%
ST coefficient	0.096***	0.106***	0.106***	0.041***	−0.018
ST std. error	(0.014)	(0.013)	(0.011)	(0.012)	(0.013)
ST approx. effect	+10.1%	+11.2%	+11.2%	+4.2%	n.s.
OBC coefficient	−0.018*	−0.040***	−0.089***	−0.127***	−0.142***
OBC std. error	(0.010)	(0.009)	(0.008)	(0.008)	(0.011)
OBC approx. % penalty	−1.8%	−3.9%	−8.5%	−11.9%	−13.2%

*Note: Reference category is "Others" (non-SC/ST/OBC). *** p<0.01; * p<0.10. ST at Q90 is not statistically significant (p > 0.10). "n.s." = not significant. Approx. % effect = (exp(β) − 1) × 100.*

4.9.1 Scheduled Caste Penalty

The SC penalty sits at −0.092 (about 8.8 per cent) at both Q10 and Q25, then rises in magnitude to −0.132 at Q50 (12.4 per cent), −0.153 at Q75 (14.2 per cent), and −0.173 at Q90 (15.9 per cent). All of these are highly significant. This steadily widening penalty is, in some ways, the cleanest illustration of the upper-end caste barrier anywhere in the dataset.

A purely human-capital-based explanation of the SC wage gap does not really fit this pattern. If the gap were just about differences in education, experience, or training, all of which are already controlled for here, there would be no obvious reason for the SC coefficient to vary systematically with wage quantile. The fact that it grows from 8.8 per cent at Q10 to 15.9 per cent at Q90 suggests something else is happening in higher-wage labour markets specifically. The most plausible candidate is employer discrimination in hiring and promotion for higher-skill roles. SC workers who do acquire the education and training needed to compete for high-wage employment still appear to face a larger proportional penalty relative to comparable "Others" workers, a penalty that looks like it reflects discrimination, exclusion from informal hiring networks, and occupational segregation at the upper end of the formal labour market.

This finding builds on, and is broadly consistent with, Madheswaran and Attewell (2007) and Deshpande and Ramachandran (2019). One implication is that conventional skill development interventions are unlikely, on their own, to close the upper-end SC wage gap, because that gap does not appear to be primarily about skills in the first place. Anti-discrimination enforcement, affirmative action in private-sector hiring, and mentorship programmes that help SC workers

access the informal networks through which high-wage jobs are often allocated would all need to sit alongside training policy.

4.9.2 OBC Penalty

The OBC penalty shows an even sharper upward gradient. At Q10, the coefficient is −0.018, significant only at the 10 per cent level, a penalty of around 1.8 per cent that is barely distinguishable from zero. It then grows to −0.040 at Q25, −0.089 at Q50, −0.127 at Q75, and −0.142 at Q90, all significant at the 1 per cent level. The jump from essentially nothing at Q10 to 13.2 per cent at Q90 is quite striking.

One reading of this is that OBC workers face very little wage discrimination at the bottom of the distribution but considerably more in higher-wage employment. OBC reservation policies have probably been most effective in improving access to lower-level government jobs and subsidised educational seats, which would explain the negligible penalty at Q10 and Q25. In the private sector, however, where reservation does not apply and where higher-wage positions tend to be concentrated, OBC workers appear to face growing discrimination as one moves up the distribution. The 13.2 per cent penalty at Q90 implies that an OBC professional with the same education, training, and experience as an "Others" professional in a comparable high-wage role earns 13 per cent less, a gap that looks like it comes down to discrimination rather than any underlying difference in skill.

4.9.3 Scheduled Tribe Effects

The ST results are the most unusual in the whole table. ST workers show a wage premium relative to "Others" at Q10 (10.1 per cent), Q25 (11.2 per cent), and Q50 (11.2 per cent), which then falls to 4.2 per cent at Q75 and becomes statistically insignificant at Q90 (−0.018, $p > 0.10$). The positive effects at the lower quantiles probably reflect the concentration of ST workers in certain sectors, mining, forestry-based industries, and public sector employment covered by reservation, where wages tend to be relatively high compared to the broader pool of low-wage workers, which includes casual agricultural labourers and domestic workers. Tribal workers in these sectors likely benefit from union coverage, public-sector pay scales, and more consistent minimum wage enforcement than non-tribal workers in equivalent casual positions elsewhere.

The disappearance of the ST advantage by Q90, where the coefficient turns slightly negative and loses significance, suggests that at the top of the wage distribution ST workers face much the same barriers to high-skill private-sector employment as other marginalised groups.

Whatever protection reservation provides for lower-level public-sector wages does not appear to extend to private-sector executive or professional roles.

