A Quantile Regression Modelling Approach to Study Gender Wage Gap in India

Samapriya Trivedi[1] | *University of Lucknow, Lucknow, India*
Shambhavi Mishra[2] | *University of Lucknow, Lucknow, India*

Received 5.4.2024 (revision received 30.7.2024)**, Accepted** (reviewed) 10.8.2024, **Published** 14.3.2025

Abstract

The Indian labour market exhibits significant gender wage disparities, particularly among regular/salaried employees and casual workers. To study these disparities comprehensively, we present a dual-methodological approach by combining Quantile Regression (QR) and Melly-Machado-Mata (MMM) decomposition. Using secondary data from India's Periodic Labour Force Survey (PLFS) 2020–21, the study highlights the intricate interplay of various demographic, personal, and occupational characteristics on wage distributions. The findings highlight the persistence of the gender wage gap across different quantile levels for both employment types. The decomposition results reveal that discrimination significantly contributes to the wage gap, particularly at lower income levels, indicating a "sticky floor" effect for regular/salaried employees. Conversely, casual workers face a consistent wage gap across all quantiles, with discrimination remaining a crucial factor. This research highlights the robustness and precision of QR modelling and decomposition, providing a comprehensive framework for scientifically assessing the gender-based wage gap and exploring policy interventions to address these inequalities.

Keywords	***DOI***	***JEL code***
Quantile regression, decomposition, Oaxaca-Blinder, Melly-Machado-Mata, PLFS	*https://doi.org/10.54694/stat.2024.18*	*C10, C40, C50, J31*

INTRODUCTION

The persistent wage gap remains a prominent concern within labour economics, reflecting its complex nature as a multifaceted bias embedded within market mechanisms. This discrepancy in earnings, often measured by comparing wages across groups differentiated by gender, education, and other factors,

[1] Department of Statistics, University of Lucknow, Lucknow, Uttar Pradesh, 226007, India. E-mail: s.p.trivedi796@gmail.com.
[2] Department of Statistics, University of Lucknow, Lucknow, Uttar Pradesh, 226007, India. Corresponding author: e-mail: shambhavimishra.lko@gmail.com. ORCID: <https://orcid.org/0000-0002-7057-067X>.

carries a profound weight in academic inquiry and professional practice. While human capital differences and market discrimination contribute to the gap, its persistence warrants further investigation even after controlling for such factors. The gender wage gap exemplifies this complexity, highlighting the need to address its multifaceted nature. Factors such as gender, caste, religion, region, and education intertwine to depress females' earning potential, even within comparable roles and skills. India's labour market also exhibits a stark "duality" between regular or salaried and casual workers. Regular, also called salaried employment, offers stability, social security, and compliance with labour norms, while casual employment entails instability, low wages, and limited entitlements. Despite recent growth, casual workers' wages remain significantly lower (India Wage Report, 2018). Also, despite laws demanding equal pay, discriminatory practices stubbornly stand in the way of fairness for females. Further interventions are crucial to dismantle discriminatory practices and achieve true economic parity for all.

Previous analyses of wage distribution, primarily relying on models based on means, have yielded limited insights into the intricate dynamics underlying these distributions. QR, as highlighted by Roger Koenker and Gilbert Bassett (Koenker and Bassett, 1978), offers a more comprehensive approach. It reveals how the influence of several factors on wages changes across the spectrum, from the lowest earners to the highest. Unlike ordinary least squares (OLS) regression, QR estimates the entire conditional wage distribution, not just the mean. This approach enables robust analysis of skewed wage data and outliers, providing valuable insights into distributional inequality and heterogeneity. While regression models offer useful insights into the relationship between wages and several factors, they often fail to explain the observed wage differences between groups comprehensively. Decomposition techniques like the Oaxaca-Blinder (OB) method (Blinder, 1973; Oaxaca, 1973) and Machado-Mata (MM) decomposition (Machado and Mata, 2005) address this limitation by statistically dissecting these differences. These techniques divide the wage disparity into two parts: the explained component, which addresses variances in personal attributes between groups, and the unexplained component, which signifies the gap that may be attributed to discrimination or undisclosed factors. This breakdown helps us identify the wage gap and measure the impact of several factors, providing insights for addressing inequalities and discriminatory practices.

This study employs QR, enabling a comprehensive analysis of how various worker and labour market characteristics influence the entire distribution of employee wages in India. This approach transcends the limitations of traditional mean-based analysis, which only captures the average impact. Additionally, decomposition techniques are utilised to dissect the relative contributions of observable and unobservable factors to wage disparities across different quantile levels, potentially revealing a "sticky floor" and the glass ceiling effects (Das, 2018). This nuanced approach offers valuable insights into the intricate dynamics of the disparities in workers' wages for men and women in an Indian context, potentially holding broader applicability for wage analysis across other countries. Furthermore, our methodological approach facilitates optimal model selection for future studies, crucial for informing policy interventions to narrow India's persistent wage gap.

1 LITERATURE SURVEY

Since the mid-1990s, a pronounced gender wage gap in India has sparked a considerable academic interest in gender-based wage discrimination. Economists and statisticians have extensively researched labour market discrimination within the context of developing countries, particularly India.

A substantial body of research has examined India's wage structure, exploring a range of factors influencing wages and the challenges of discrimination faced by minority groups and females. Kingdon and Unni (2001) analysed 1987–88 National Sample Survey Office (NSSO) data, finding a significant disparity in education returns. Building on this, Madheswaran and Attewell (2007) emphasised occupational discrimination over wage discrimination for disadvantaged groups like scheduled tribes and castes. Agrawal (2013) used 2005 India Human Development Survey (IHDS) data to show that while endowment differences partly explained the wage gap between social groups, labour market discrimination was the primary driver. According to Chakraborty and Mukherjee (2014), a significant gap in wages between different genders in India's industries and professions using NSSO's Employment-Unemployment Survey (EUS) 2009–10 data, indicating wage discrimination against women in rural and urban areas. Sengupta and Das (2014) demonstrated that economically disadvantaged caste women and religious minorities faced greater discrimination. Duraisamy and Duraisamy (2016) analysed the NSSO data from 1983 to 2012, finding a decreasing raw wage gap and signs of gender convergence in productive characteristics. Pala and Nongspung (2022) identified slow wage growth for regular workers but noted faster long-term growth for casual workers, suggesting convergence. These studies underscore the continued existence of the gender wage gap, highlighting its variable impact across different social groups and the intricate interplay between discriminatory practices and individual characteristics. Madan and Mor (2022) analysed the persistence of the gender earnings gap in India using the PLFS 2017−18 dataset, employing Generalized Linear Models (GLM) and Analysis of Covariance to estimate marginal mean earnings, revealing a significant gender earnings gap across occupational groups and work statuses, with males earning 1.744 times more than females, even after controlling for education.