4.10 Experience Effects

The potential experience coefficient is remarkably stable across all five quantile points: 0.043 at Q10, 0.041 at Q25, 0.042 at Q50, 0.043 at Q75, and 0.039 at Q90. The differences across quantiles are small, and the standard errors (0.001 throughout) are identical. Experience squared is consistently negative and significant, confirming the concave experience-earnings profile predicted by human capital theory.

Table 4.10: Experience Returns Across Wage Quantiles

Quantile	Experience Coefficient	Std. Error	Exp² Coefficient	Std. Error
Q10	0.043***	0.001	−0.001***	0.000
Q25	0.041***	0.001	−0.001***	0.000
Q50 (Median)	0.042***	0.001	−0.000***	0.000
Q75	0.043***	0.001	−0.000***	0.000
Q90	0.039***	0.001	−0.000***	0.000
OLS (Mean)	0.042***	0.001	−0.001***	0.000

*Note: *** p<0.01. Experience = age − edu_years − 6. The near-identical estimates across quantiles confirm stable experience returns throughout the wage distribution.*

The stability of these estimates, ranging only from 3.9 per cent at Q90 to 4.3 per cent at Q10 and Q75, suggests that labour market experience does not really contribute to distributional wage divergence in the way education and training do. Each additional year of experience raises wages by roughly 4 per cent regardless of where a worker sits in the distribution. This is a sharp contrast with education (rising from 6.9 to 11.5 per cent) and training (ranging from 10.5 to 18.2 per cent). The finding fits comfortably with Mincer's (1974) original prediction of stable experience returns, and with the broader interpretation that on-the-job learning operates as a fairly general process rather than one that complements particular skill levels.

The plateauing of education returns near the top of the distribution, Q75 at 0.114 and Q90 at 0.115, hints at diminishing marginal signalling returns at high qualification levels: the earnings gap between a graduate and a post-graduate appears proportionately smaller than the gap between a secondary completer and a graduate. At Q10, the lower coefficient also reflects the limited scope formal credentials have to change employer behaviour in highly informal hiring contexts, domestic employers, construction supervisors, and agricultural contractors typically hire through personal networks rather than by screening credentials, which reduces the schooling premium at this end of the distribution.

None of this amounts to an argument against educational investment at any wage level, the 7.1 per cent annual return at Q10 is real and worth having. It is more an argument against assuming educational expansion is automatically equalising. At current rates of return, expanding secondary completion raises wages across the board, but by more at the top than at the bottom. If reducing wage inequality is a specific goal, vocational training, with its considerably more equitable distributional footprint, probably deserves a larger share of attention in the ongoing skill-versus-schooling allocation debate.

In practical terms, this also suggests that policies aimed at smoothing young workers' transition into stable employment, by reducing entry barriers, improving information about vacancies, and reducing hiring discrimination, would benefit workers across the wage distribution in a fairly even-handed way, since the experience returns these policies would help unlock are themselves distributed quite uniformly.

4.11 Summary of Key Findings

Six findings emerge from the quantile regression analysis taken as a whole. First, education widens wage inequality: returns rise monotonically from 6.9 per cent per year at Q10 to 11.5 per cent at Q90, confirming H1 and marking education out as a skill-biased force in the Indian wage distribution at current rates of return. Second, vocational training follows an inverted-U pattern, peaking at the median (18.2 per cent) and moderate at both tails (12.1 per cent at Q10, 10.5 per cent at Q90); training is more egalitarian than education, though not concentrated at the very bottom, so H2 is confirmed only partially. Third, gender penalties are largest at the bottom, a clear sticky floor: the female penalty falls from 49.3 per cent at Q10 to 17.6 per cent at Q90, confirming H3, and pointing to the worst gender discrimination occurring precisely where workers are most economically vulnerable. Fourth, SC and OBC disadvantages widen at higher wage levels, confirming the upper-end caste barrier: SC penalties grow from 8.8 per cent at Q10 to 15.9 per cent at Q90, and OBC penalties from 1.8 per cent to 13.2 per cent, confirming H4 and suggesting caste discrimination is most binding in higher-skill, higher-wage employment. Fifth, the urban premium is largest for low-wage workers, declining from 26.1 per cent at Q10 to 18.6 per cent at Q90, which suggests urban labour markets offer proportionately greater gains to the least well-paid workers. Sixth, experience returns are stable across quantiles at roughly 4 per cent per year and do not appear to contribute to distributional divergence at all.

CHAPTER 5: CONCLUSION AND POLICY IMPLICATIONS

5.1 Summary of Findings

This dissertation set out to examine the distributional wage effects of skill development in India using quantile regression applied to PLFS 2023-24 data for 46,294 wage workers. The basic argument running through it is that skill development does improve wages, that much is not really in question, but the improvement is highly uneven across the wage distribution, and that unevenness carries fairly direct implications for how training programmes ought to be designed and targeted going forward.

Formal education raises wages at every quantile, but it does so far more powerfully toward the top, 6.9 per cent per year at Q10 rising to 11.5 per cent at Q90, which confirms education as a force for wage divergence at current rates of return. Vocational training, by contrast, traces out an inverted-U: the premium peaks at the median (18.2 per cent), is more moderate at the lower tail (12.1 per cent at Q10), and declines somewhat at the upper tail (10.5 per cent at Q90), making it noticeably more egalitarian in its distributional footprint than formal education. Gender wage penalties are severe and concentrated at the bottom of the distribution, 49.3 per cent at Q10, consistent with a sticky floor. SC and OBC penalties widen at higher wage levels, the SC penalty growing from 8.8 to 15.9 per cent, pointing to caste-based exclusion operating most powerfully in higher-skill employment. Urban workers see the largest relative advantage at the bottom of the distribution (26.1 per cent at Q10), and experience returns remain stable and broadly uniform across quantiles at around 4 per cent per year.

5.2 Objective-wise Conclusions

The first objective was to examine how skill development affects wages across the distribution using PLFS 2023-24. What emerges is that both education and vocational training are associated with significantly higher wages at every quantile, but with quite different distributional profiles. Education reinforces the advantage already held by high-wage workers, while vocational training looks considerably more egalitarian, with returns clustering around the median. OLS estimates, which earlier studies have largely relied on, conceal this heterogeneity entirely.

The second objective was to assess whether skill development reduces or widens wage inequality, and the honest answer here is that it depends on which form of skill investment one is talking about. Formal education, at current rates of return, appears to widen inequality: the gap between Q10 and Q90 education returns (6.9 versus 11.5 per cent per year) means

educational expansion raises wages faster at the top than at the bottom. Vocational training looks more equitable, the Q10 premium (12.1 per cent) is not dramatically below the Q90 premium (10.5 per cent), and the gap between them is much narrower than the corresponding education gap. Where India's overall skill investment lands on the inequality question probably depends on the balance between education expansion and training investment, and on whether training access genuinely reaches low-wage workers, who currently have limited programme reach.

The third objective was to look at heterogeneity in skill returns across gender, location, and social group, and here the results turn out to be fairly substantial and policy-relevant. Female workers at the bottom of the distribution face by far the largest gender penalty. SC and OBC workers face worsening disadvantage as wages rise. Rural workers at the bottom gain the most from being in an urban location, which suggests that expanding rural training infrastructure could deliver something close to a double dividend. These intersections point fairly clearly toward differentiated policy responses rather than a one-size-fits-all programme design.

The fourth objective was to draw out policy implications, which are developed more fully in section 5.4 below. The central insight, if there is one, is that making skill development more progressive is not just a matter of expanding access to training. It also requires rethinking how training is targeted, what it actually includes, and what labour market conditions need to be in place for training to translate into real wage gains, particularly for women, SC/ST workers, and low-wage rural youth.

5.3 Theoretical Implications

The findings speak to at least three ongoing theoretical debates. First, they offer fairly direct empirical support for the SBTC hypothesis in the Indian context: education returns rising monotonically across quantiles is consistent with technological change that complements high-skill labour and raises relative demand for workers with strong cognitive ability. This extends Kijima's (2006) earlier analysis to PLFS 2023-24 using methods that are sensitive to distribution in a way his were not.

Second, the inverted-U shape of vocational training returns raises some awkward questions for a simple human capital story, which would predict either flat returns or returns that vary monotonically in one direction. The peak at the median sits more comfortably within a signalling framework, where VT credentials carry the most informational value for "middle-skill" positions, where employers use certification as a primary screening device, but have diminishing signalling value at high-wage positions (where education credentials already

dominate) and limited traction at the bottom (where informal hiring and discrimination tend to override whatever signal a credential might send). One implication worth flagging is that strengthening the labour market recognition of VT credentials, through industry-validated certification, placement linkages, and Recognition of Prior Learning, might not only raise average returns but could shift the peak of the inverted-U downward toward Q25, making training more progressive than it currently is.