Previous studies on wages in India have relied on mean-based methodologies, overlooking the labour market's nuanced dynamics and varied composition. These approaches have disregarded extreme wage structures and failed to encompass the entirety of the wage distribution for different employment types. This research gap emphasises the necessity for an all-inclusive assessment that includes the entire workforce spectrum and addresses gender disparities within these segments. Studies have advocated the use of QR modelling to obtain a deeper insight into wage differentials, particularly those concerning gender (Fitzenberger et al., 2021; Waldmann, 2018). Khanna (2012) identified a persistent "sticky floor" effect in India's wage distribution, where gender pay gaps are more pronounced at lower income levels. Deshpande and Sharma (2015) corroborated this finding, revealing significant income disparities within India's wage structure. Their research underscored the "sticky floor" phenomenon, hindering labour market access for low-income earners while also unveiling a glass ceiling impeding women's advancement toward higher-paying positions. Azam (2012) utilised the MM procedure to scrutinise the evolution of urban wage structures in India between 1983 and 2004, drawing on NSSO data. Azam and Prakash (2015) extended this approach to investigate public-private wage differentials within India's labour market in 2004–05, again relying on NSSO data. Mitra (2016) employed Augmented Mincerian equations to analyse how education and other factors affect the salaries of different worker groups (regular/casual, male/female) in India, highlighting the interplay between education and earnings across various segments. Sengupta and Puri (2021) examined the gender pay gap using NSSO data from the same period, employing OLS and linear QR methods. While their OLS decomposition provided insights at the mean level, it was limited in distributional analysis. To overcome these constraints, our study suggests using advanced decomposition techniques, such as MM or MMM, to illuminate how this gap affects wage distribution entirely. This approach would deepen our understanding of this complex phenomenon.

Drawing upon the extensive unit-by-unit data representing the entire nation from the PLFS conducted by the Ministry of Statistics and Programme Implementation (MoSPI), our study examines gender-based wage disparities between the regular/salaried and casual employees in India. Launched in 2017, the PLFS is an annual survey implemented by the Government of India's National Statistical Office (NSO). To ensure the findings are generalisable to the national demographics, the PLFS incorporates survey weights within its analysis. This rigorous design has established the PLFS as a valuable source for investigating employment trends, income patterns, and wage disparities within the Indian labour market (Pala and Nongspung, 2022).

2 METHODS

2.1 Quantile Regression Model

Quantile regression, recognised for its robustness and flexibility compared to traditional OLS regression, is frequently employed in wage analysis to explore the explanatory variables that heterogeneously affect conditional wage distribution. It measures how explanatory variables affect a specific part of the dependent variable's distribution without assuming a particular shape for that distribution (Waldmann, 2018). QR offers several advantages, including reduced sensitivity to outliers and misspecified error distributions commonly encountered in wage data (Huang et al., 2017; Patidar et al., 2023). Additionally, it can handle situations where the error variance varies with the explanatory variables, a scenario where OLS estimates may lack reliability (Porter, 2014). Although other techniques, such as Generalized Least Squares (GLS) and sandwich estimators, offer robust alternatives for heteroscedasticity, they have limitations. GLS is a statistical method used to estimate parameters in linear regression models, particularly in the presence of heteroscedasticity or autocorrelation among residuals. By transforming the original data, GLS addresses these issues, allowing for the efficient application of OLS on the transformed data. The sandwich estimator is another robust technique for estimating the variance of parameter estimates in regression models. It is commonly used alongside GLS to provide robust standard errors that remain valid even when assumptions like homoscedasticity are violated. However, GLS has limitations, as it assumes a specific form of heteroscedasticity or correlation among residuals and necessitates data transformation to satisfy OLS assumptions. Similarly, sandwich estimators are primarily applied to adjust standard errors in mean regression models. In contrast, quantile regression presents a robust alternative, with a more comprehensive data distribution analysis, by providing insights beyond the mean, which is the focus of both GLS and sandwich estimator methods.

The equation below estimates the QR model's coefficients:

$$y_i = X_i \beta_\theta + \mu_{\theta i}, \tag{1}$$

with:

$$Q_{\theta i}\left(y_i | X_i\right) = X_i \beta(\theta), \tag{2}$$

here: y_i is ln (daily pay) and X_i is the covariates related to workers, β is the coefficient vector, θ represents the specified quantile of the wage distribution ($0 < \theta < 1$) and $\mu_{\theta i}$ is the random error term which accounts for erratic components in y_i.

The θ^{th} QR estimator, $\hat{\beta}(\theta)$ minimises over $\beta(\theta)$, for the value of β. In this case, the objective function is:

$$\theta\big(\beta(\theta)\big) = \sum_{i \in \{i: y_i \geq X_i \beta\}}^{N} \theta \left| y_i - X_i \beta_\theta \right| + \sum_{i \in \{i: y_i < X_i \beta\}}^{N} (1-\theta) \left| y_i - X_i \beta_\theta \right|. \tag{3}$$

The estimation method for the QR relies on a linear programming approach. In STATA, the "QREG" command is used for this analysis, where the minimisation problem is formulated as a linear programming problem. This approach is consistent with the methodology suggested by Armstrong et al. (1979) and comprehensively described by Koenker (2005). It uses the simplex method, which iteratively improves the objective function at each step until the optimal solution is achieved. Thus, the coefficients from the QR model reveal how various factors influence wages at different points in the wage distribution, demonstrating how these factors' effects vary across different wage levels (Mitra, 2016).

Koenker and Machado (1999) developed a quantile-specific goodness-of-fit measure for QR models. This metric addresses the limitations of traditional, global measures by assessing model fit at individual quantiles, enabling a more localised evaluation of model performance. The pseudo-R^2, which ranges from 0 to 1, is calculated using the Residual Absolute Sum of Weighted Differences (RASW) and the Total Absolute Sum of Squared Differences (TASW) according to the provided formula for the specific quantile (θ):

$$pseudo\, R_\theta^2 = 1 - \frac{RASW_\theta}{TASW_\theta}\ , \tag{4}$$

where:

$$RASW_\theta = \sum_{y_i \geq X_i\hat{\beta}(\theta)} \theta \left| y_i - X_i\hat{\beta}_\theta \right| + \sum_{y_i < X_i\hat{\beta}(\theta)} (1-\theta)\left| y_i - X_i\hat{\beta}_\theta \right|,$$

$$TASW_\theta = \sum_{y_i \geq \theta} \theta \left| y_i - \hat{\theta} \right| + \sum_{y_i < \theta} (1-\theta)\left| y_i - \hat{\theta} \right|.$$

In the above equations, $X_i\hat{\beta}_\theta$ represents the predicted dependent variable for the i^{th} recording at quantile °θ and $\hat{\theta}$ denotes the estimated quantile value. $RASW_\theta$ differs from the absolute quantile function presented in Formula (3) as it is utilized to assess the QR model is good fit. In contrast, the absolute quantile function is integral to the optimisation process for estimating the model parameters.

It is essential to acknowledge that various methodologies exist for analysing wage distribution, each with limitations and unique features. Unconditional Quantile Regression (UQR) assesses covariates' influence on the dependent variable's unconditional quantiles, which might offer a limited understanding of the conditional distribution of wages (Adireksombat et al., 2010). The Heckman Correction Model addresses sample selection bias but is sensitive to the choice of instruments and the specification of the selection equation. GLMs primarily focus on the mean of the outcome variable, potentially overlooking significant distributional aspects of the wage gap and often relying on assumptions about the distribution of error terms, which may not always be valid (Madan and Mor, 2022). Kernel regression, being non-parametric, can model complex relationships but does not offer the same level of detail about the conditional distribution of wages as QR. Therefore, the application of QR to study the gender wage distribution among the two categories of employees is well-justified.

2.2 Oaxaca-Blinder Decomposition

The OB Decomposition method extends regression analysis by decomposing the average difference in outcome between two groups into attributable components related to group differences in independent variable endowments and the effects of these variables. This method employs OLS regressions for each group separately, predicting the dependent variable using the same explanatory variables. The decomposition then isolates the contributions of endowment and coefficient effects. Endowment (explained) effects capture the influence of disparities in the groups' average levels of explanatory variables. Conversely, coefficient (unexplained) effects isolate how the explanatory variables impact the outcome variable. However, each group's impact is measured separately based on their respective coefficients estimated in the regressions.

The approach entails calculating wage equations independently for individuals belonging to the male (m) and female (f) groups:

$$y_{gi} = \beta_{g0} + \sum_{i=1}^{p} X_{gki}\beta_{gk} + \mu_{gi} \,, \tag{5}$$

where g represents the two groups (male and female), k represents the independent variable, while all other variables maintain the same meanings as defined in Formula (1). Considering that the residuals from OLS regression have a mean of zero, the equation below calculates the discrepancy in average wages across both genders by comparing the predicted wages for each group:

$$\overline{y}_m - \overline{y}_f = \left(\hat{\beta}_{m0} + \sum_{k=1}^{p} \overline{X}_{mk}\hat{\beta}_{mk}\right) - \left(\hat{\beta}_{f0} + \sum_{k=1}^{p} \overline{X}_{fk}\hat{\beta}_{fk}\right). \tag{6}$$

The assumption that the non–discriminatory wage framework applied to males was used to construct a counterfactual (CF) average wage for females using the coefficients estimated for males:

$$CF_f = \hat{\beta}_{m0} + \sum_{k=1}^{p} \overline{X}_{fk}\hat{\beta}_{fk} \,. \tag{7}$$

We can estimate the effect of discrimination on wages by creating a hypothetical scenario where male employees earn the same as female employees for observable characteristics. Now, adding and subtracting Formula (7) from Formula (6), we get:

$$\overline{y}_m - \overline{y}_f = \left(\hat{\beta}_{m0} - \hat{\beta}_{f0}\right) + \sum_{k=1}^{p} \overline{X}_{fk}\left(\hat{\beta}_{mk} - \hat{\beta}_{fk}\right) + \sum_{k=1}^{p}\left(\overline{X}_{mk} - \overline{X}_{fk}\right)\hat{\beta}_{mk} \,. \tag{8}$$

The first two terms decompose the coefficient effects, reflecting wage discrimination through discrepancies in returns for each gender's attributes (separate regression coefficients). These disparities create a wage gap despite controlling for average covariate levels. The final term isolates the unexplained portion due to the unequal distribution of individual characteristics across genders. It captures the average log wage difference attributable to gender differences in average covariate levels (Deshpande et al., 2017).

2.3 Melly-Machado-Mata Decomposition

This study used Melly's refined MM methodology to achieve a quantile-specific decomposition of the gender-based employment wage gap. This approach breaks down the observed disparity into components specific to each wage distribution quantile. Doing so separates the influence of individual worker characteristics from the impact of wage structures associated with those characteristics. This facilitates an in-depth wage disparity analysis across the whole wage distribution for male and female employees, which moves beyond the limitations of mean-focused decomposition techniques, revealing the dynamic nature of the disparity at diverse wage spectrum quantiles (Azam, 2012).

From Formula (2), for each group, the conditional quantile function can be stated as:

$$Q_\theta\left(y_{gi}|X_{gi}\right) = X_{gi}\beta_g\left(\theta\right);\ \theta \epsilon \left(0.1\right). \tag{9}$$

In accordance with MM methodology, the following steps outline the decomposition process:

i. Generate a random sample from a uniform distribution $U[0.1]$. This step leverages the probability integral transformation theorem, which asserts that if U is a uniform random variable on [0.1], then $F^{-1}(U)$ follows the distribution F. By applying the inverse cumulative distribution function (CDF) F^{-1} of the wage distribution to these uniform random variables, the transformation $F^{-1}(\theta_i)$ yields the conditional quantiles of wages. This process effectively simulates a sample from the estimated conditional salary distribution, given a set of covariates.
ii. Estimate n distinct QR coefficient vectors for males and females separately.
iii. Draw independent random samples with replacement from the covariate distributions of males and females.
iv. Construct the counterfactuals by multiplying various combinations of estimated quantile coefficients and respective covariate distributions across genders, i.e., $y_j^{cf} = \tilde{X}_j^m \beta_{\mu j}^f$.

The final decomposition model is as follows:

$$\hat{Q}_m(\theta) - \hat{Q}_f(\theta) = \left(\hat{Q}_m(\theta) - \hat{Q}_{cf}(\theta)\right) + \left(\hat{Q}_{cf}(\theta) - \hat{Q}_f(\theta)\right). \tag{10}$$

The observed disparity can be divided into two components: the characteristics (explained) component, shown as the initial term on the right, measures how much of the difference can be accounted for by characteristic variations. It is calculated as the difference between the quantile regression estimates for males and the counterfactual distribution, which represents what the female wage distribution would be if females had the same characteristics as males but were paid according to the male wage structure. The second component is the coefficients (unexplained), which describe the remainder of the difference that cannot be explained by the measured characteristics and represent discrimination or bias, indicating that even if females had the same characteristics as males, they would still face a wage gap due to differences in how these characteristics are valued in the labour market (Deshpande et al., 2017; Khanna, 2012). As specified in Formula (10), the MM decomposition model utilises coefficient estimates obtained from quantile regression for both male and female employees at selected quantiles of the wage distribution. In contrast, the OB model, outlined in Formula (8), employs a linear regression framework that concentrates on the mean of the wage distribution. While both decompositions are used to analyse the gender wage gap, the OB model emphasises mean differences, whereas the MM decomposition, with its use of quantile-specific coefficient estimates, provides a more comprehensive analysis across the entire distribution, capturing variations at different quantiles.