Third, the sticky floor finding for gender and the escalating caste penalty toward the top of the distribution together pose a more fundamental challenge to human capital theory as a complete account of Indian wage determination. These patterns are highly significant and survive the full set of controls, which makes them difficult to reconcile with any theory in which wages simply reflect marginal productivity. They sit much more comfortably within models of statistical discrimination, taste-based discrimination, and occupational segregation, models that can generate persistent wage gaps even between workers with identical observable characteristics. For theory, this suggests that any realistic account of Indian wage determination needs to treat social identity variables not as add-ons bolted onto a production-based framework, but as primary determinants of which segments of the labour market different workers can actually access.

5.4 Policy Recommendations

Five specific recommendations follow reasonably directly from the empirical findings.

The first is to redirect training enrolment toward low-wage workers through active targeting. The inverted-U pattern in training returns, combined with the sticky floor and urban premium findings, suggests that the workers who most need a wage boost, those at Q10 and Q25, face structural barriers that prevent them from realising the full potential benefit of training. DDU-GKY already targets rural BPL youth fairly explicitly; PMKVY could implement something analogous for urban informal workers, with stipends to offset the opportunity cost of training time, mobile training units for workers who cannot travel to fixed centres, and women-only training options that address some of the social norm barriers to female participation. What seems to be needed is active outreach rather than simply open application windows, if Q10 workers are to be reached at all.

The second is to design gender-disaggregated skill programmes. The 49.3 per cent wage penalty facing women at Q10 does not reflect a skills deficit, it reflects structural discrimination and occupational segregation. Gender-neutral programme design is unlikely to close this gap on its own. What seems to be needed instead includes dedicated enrolment

targets for women, with outreach specifically into domestic service, garment production, and casual agriculture; training in sectors where gender discrimination appears to be less severe; placement services that actively monitor gender wage gaps among placed candidates; and non-discrimination commitments from employer partners. Maitra and Mani (2017), a randomised evaluation showing positive earnings effects for female trainees in a placement-linked programme, demonstrates that well-designed interventions of this kind can work, the harder question is how to scale them up.

The third is to reform VT curriculum and certification with an eye toward maximising labour market relevance, particularly for median and lower-wage workers. The inverted-U peaking at Q50 suggests training works best where there is a clear, identifiable market for the specific skill being certified. Industry-aligned curriculum development, regular revisions to reflect changing employer demand, and a more robust Recognition of Prior Learning system, which would let workers already doing a job informally get formally certified, could raise the market value of VT credentials precisely where informal workers are most concentrated, at the bottom of the distribution. Stackable credentials, allowing trainees to build up qualifications progressively rather than needing to complete a single long course, would also help reduce the opportunity cost barrier facing the lowest-wage workers.

The fourth is to expand digital and vocational training infrastructure in rural areas. The declining urban premium, from 26.1 per cent at Q10 down to 18.6 per cent at Q90, implies the location-based disadvantage facing rural workers is largest precisely at the bottom of the distribution. Broadband expansion, upgrades to rural ITIs, mobile training vans, and digital literacy modules built into PMKVY curricula could simultaneously address skill gaps and reduce the spatial wage penalty for the most disadvantaged workers. This fits with the existing rural focus of DDU-GKY and with the broader Digital India and Skill India programmes; what seems to be missing is better evidence-based coordination across these initiatives.

The fifth is to pair skill training with anti-discrimination enforcement, particularly in higher-wage private sector employment. The widening SC penalty (8.8 per cent at Q10 to 15.9 per cent at Q90) and OBC penalty (1.8 per cent to 13.2 per cent) make it fairly clear that the main barrier facing these groups in higher-wage employment is not a lack of skills but discrimination. Training cannot, on its own, address a penalty that grows precisely as skills improve. Stronger enforcement of the Equal Remuneration Act, mandatory pay transparency reporting for larger firms, affirmative action incentives in the private sector such as tax credits for SC/OBC hiring into managerial roles, and mentorship or networking programmes that help SC and OBC workers access the informal channels through which high-wage jobs are often

allocated, all seem like necessary complements to skill training policy rather than optional extras.

Running through all five of these recommendations is a more general point about measurement: programme monitoring needs to track distributional outcomes, not just averages. The standard metrics for PMKVY and DDU-GKY evaluations tend to be placement rate and average wage of placed candidates. Adding distributional monitoring, reporting wage outcomes at the 10th, 25th, and 50th percentiles of the placed candidate distribution, disaggregated by gender, caste, and rural/urban location, would let programme managers see whether outcomes are actually improving at the bottom of the distribution, and whether placement wages for female and SC/OBC trainees are comparable to those for male and upper-caste trainees. This kind of monitoring is relatively low-cost and would substantially improve the equity relevance of how these programmes are evaluated.