The MM decomposition provides valuable insights into wage inequality. However, its reliance on computationally expensive Monte Carlo simulations for counterfactual wage estimation is a significant limitation. Melly proposed an alternative quantile-based decomposition approach that addresses this limitation by directly integrating the conditional wage distribution across the relevant variable space, eliminating the need for computationally expensive simulations. This methodological improvement allows for a more efficient and statistically robust analysis of wage inequality. Melly's framework disaggregates wage inequality at specific wage distribution quantiles into distinct characteristic and coefficient components. The characteristic component captures disparities that can be credited to differences in worker features, while the coefficients component isolates disparities in wage returns due to other factors. Crucially, the quantile-based approach demonstrates convergence to the MM results under the assumption of infinitely many simulations, confirming its accuracy and computational efficiency.

For alternative approaches, the following suggestions could be considered for different studies within the same domain. One method proposed by Ñopo (2008) involves decomposing the wage gap by comparing individuals with similar characteristics. This method overtly explains the differences in the supports of the characteristics' distribution. However, it may encounter issues related to dimensionality and the availability of suitable instruments (Ñopo et al., 2012). Additionally, copula analysis within the decomposition framework can enhance the accuracy and flexibility of gender wage gap analysis by effectively modelling the dependence structure between variables and adjusting for sample selection bias. This approach is particularly useful for addressing sample selection issues and providing a more flexible and accurate decomposition of wage gaps across different quantiles. This is crucial for accurately estimating the wage distribution, especially when dealing with non-random selection into the labour force (Arellano and Bonhomme, 2017; Biewen and Erhardt, 2021).

2.4 Gini coefficient of inequality

The Gini coefficient, a statistical tool, is the most utilised measure of inequality. It effectively summarises the extent of a population's income disparity. An alternative method to calculate the Gini coefficient is the relative mean absolute difference, which is complementary to the traditional Lorenz curve-based method. However, the most used formula for the Gini coefficient, introduced by Stephen P. Jenkins in 1999, is more computationally efficient and captures the same concept as the relative mean absolute difference by utilising the ranks of incomes (Jenkins, 1999). This formula, employed in our study, is given by:

$$G = 1 + \frac{1}{N} - \frac{2}{aN^2} \sum_{i=1}^{n} \left(N - i + 1 \right) w_i, \tag{11}$$

here: N represents the total number of employees, w is the i^{th} and j^{th} individuals' incomes, and a is the arithmetic mean of the income (Gazeley et al., 2018). The Gini coefficient, which measures relative rather than absolute inequality, spans from 0 to 1. A value of 0 suggests perfect equality, and a 1 signifies complete inequality.

3 RESULTS

3.1 Data description

This research utilised Unit-Level data from the PLFS Schedule 10.4 (first visit) for the period between July 2020 and June 2021, conducted by the NSO and obtained from MoSPI, to evaluate the gender-based wage gap in India's regular/salaried and casual-basis employees. Worker classifications followed the NSSO activity status definitions, i.e., regular/salaried employees receive fixed wages (not based on daily contracts), while casual workers work irregularly and are paid per day or contract. Self-employment was excluded due to challenges in separating profit and wage components.

We analysed factors influencing the distribution of daily wages of workers in both genders aged 15–59 in regular/salaried and casual employment in India. We used the natural logarithm of daily wage, derived from weekly salary and days worked reported in the NSO, as the dependent variable. To understand the effects on wage distribution across the spectrum, we systematically assessed predictor impacts at five quantiles (10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th}) for both regular/salaried employees and casual workers. This approach revealed key distributional characteristics, such as shape, spread, and central tendency, facilitating a comprehensive analysis of associated differentials and enhancing study clarity and consistency.

The QR models and decomposition of the wage gap for the regular/salaried employees and casual workers included various individual characteristics, such as age, residential location, social group, marital status, educational levels (general and technical), and occupational characteristics (types of occupation and industry). Age and its squared term (divided by 100) were included to capture non-linear effects on wages (Bai and Veall, 2023; Si and Li, 2023). The National Classification of Occupations-2004 from the Directorate General of Employment and Training was aggregated into three major categories: white-collar (NCO 1 to 4), blue-collar (NCO 5 to 9), and agricultural (NCO 6) occupations (Hnatkovska et al., 2012). Similarly, the National Industrial Classification (MoSPI, 2008) was further grouped into five major types with NIC codes: Production and Extraction (NIC 1 to 3), Infrastructure and Utilities (NIC 4 to 6), Goods and Service Distribution (NIC 7 to 9), Knowledge and Service-Based (NIC 10 to 15), and Public and Social Services (NIC 16 to 20). These classifications encompass most regular/salaried and casual employment occupations and industries in India. The categorical variables were used as dummy variables in the regression modelling approach, with one category as the reference (Alkharusi, 2012). All analyses used Stata v. 13 (StataCorp, 2013) on Windows x64, incorporating sample weights for population representativeness.

After data pre-processing to eliminate missing wage/salary entries, 63 704 respondents (23.92% female, 76.08% male) were retained for model fitting. The table in the Annex presents the detailed distribution of male and female workers across various characteristics. Daily wage disparity persisted between genders across both regular/salaried employees and casual worker categories. Male regular/salaried workers earned a significantly higher average daily wage (Rs. 682.84) than females (Rs. 540.62). A similar disparity was observed for casual workers, with males earning Rs. 357.68 on average and females earning Rs. 225.47. A Mann-Whitney test showed significant gender wage disparities among India's regular/salaried and casual workers (p-value<0.001). Regular employees had a Gini coefficient of 0.450, indicating moderate income inequality. In contrast, casual workers had a more equitable distribution with a Gini coefficient of 0.234. Female regular employees had a higher wage disparity (Gini = 0.509) than male regular employees (Gini = 0.425). This difference was not seen among casual workers, where males (Gini = 0.210) and females (Gini = 0.201) had lower Gini coefficients than regular employees, suggesting a more equitable wage distribution, especially for females.