5.5 Limitations

A number of limitations are worth setting out plainly. First, the cross-sectional design rules out causal inference, the coefficients reflect conditional correlations rather than treatment effects. Workers who receive training likely differ from those who do not in ways that are not directly observable, motivation, family support, geographic proximity to training centres, all of which could also affect wages independently. This positive selection bias probably inflates both the OLS and QR estimates of training returns relative to whatever the true causal effect might be. Future work using genuine programme evaluation designs, regression discontinuity around PMKVY eligibility thresholds, or difference-in-differences using administrative data, would provide considerably stronger causal evidence than is possible here.

Second, the binary skill dummy cannot capture the substantial heterogeneity in training type, duration, quality, or provider. A three-month PMKVY short course and a two-year ITI diploma are both coded as skill=1. The relatively modest Q10 training return (12.1 per cent) may partly reflect the fact that shorter, lower-quality training is more dominant at the bottom of the distribution. Disaggregating by training duration and provider type would be a natural next step.

Third, restricting the sample to wage workers excludes a very large share of the Indian workforce in self-employment and unpaid family work, segments that include many of the poorest workers in the country. The wage-worker sample is therefore not representative of the labour force as a whole. The supervisor's chapter addresses selection into wage employment

using Heckman's two-step method, but PLFS 2023-24 does not provide the household composition variables needed to replicate that correction here.

Fourth, the Mincerian experience proxy (age − edu_years − 6) is fairly imprecise, particularly for women who have had career interruptions, and any resulting attenuation bias in the experience coefficient could slightly inflate the coefficients on other regressors.

Fifth, all estimates here are nominal. PLFS wages are not adjusted for regional cost-of-living differences, which in India are substantial. An urban worker earning ₹10,000 a month in Mumbai has considerably less real purchasing power than a rural worker earning the same amount in Bihar. The urban premium estimated here, somewhere between 18.6 and 26.1 per cent, may partly reflect cost-of-living compensation rather than a pure productivity advantage.

A sixth limitation concerns the social group categories themselves. SC, ST, and OBC as used in PLFS aggregate together highly diverse communities with very different occupational profiles, regional concentrations, and historical relationships to the labour market. Scheduled Castes in Punjab face a different labour market reality from Scheduled Castes in Tamil Nadu, to take one obvious example. The aggregate coefficients reported here average across all of this diversity, and inevitably conceal regional and community-level variation that a more granular state-level or community-level analysis would be better placed to capture.

5.6 Future Research Directions

A few extensions of this work seem worth pursuing. Panel or repeated cross-section analysis using multiple PLFS rounds, from 2017-18 through to 2023-24, would allow the distributional pattern of skill returns to be tracked over time as PMKVY has scaled up, and could support difference-in-differences estimates of programme effects for specific training cohorts.

Disaggregating by training programme type, ITI diploma versus PMKVY short course versus DDU-GKY placement-linked training, would help establish whether the inverted-U pattern observed here is driven by underlying heterogeneity in training quality, or whether it reflects something more fundamental about how the wage-skill relationship works across the distribution regardless of programme type.

A wage distribution decomposition using either the Machado-Mata (2005) counterfactual approach or the Firpo-Fortin-Lemieux (2009) unconditional quantile regression method would help separate out how much of any change in the wage distribution is due to changes in skill endowments, composition effects, versus changes in the returns to those skills, price effects. This would give a fuller picture of whether rising wage inequality in India is being driven more by widening skill gaps or by rising skill premia.

Perhaps most directly useful from a policy standpoint would be a qualitative study of employer hiring practices for SC, OBC, and female candidates in higher-wage positions, drawing on employer interviews and HR process audits. This kind of mechanistic evidence is really what would be needed to design anti-discrimination interventions capable of addressing the upper-end caste barrier and sticky floor identified in this study.

A fifth direction worth mentioning would involve linking administrative data from PMKVY enrolment records to PLFS respondents, if this proves feasible through geographic or individual-level matching. If unique identifiers or district-level matching could connect training programme records to wage survey respondents, this would open up the possibility of difference-in-differences or propensity score matching designs that address at least some dimensions of the selection bias problem. NSDC already collects enrolment and certification data; linking these records to PLFS would create what could be one of the richest resources currently available for causal evaluation of India's flagship skill programme.

5.7 Final Conclusion

India's skill development programmes do appear to improve wages. That much is reasonably clear from the data, and it justifies continued public investment on its own terms. What is less clear, and what this dissertation has tried to address, is for whom those improvements are largest, and whether the current distribution of skill returns sits comfortably with the redistributive goals that are often used to justify the investment in the first place.

The answer that emerges here is, inevitably, a nuanced one. Vocational training, with its inverted-U profile peaking at the median, turns out to be more equitable in its wage effects than formal education, which rewards higher-wage workers more or less monotonically. But neither education nor training, as currently accessed and delivered, is systematically concentrated at the bottom of the distribution, where the need is arguably greatest. And the structural barriers of gender discrimination and caste exclusion compound the problem further: a woman at the 10th percentile of the wage distribution faces a wage penalty of close to 50 per cent relative to a comparable man, and an SC worker at the 90th percentile faces a growing penalty of around 16 per cent, one that no amount of skill investment is likely to eliminate so long as discrimination in high-wage hiring persists.

Realising India's demographic dividend will require more than simply more training. It will require training that is better targeted, better designed, and better supported, alongside anti-discrimination enforcement and broader labour market institution-building that training on its own cannot substitute for. What this dissertation offers is a distributional evidence base for

that kind of reform agenda, and points, hopefully with some specificity, toward the interventions most likely to make India's skill development investment a more equitable one in practice.

REFERENCES

Acemoglu, D. (2002). Directed technical change. The Review of Economic Studies, 69(4), 781 to 809.

Acemoglu, D., & Pischke, J. S. (1998). Why do firms train? Theory and evidence. Quarterly Journal of Economics, 113(1), 79 to 119.

Acemoglu, D., & Pischke, J. S. (1999). Beyond Becker: Training in imperfect labour markets. The Economic Journal, 109(453), F112 to F142.

Acharya, S. (2016). Malthus at the heart of India. Business Standard, 13 October.

Albert, C., García-Serrano, C., & Hernanz, V. (2010). On-the-job training in Europe: Determinants and wage returns. International Labour Review, 149(3), 315 to 341.

Almeida, R., & Faria, A. (2014). The wage returns to on-the-job training: Evidence from matched employer-employee data. IZA Discussion Paper No. 8314.

Almeida-Santos, F., Chzhen, Y., & Mumford, K. (2010). Employee training and wage dispersion: White and blue collar workers in Britain. IZA Discussion Paper No. 4821.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2005). Rising wage inequality: The role of composition and prices. NBER Working Paper No. 11628.

Azam, M., Chin, A., & Prakash, N. (2013). The returns to English-language skills in India. Economic Development and Cultural Change, 61(2), 335-367.

Barron, J. M., Berger, M. C., & Black, D. A. (1999). Do workers pay for on-the-job training? Journal of Human Resources, 34(2), 235 to 252.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of Political Economy, 70(5, Part 2), 9 to 49.

Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. Columbia University Press.

Bhattacharya, P. C. (2011). Economic development, gender inequality and demographic transitions: India and sub-Saharan Africa: population and Development Review, 37(2), 309 to 340.

Bloom, D. E., & Williamson, J. G. (1998). Demographic transitions and economic miracles in emerging Asia. The World Bank Economic Review, 12(3), 419 to 455.

Blundell, R., Dearden, L., & Meghir, C. (1996). The determinants and effects of work-related training in Britain. Institute for Fiscal Studies.

Booth, A. L. (1993). Private sector training and graduate earnings—the Review of Economics and Statistics, 75(1), 164 to 170.

Buchinsky, M. (1994). Changes in the U.S. wage structure, 1963 to 1987: Application of quantile regression. Econometrica, 62(2), 405 to 458.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), Handbook of Labour Economics (Vol. 3, pp. 1801 to 1863). Elsevier.

Chattopadhyay, S. (2016). Return to general education and vocational education and training in the Indian context—International Journal of Educational Management, 30(3), 370 to 385.

Das, P. (2008). Wage inequality in India: Decomposition by sector, gender and activity status—the Indian Journal of Labour Economics, 51(3), 481 to 504.

Dearden, L., Reed, H., & Van Reenen, J. (2006). The impact of training on productivity and wages: Evidence from British panel data. Oxford Bulletin of Economics and Statistics, 68(4), 397 to 421.

Deshpande, A., & Ramachandran, R. (2019). The changing contours of inter-group discrimination: Evidence from India's labour markets, 2004 to 2012—Journal of Development Studies, 55(5), 888 to 908.

Dutta, P. V. (2006). Returns to education: New evidence for India, 1983 to 1999: Education Economics, 14(4), 431 to 451.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. Econometrica, 77(3), 953 to 973.

Gill, S., Sharma, M., & Kaur, R. (2024). Skill development and youth employment in India: A review of programme evidence—Journal of Asian Economics, 89, 101657.

Gorlitz, K. (2010). Continuous training and wages: An empirical analysis using a comparison-group approach. Ruhr Economic Paper No. 197.