3.2 Wage effects across quantiles

QR models were separately estimated for male and female regular/salaried employees, as presented in Tables 1 and 2, and for male and female casual workers, as presented in Tables 3 and 4. The models utilised the daily wage's natural logarithm as the dependent variable and included the same independent predictors. Applying a logarithmic transformation to daily wages ensures that the estimated coefficients represent the change (percentage) in the daily wages log for a one-unit alteration in a continuous independent variable and the percentage difference in the log of daily wages between the reference category and the category in question for categorical independent variables. They were exponentiated to revert the coefficients to the original scale (e^{β}). The exponentiated coefficients were then interpreted as the percentage increase, as $(e^{\beta}-1)\times 100\%$, in the dependent variable attributable to the dummy variable. In cases where there was a decrease, it was interpreted as a percentage decrease, disregarding the sign. This method ensures that the interpretation of the coefficients remains consistent with the original scale of the dependent variable, thereby providing meaningful insights into the effects of the independent variables across different quantiles of the wage distribution.

The wage-age analysis revealed a concave relationship for all workers, with wages initially increasing before decelerating with age. This trend was more pronounced for females and in higher wage quartiles. The wage-age profile also varied across the wage distribution, particularly for male casual workers in the top half. Furthermore, the study found a significant urban wage premium for regular employees. Males experienced a 23.9% increase, while females had a 26.4% increase at the 10th percentile, which reversed at the 90th percentile to 13.5% (males) and 21.9% (females). Conversely, casual workers had a lower raw wage premium, ranging from 8.0% to 22.8% for males and 9.6% to 18.8% for females at higher percentiles. Regular employees, especially females, gained more across the wage spectrum than male casual workers. These findings highlight the gender-specific impact of urbanisation on wages across different wage levels and job types, supporting prior research on rural wage disparities, particularly for females (Dutta, 2006; Khanna, 2012).

The analysis revealed persistent wage disparities across caste groups. Scheduled Tribes (STs), Scheduled Castes (SCs), and Other Backward Classes (OBCs) earned less compared to the 'Others' category. Male STs experienced a 13.7% wage disadvantage among regular employees, while female STs displayed a surprising 19.1% wage advantage at the lowest income levels. SCs faced the most significant gaps, with male incomes lagging 5% to 10% and female incomes falling behind 6% to 11%. OBCs exhibited minor discrepancies. In the casual workforce, STs were the most disadvantaged, with male and female incomes falling behind the 'Others' category by 22.9% to 9.4% and 11% to 33.6% from the lower-to-upper wage spectrum, respectively. This suggests that STs faced greater disparities in casual employment, while SCs encountered more challenges in regular jobs. OBCs performed comparatively better in both job types. These findings underscore the persistence of wage gaps for socially disadvantaged groups in India, highlighting the complex interplay between worker type, gender, and income level, aligning with Madheswaran and Attewell's (2007) findings, who observed similar disparities among reserved categories.

Table 1 Quantile Regression Model for male regular/salaried employees

Independent variables	**Percentiles**				
	θ = 0.10	**θ = 0.25**	**θ = 0.50**	**θ = 0.75**	**θ = 0.90**
Age	0.049*** (0.01)	0.026*** (0.01)	0.015** (0.01)	0.014*** (0.00)	0.024*** (0.01)
$Age^2/100$	–0.053*** (0.01)	–0.018* (0.01)	0.002 (0.01)	0.008 (0.01)	–0.000 (0.01)
Residential area (rural)					
Urban	0.214*** (0.02)	0.191*** (0.02)	0.177*** (0.01)	0.152*** (0.01)	0.127*** (0.02)
Social group (others)					
STs	–0.147** (0.05)	–0.080* (0.03)	–0.032 (0.02)	0.011 (0.02)	0.038 (0.03)
SCs	–0.104*** (0.03)	–0.065** (0.02)	–0.060*** (0.02)	–0.055** (0.02)	–0.057** (0.02)
OBCs	–0.053* (0.02)	–0.055*** (0.02)	–0.052** (0.02)	–0.045** (0.02)	–0.035 (0.02)
Marital status (never married)					
Currently married	0.167*** (0.03)	0.153*** (0.02)	0.135*** (0.02)	0.132*** (0.02)	0.098*** (0.02)
Widowed	0.076 (0.04)	–0.030 (0.12)	0.006 (0.08)	0.033 (0.04)	–0.017 (0.07)
Divorced/separated	–0.117 (0.24)	–0.025 (0.15)	–0.117*** (0.03)	–0.063 (0.13)	0.013 (0.23)
General education (illiterate)					
No formal schooling	0.150* (0.06)	0.047 (0.05)	0.183*** (0.04)	–0.032 (0.04)	–0.253 (0.48)
Up to primary school	0.022 (0.05)	0.059 (0.04)	0.082** (0.03)	0.087*** (0.03)	0.141** (0.04)
Middle school	0.133** (0.05)	0.158*** (0.04)	0.241*** (0.02)	0.244*** (0.02)	0.286*** (0.03)
Secondary school	0.281*** (0.05)	0.280*** (0.04)	0.332*** (0.03)	0.364*** (0.03)	0.399*** (0.03)
Higher secondary	0.274*** (0.05)	0.286*** (0.04)	0.393*** (0.03)	0.454*** (0.03)	0.529*** (0.03)
Graduate	0.439*** (0.05)	0.477*** (0.04)	0.613*** (0.03)	0.693*** (0.03)	0.708*** (0.03)
Postgraduate & above	0.598*** (0.06)	0.766*** (0.06)	0.903*** (0.04)	0.940*** (0.04)	0.951*** (0.04)
Technical education (not received)					
Have technical edu.	0.254*** (0.04)	0.235*** (0.03)	0.236*** (0.03)	0.191*** (0.02)	0.249*** (0.03)
Occupation (agricultural)					
White-collar	0.225 (0.16)	0.383*** (0.04)	0.352*** (0.03)	0.322*** (0.09)	0.330*** (0.03)
Blue-collar	0.038 (0.16)	0.110** (0.04)	0.052* (0.02)	0.026 (0.08)	0.033 (0.02)
Industry (public & social services)					
Production & extraction	0.260*** (0.04)	0.176*** (0.03)	0.005 (0.03)	–0.020 (0.02)	–0.043 (0.02)
Infrastructure & utilities	0.378*** (0.04)	0.264*** (0.04)	0.130*** (0.04)	0.097** (0.03)	0.167*** (0.05)
Goods & services dist.	0.171*** (0.04)	0.105*** (0.03)	–0.056* (0.03)	–0.082*** (0.02)	–0.058* (0.03)
Knowledge & service based	0.323*** (0.04)	0.285*** (0.03)	0.190*** (0.03)	0.220*** (0.02)	0.254*** (0.02)
Intercept	3.720*** (0.21)	4.375*** (0.11)	4.942*** (0.10)	5.262*** (0.11)	5.285*** (0.10)
pseudo-R^2	0.133	0.181	0.268	0.348	0.343