Goux, D., & Maurin, E. (2000). Returns to firm-provided training: Evidence from French worker-firm matched data: Labour Economics, 7(1), 1 to 19.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. Econometrica, 45(1), 1 to 22.

Heckman, J. J. (1979). Sample selection bias as a specification error: Econometrica, 47(1), 153 to 161.

Kapoor, R. (2019). Technology, jobs and inequality: Evidence from India's manufacturing sector. The Journal of Development Studies, 55(12), 2555 to 2573.

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963 to 1987: Supply-and-demand factors. The Quarterly Journal of Economics, 107(1), 35 to 78.

Kijima, Y. (2006). Why did wage inequality increase? Evidence from urban India, 1983 to 99—Journal of Development Economics, 81(1), 97 to 117.

Koenker, R. (2005). Quantile regression. Cambridge University Press.

Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33 to 50.

Konings, J., & Vanormelingen, S. (2010). The impact of training on productivity and wages: Firm-level evidence. IZA Discussion Paper No. 4731.

Krafft, C. (2013). Is school the best route to skills? Returns to vocational school and vocational skills in Egypt. Working Paper No. 2013-09. Economic Research Forum.

Laukkanen, E. (2010). Wage returns to training: Evidence from Finland: Doctoral Thesis, Abo Akademi, Finland.

Lechner, M. (2000). An evaluation of public-sector-sponsored continuous vocational training programs in East Germany. Journal of Human Resource, 35(2), 347-375.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? American Economic Review, 96(3), 461-498.

Leuven, E., & Oosterbeek, H. (2008). An alternative approach to estimate the wage returns to private-sector training. Journal of Applied Econometrics, 23(4), 423 to 434.

Lillard, L. A., & Tan, H. W. (1992). Private sector training: Who gets it and what are its effects? Research in Labour Economics, 13, 1 to 62.

Lynch, L. M. (1992). Private-sector training and the earnings of young workers. The American Economic Review, 82(1), 299 to 312.

Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression—Journal of Applied Econometrics, 20(4), 445 to 465.

Madheswaran, S., & Attewell, P. (2007). Caste discrimination in the Indian urban labour market: Evidence from the National Sample Survey. Economic and Political Weekly, 42(41), 4146 to 4153.

Maitra, P., & Mani, S. (2017). Learning and earning: Evidence from a randomised evaluation in India. Labour Economics, 45, 116 to 130.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. Quarterly Journal of Economics, 107(2), 407 to 437.

Mehrotra, S., & Parida, J. K. (2019). India's employment crisis: Rising education levels and falling non-agricultural employment growth. CSE Working Paper 2019-04. Azim Premji University.

Mincer, J. (1958). Investment in human capital and personal income distribution. Journal of Political Economy, 66(4), 281 to 302.

Mincer, J. (1962). On-the-job training: Costs, returns and some implications. Journal of Political Economy, 70(5), S50 to S79.

Mincer, J. (1974). Schooling, experience, and earnings. National Bureau of Economic Research.

Ministry of Skill Development and Entrepreneurship. (2019). Evaluation of Pradhan Mantri Kaushal Vikas Yojana 2.0: Final report. Government of India.

Monk, C., Sandefur, J., & Teal, F. (2008). Does doing an apprenticeship pay off? Evidence from Ghana. CSAE Working Paper 2008-08. University of Oxford.

MoSPI. (2024). Periodic Labour Force Survey (PLFS) Annual Report 2023 to 24. Ministry of Statistics and Programme Implementation, Government of India.

Mulligan, C. B., & Sala-i-Martin, X. (1995). A labour-income-based measure of the value of human capital: An application to the states of the United States. NBER Working Paper No. 5018.

Parida, J. K., & Awasthi, D. N. (2020). Gender wage differentials in India: The role of human capital, institutional factors, and labour market segmentation: Economic and Political Weekly, 55(9), 30 to 38.

Parent, D. (2003). Employer-supported training in Canada and its impact on mobility and wages. Empirical Economics, 28(3), 431 to 459.

Pischke, J. S. (2000). Continuous training in Germany. IZA Discussion Paper No. 137.

Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update: Education Economics, 12(2), 111 to 134.

Sarmah, P., & Ang, J. B. (2019). Formal vocational education and training and wages in East and Southeast Asia: Asian Economic Papers, 18(2), 142 to 165.

Schultz, T. W. (1961). Investment in human capital. American Economic Review, 51(1), 1 to 17.

Schultz, T. W. (1963). The economic value of education. Columbia University Press.