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

Table 2 Quantile Regression Model for female regular/salaried employees

Independent variables	Percentiles				
	θ = 0.10	θ = 0.25	θ = 0.50	θ = 0.75	θ = 0.90
Age	0.082*** (0.02)	0.068*** (0.01)	0.044*** (0.01)	0.011 (0.01)	0.015 (0.01)
$Age^2/100$	–0.096*** (0.02)	–0.076*** (0.02)	–0.040** (0.01)	0.005 (0.02)	0.009 (0.02)
Residential area (rural)					
Urban	0.234*** (0.04)	0.282*** (0.03)	0.304*** (0.03)	0.254*** (0.03)	0.198*** (0.03)
Social group (others)					
STs	0.175** (0.06)	0.037 (0.06)	0.028 (0.04)	–0.011 (0.06)	–0.010 (0.11)
SCs	–0.064 (0.06)	–0.118*** (0.03)	–0.085* (0.04)	–0.067* (0.03)	–0.122** (0.05)
OBCs	0.030 (0.04)	–0.076* (0.03)	–0.033 (0.03)	–0.093*** (0.03)	–0.154*** (0.04)
Marital status (never married)					
Currently married	0.000 (0.05)	0.120*** (0.04)	0.160*** (0.04)	0.226*** (0.04)	0.097* (0.04)
Widowed	0.150* (0.07)	0.228*** (0.06)	0.258*** (0.06)	0.316*** (0.05)	0.166** (0.06)
Divorced/separated	–0.030 (0.10)	0.055 (0.07)	–0.102 (0.07)	0.178 (0.19)	0.084 (0.14)
General education (illiterate)					
No formal schooling	0.229 (0.19)	–0.149 (0.11)	0.373*** (0.04)	0.015 (0.79)	0.169 (0.14)
Up to primary school	0.164 (0.09)	0.094* (0.04)	0.170*** (0.05)	0.160* (0.08)	0.187 (0.10)
Middle school	0.242** (0.09)	0.357*** (0.04)	0.332*** (0.04)	0.305*** (0.06)	0.300** (0.11)
Secondary school	0.488*** (0.08)	0.554*** (0.05)	0.552*** (0.05)	0.494*** (0.05)	0.498*** (0.12)
Higher secondary	0.644*** (0.10)	0.647*** (0.04)	0.701*** (0.06)	0.730*** (0.08)	0.793*** (0.12)
Graduate	0.853*** (0.09)	0.934*** (0.06)	1.055*** (0.06)	1.163*** (0.07)	1.159*** (0.11)
Postgraduate & above	1.126*** (0.10)	1.280*** (0.08)	1.508*** (0.07)	1.500*** (0.07)	1.377*** (0.11)
Technical education (not received)					
Have technical edu.	0.318*** (0.07)	0.328*** (0.05)	0.328*** (0.04)	0.267*** (0.04)	0.255*** (0.06)
Occupation (agricultural)					
White-collar	0.192 (0.51)	0.228 (0.45)	0.129 (0.13)	0.146 (0.08)	0.375* (0.15)
Blue-collar	–0.041 (0.51)	0.027 (0.45)	–0.148 (0.13)	–0.200** (0.08)	–0.046 (0.15)
Industry (public & social services)					
Production & extraction	0.648*** (0.05)	0.528*** (0.03)	0.465*** (0.03)	0.362*** (0.03)	0.274*** (0.03)
Infrastructure & utilities	0.492*** (0.04)	0.376** (0.13)	0.692*** (0.11)	0.603*** (0.08)	0.479*** (0.04)
Goods & services dist.	0.429*** (0.06)	0.437*** (0.04)	0.386*** (0.04)	0.299*** (0.04)	0.241*** (0.04)
Knowledge & service based	0.413*** (0.04)	0.486*** (0.04)	0.500*** (0.04)	0.406*** (0.03)	0.357*** (0.04)
Intercept	2.317*** (0.61)	2.768*** (0.52)	3.607*** (0.24)	4.647*** (0.25)	4.825*** (0.28)
pseudo-R^2	0.211	0.238	0.298	0.372	0.353

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

Our study also identified a marriage premium for married regular employees of both genders compared to unmarried individuals. The premium was more pronounced for low-income males, decreasing from 18.2% at the 10th percentile to 10.3% at the 90th. Married females also experienced a wage advantage, ranging from 12.7% to 25.4%, with a higher premium observed in the upper-income bracket. Conversely, widowed and divorced/separated individuals did not experience significant effects, except for widowed females, who saw a larger premium, ranging from 16.2% to 37.2%. This suggests a potential role of job security in the form of wage advantage for widowed females. Interestingly, among casual workers, marital status had the opposite effect. Married males experienced a wage decrease, ranging from 4.1% to 15.2% at higher percentiles, while married females saw a positive effect at the same percentiles. These findings support Spence's (1974) theory that marriage, especially for employees in regular/salaried employment, can signal stability and lead to a wage premium, with variations based on gender and employment type.

QR models revealed a significant positive effect of education levels on wages. Individuals with higher education, including those with informal schooling, consistently earned more than those without formal education. Among regular/salaried workers, male higher-secondary school graduates enjoyed a wage premium of 31.5% to 69.7%, while females experienced a larger premium of 90.4% to 121.0%. Postgraduates had even greater advantages, with males experiencing a premium of 81.8% to 158.8% and females a premium of 208.3% to 296.3%. Education also positively affected casual workers, primarily influencing male wages, while only informal education significantly impacted female casual workers' wages. Among male casual workers, postgraduate degrees offered the highest advantage, ranging from 10% to 26.6%. These findings underscore the significance of education for regular/salaried employees, particularly females who benefit significantly. However, for casual workers, especially females, educational attainment had minimal wage impact, suggesting a need for further research and targeted interventions. This aligns with Chakraborty and Mukherjee's (2014) study, which found that education equalised wages for both genders, enabling females to earn more. In regular employment, technical education increased earnings for both genders, particularly lower-income females. However, this trend was less pronounced for casual workers, where only males significantly benefited from technical education.

The study revealed significant wage variations across occupations, employment types, and genders. For regular/salaried employees, white-collar occupations had a more substantial and consistent positive impact on males across percentiles, with 46.7% at the 25th percentile decreasing to 39.1% at the 90th percentile, compared to females, where the effect was significant only at the 90th percentile. Blue-collar occupations showed a negative impact of 18.1% for females at the 75th percentile, while the impact was more negligible and less consistent for males, with 11.6% at the 25th percentile.

Among casual workers, males experienced significant negative impacts from both white-collar (19.3% at the 10th percentile) and blue-collar occupations, particularly at the lower (19.3% at the 10th and upper (20.7% at the 90th) percentiles. For females, the impact of white-collar occupations was mixed, with a positive impact at the 10th percentile (24.0%) and 90th percentile (23.5%) but negative at the median (20.0%), while blue-collar occupations consistently showed negative impacts, with significant decreases at the 25th and 75th percentiles (16.9% and 14.2%, respectively). Gender differences in the two occupational categories' impact on earnings compared to agricultural occupations were evident in both employment categories. Regular/salaried males benefited more from white-collar occupations than females, while casual males faced more negative impacts from both occupation types compared to females.

Table 3 Quantile Regression Model for male casual workers

Independent variables	Percentiles				
	θ = 0.10	**θ = 0.25**	**θ = 0.50**	**θ = 0.75**	**θ = 0.90**
Age	0.010*** (0.00)	0.014*** (0.00)	0.027*** (0.00)	0.031*** (0.00)	0.043*** (0.00)
$Age^2/100$	−0.012*** (0.00)	−0.016*** (0.00)	−0.029*** (0.00)	−0.030*** (0.01)	−0.041*** (0.01)
Residential area (rural)					
Urban	0.077*** (0.02)	0.074*** (0.01)	0.124*** (0.01)	0.163*** (0.01)	0.205*** (0.01)
Social group (others)					
STs	−0.260*** (0.01)	−0.170*** (0.01)	−0.137*** (0.02)	−0.119*** (0.03)	−0.099*** (0.03)
SCs	−0.004 (0.01)	0.006 (0.01)	−0.001 (0.01)	−0.026 (0.02)	−0.032 (0.03)
OBCs	−0.024*** (0.01)	0.000 (0.01)	0.024 (0.01)	0.056** (0.02)	0.081** (0.03)
Marital status (never married)					
Currently married	−0.003 (0.01)	0.002 (0.01)	−0.042* (0.02)	−0.084*** (0.02)	−0.165*** (0.02)
Widowed	−0.008 (0.02)	−0.040 (0.08)	−0.113 (0.06)	−0.060 (0.08)	−0.186*** (0.03)
Divorced/separated	−0.312*** (0.03)	−0.113 (0.18)	−0.011 (0.04)	−0.087* (0.04)	−0.079* (0.04)
General education (illiterate)					
No formal schooling	0.005 (0.04)	−0.072 (0.08)	0.022 (0.09)	−0.060* (0.03)	0.012 (0.08)
Up to primary school	0.005 (0.01)	0.017* (0.01)	0.040** (0.01)	0.092*** (0.02)	0.120*** (0.02)
Middle school	0.018 (0.01)	0.035*** (0.01)	0.058*** (0.01)	0.133*** (0.02)	0.172*** (0.02)
Secondary school	0.030** (0.01)	0.075*** (0.01)	0.114*** (0.01)	0.187*** (0.02)	0.216*** (0.03)
Higher secondary	0.021** (0.01)	0.025 (0.02)	0.072*** (0.02)	0.160*** (0.02)	0.197*** (0.04)
Graduate	0.042*** (0.01)	0.087*** (0.01)	0.094* (0.04)	0.127** (0.04)	0.157** (0.05)
Postgraduate & above	0.100 (0.15)	0.025 (0.03)	0.098*** (0.02)	0.264*** (0.05)	0.266*** (0.04)
Technical education (not received)					
Have technical edu.	0.131 (0.07)	0.190*** (0.02)	0.292*** (0.07)	0.407*** (0.06)	0.300*** (0.05)
Occupation (agricultural)					
White-collar	−0.215*** (0.04)	−0.085 (0.05)	−0.101 (0.12)	0.063 (0.10)	0.021 (0.04)
Blue-collar	−0.215*** (0.01)	−0.116*** (0.03)	−0.104* (0.04)	−0.178* (0.09)	−0.232*** (0.02)
Industry (public & social services)					
Production & extraction	−0.118*** (0.01)	−0.178** (0.05)	−0.150** (0.06)	−0.101*** (0.02)	0.001 (0.07)
Infrastructure & utilities	0.134*** (0.01)	0.073 (0.05)	0.054 (0.05)	0.077*** (0.02)	0.181* (0.07)
Goods & services dist.	0.037* (0.02)	0.023 (0.05)	0.012 (0.06)	0.064 (0.03)	0.161* (0.07)
Knowledge & service based	−0.108 (0.22)	−0.013 (0.04)	−0.005 (0.09)	−0.045 (0.04)	−0.045 (0.08)
Intercept	5.417*** (0.05)	5.443*** (0.07)	5.356*** (0.09)	5.449*** (0.11)	5.364*** (0.10)
pseudo-R^2	0.101	0.125	0.110	0.085	0.128

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

Table 4 Quantile Regression Model for female casual workers

Independent variables	Percentiles				
	θ = 0.10	**θ = 0.25**	**θ = 0.50**	**θ = 0.75**	**θ = 0.90**
Age	0.008 (0.01)	0.002 (0.00)	−0.000 (0.02)	0.017** (0.01)	0.010 (0.01)
$Age^2/100$	−0.013 (0.01)	−0.003 (0.01)	0.000 (0.02)	−0.026*** (0.01)	−0.013 (0.01)
Residential area (rural)					
Urban	−0.003 (0.02)	0.092** (0.03)	0.182 (0.15)	0.172*** (0.02)	0.149*** (0.01)
Social group (others)					
STs	−0.117* (0.06)	−0.015 (0.06)	−0.105 (0.17)	−0.214*** (0.04)	−0.409*** (0.06)
SCs	0.033 (0.04)	0.090 (0.06)	0.000 (0.08)	−0.148*** (0.02)	−0.275*** (0.05)
OBCs	−0.054 (0.05)	−0.006 (0.06)	−0.000 (0.07)	−0.153*** (0.02)	−0.266*** (0.05)
Marital status (never married)					
Currently married	−0.031 (0.04)	−0.003 (0.04)	0.000 (0.09)	0.057** (0.02)	0.092** (0.03)
Widowed	−0.043 (0.05)	−0.006 (0.05)	0.000 (0.10)	0.034 (0.03)	0.080* (0.04)
Divorced/separated	−0.000 (0.04)	−0.008 (0.09)	0.000 (0.16)	0.053 (0.11)	0.235 (0.25)
General education (illiterate)					
No formal schooling	0.279 (0.16)	0.094*** (0.03)	−0.118 (.)	−0.291** (0.09)	−0.176 (0.18)
Up to primary school	−0.049* (0.02)	−0.005 (0.02)	−0.000 (0.05)	−0.010 (0.02)	0.010 (0.01)
Middle school	−0.027 (0.03)	−0.004 (0.03)	−0.000 (0.05)	−0.055 (0.04)	−0.003 (0.01)
Secondary school	−0.091 (0.11)	−0.012 (0.04)	−0.000 (0.10)	0.000 (0.02)	0.021 (0.08)
Higher secondary	−0.036 (0.06)	−0.001 (0.07)	−0.000 (0.12)	−0.006 (0.14)	0.108 (0.07)
Graduate	0.021 (0.07)	0.088 (0.15)	−0.223 (0.15)	−0.107 (0.12)	−0.085 (0.12)
Postgraduate & above	0.277 (1.13)	0.173 (.)	−0.105 (0.78)	−0.178 (0.62)	−0.731 (1.75)
Technical education (not received)					
Have technical edu.	0.001 (0.52)	−0.006 (0.17)	0.329 (0.82)	0.172 (0.18)	0.809 (1.03)
Occupation (agricultural)					
White-collar	0.215 (0.87)	−0.094 (0.09)	−0.223 (0.18)	−0.123 (0.46)	0.211 (0.23)
Blue-collar	−0.077 (0.05)	−0.185*** (0.03)	−0.223 (0.18)	−0.153*** (0.02)	−0.004 (0.02)
Industry (public & social services)					
Production & extraction	0.338*** (0.07)	−0.001 (0.11)	−0.000 (0.05)	−0.162 (0.10)	−0.157*** (0.02)
Infrastructure & utilities	0.449*** (0.06)	0.185 (0.11)	0.223 (0.13)	0.057 (0.10)	0.108** (0.04)
Goods & services dist.	0.550*** (0.08)	0.181 (0.12)	0.318** (0.11)	0.102 (0.10)	0.098 (0.12)
Knowledge & service based	0.746 (0.41)	0.414* (0.17)	0.247 (0.58)	−0.025 (0.13)	−0.169* (0.09)
Intercept	4.628*** (0.16)	5.185*** (0.16)	5.521*** (0.29)	5.671*** (0.15)	5.861*** (0.15)
pseudo-R^2	0.021	0.049	0.033	0.072	0.111