Singh, R., Parida, J. K., & Awasthi, D. N. (2020). Youth employment and wages in India: An analysis of PLFS 2017 to 18 data—Indian Journal of Labour Economics, 63(4), 985 to 1006.

Singh, R., Parida, J. K., & Awasthi, D. N. (2022). Skill development and youth wage outcomes in India: Distributional analysis using PLFS 2018 to 19—Journal of South Asian Development, 17(1), 35 to 61.

Spence, M. (1973). Job market signalling—the Quarterly Journal of Economics, 87(3), 355 to 374.

StataCorp. (2023). Stata statistical software: Release 17. StataCorp LLC.

Thakur, N. (2012). The demographic dividend in India: Gift or curse? A state-level analysis. International Development Working Paper No. 12-128. London School of Economics.

Thorat, S., & Newman, K. S. (Eds.). (2010). Blocked by caste: Economic discrimination in modern India. Oxford University Press.

Uma Rani, U. (2019). Digital technologies, wages and working conditions in India. ILO Asia-Pacific Working Paper Series. International Labour Organisation.

Veum, J. R. (1999). Training, wages and the human capital model. Southern Economic Journal, 65(3), 526-538.

Willis, R. J., & Rosen, S. (1979). Education and self-selection. Journal of Political Economy, 87(2), S7 to S36.

APPENDICES

Appendix I: Stata Do-File Summary Variable Construction

The following summarises the key steps for constructing variables from FINAL_Bhargavi_Dissertation_CLEANED.do, Section 11 onward. All regression results in Chapter 4 are generated by the qreg and reg commands in this do-file.

Step	Stata Command	Variable Created
Load data	use 'merged data set.dta', clear	All raw PLFS variables
Rename	capture rename b4q6_perv1 age	age
Rename	capture rename b4q10_perv1 edu_years	edu_years
Rename	capture rename b4q12_perv1 voc_training	voc_training
Rename	capture rename b6q9_perv1 wage_regular	wage_regular
Female dummy	gen female = (sex == 2)	female
Urban dummy	gen urban = (sector == 2)	urban
Skill dummy	gen skill = (voc_training == 1)	skill
SC dummy	gen sc = (caste == 2)	sc
ST dummy	gen st = (caste == 1)	st
OBC dummy	gen obc = (caste == 3)	obc
Wage construction	gen wage = wage_regular	wage
Drop zero wages	drop if wage <= 0 | wage == .	Cleaned sample
Log wage	gen ln_wage = log(wage)	ln_wage
Experience	gen experience = age - edu_years - 6	experience
Replace neg exp	replace experience = 0 if experience < 0	experience ≥ 0
Exp squared	gen experience2 = experience^2	experience2

Appendix II: Full Quantile Regression Coefficient Table with Standard Error

This appendix reproduces the complete regression coefficient table from the Quantile_results.rtf output file generated by Stata's esttab command. All values are taken directly from the Stata output.

Variable	Q10	Q25	Q50 Median	Q75	Q90
Years of education	0.069***	0.084***	0.102***	0.114***	0.115***
(Std Error)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Potential experience	0.043***	0.041***	0.042***	0.043***	0.039***
(Std Error)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Experience squared	−0.001***	−0.001***	−0.000***	−0.000***	−0.000***
(Std Error)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Skill (VT dummy)	0.114***	0.140***	0.182***	0.134***	0.105***
(Std Error)	(0.017)	(0.013)	(0.011)	(0.010)	(0.012)
Female dummy	−0.679***	−0.539***	−0.399***	−0.268***	−0.194***
(Std Error)	(0.013)	(0.010)	(0.009)	(0.009)	(0.011)
Urban dummy	0.232***	0.198***	0.187***	0.182***	0.171***
(Std Error)	(0.010)	(0.007)	(0.007)	(0.007)	(0.009)
Scheduled Caste	−0.092***	−0.092***	−0.132***	−0.153***	−0.173***
(Std Error)	(0.013)	(0.010)	(0.010)	(0.012)	(0.014)
Scheduled Tribe	0.096***	0.106***	0.106***	0.041***	−0.018
(Std Error)	(0.014)	(0.013)	(0.011)	(0.012)	(0.013)
OBC	−0.018*	−0.040***	−0.089***	−0.127***	−0.142***
(Std Error)	(0.010)	(0.009)	(0.008)	(0.008)	(0.011)
Constant	7.712***	7.879***	7.989***	8.185***	8.519***
N	46,294	46,294	46,294	46,294	46,294

Note: *Source: Stata qreg with vce(robust), exported via esttab to Quantile_results.rtf. * p<0.10; *** p<0.01. ST at Q90 is not statistically significant at conventional levels.*