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors are reported in parentheses, estimated using 300 iterations. Reference categories for independent variables are indicated in parentheses.
Source: Author's calculation based on PLFS (2020–21) dataset using Stata v.13

The public and social services sectors served as a reference when comparing industry types for employees. Among male regular/salaried employees, the impact of working in the Production and Extraction industries was only significant at lower percentiles, i.e., 29.7% at the 10^{th} and 19.2% at the 25^{th} percentiles. For females, the impact remained consistently positive throughout the wage distribution, starting at 91.2% at the 10^{th} percentile, which decreased to 31.5% at the 90^{th} percentile. For males in the Infrastructure & Utilities industries, the effect was positive across all percentiles, beginning at 45.9% at the 10^{th} percentile, which decreased to 18.2% at the 90^{th} percentile. Similarly, the effect was positive for females throughout the wage distribution, with an exceptionally positive impact of 99.8% at the median. In the Goods & Services Distribution industries, the impact for males was positively higher at lower percentiles (18.6% at the 10^{th} percentile) but became negative at higher percentiles, reaching –5.6% at the 90^{th} percentile. This suggested that males earned more than those in Public & Social Services at lower percentiles but less at higher percentiles. This effect remained consistently positive for females, starting at 53.6% at the 10^{th} percentile but decreasing to 27.3% at the 90^{th} percentile. For males in Knowledge & Service Based industries, the impact was positive across all percentiles, starting at 38.1% at the 10^{th} percentile and reaching 28.9% at the 90^{th} percentile. The effect was also positive for females, starting at 51.1% at the 10^{th} percentile and decreasing to 42.9% at the 90^{th} percentile. These results indicated that both male and female regular/salaried employees in these industries earned more than those in Public & Social Services, with variations across percentiles. Females tended to experience higher positive impacts compared to males, particularly in the Production & Extraction and Goods & Services Distribution industries. For male casual workers in the Production & Extraction industries, effects were negative across all percentiles, with the most significant decline of 16.3% at the 25^{th} percentile and a slight positive effect of 0.1% at the 90^{th} percentile. The effects were positive for females, with the highest increase of 40.2% at the 10^{th} percentile and a decrease of 14.5% at the 90^{th} percentile. In the Infrastructure & Utilities industries, male casual workers experienced mostly positive effects, with the highest increase of 19.8% at the 90^{th} percentile and a significant positive effect of 14.3% at the 10^{th} percentile. Female workers also saw positive effects at lower percentiles, with the highest impact of 56.7% at the 10^{th} percentile, which, however, decreased to 10.3% at the 90^{th} percentile. Among male casual workers in the Goods & Services Distribution industries, the effects were positive, with the highest increase of 17.5% at the 90^{th} percentile, in contrast to females, where the highest advantage was 73.3% at the 10^{th} percentile. In the Knowledge & Service-Based industries, male casual workers faced insignificant negative effects across all percentiles, and the results of female casual workers in those industries were mixed, i.e., major advantage at the 25^{th} percentile but negative at the 90^{th} percentile of the wage distribution. Overall, female casual workers experienced more positive effects from industry variables compared to male casual workers, particularly at lower percentiles. Thus, the differences in impact across percentiles indicated that industry effects varied significantly between genders and across different income levels.

The analysis of pseudo-R^2 values revealed an improved model fit at higher quantiles, indicating better predictions for higher response percentiles. Both regular employee groups (male and female) displayed higher pseudo-R^2 values than casual workers, suggesting a superior model fit for regular employees. Specifically, male regular employees had values ranging from 0.133 to 0.343, while females ranged from 0.211 to 0.353. In contrast, casual workers showed lower values, with males ranging from 0.101 to 0.128 and females from 0.021 to 0.111. However, the QR models for regular females and casual males had slightly higher average pseudoR^2 values, suggesting the potential for marginally better fits in these specific subgroups.

The fitted QR models effectively yielded statistically significant and accurate parameter estimates even under non-normal error terms and heteroscedasticity, aligning with Chen and Chalhoub-Deville (2014). Notably, a substantial portion of the significant estimates from both models exhibited exceptionally small standard errors, some even approaching or exceeding the precision of the median. This finding supports the suitability of QR models for analysing wage data, which often features non-normality and outliers, where traditional methods might be less reliable.

3.3 Decomposition of wage differentials

The results from Table 5, obtained through the OB and MMM decomposition, show a statistically significant male worker advantage over females. This is indicated by positive coefficients in the differences, coefficient, and characteristics components for both regular/salaried employees and casual workers.

Table 5 Wage differential decomposition across genders for regular/salaried employees and casual workers

Components	OB	MMM				
		θ = 0.10	θ = 0.25	θ = 0.50	θ = 0.75	θ = 0.90
			Regular/salaried employees			
Difference	0.458*** (0.02)	0.709*** (0.01)	0.593*** (0.01)	0.473*** (0.01)	0.315*** (0.01)	0.161*** (0.01)
Characteristics	(0.104***) (0.02)	0.171*** (0.04)	0.152*** (0.02)	0.105*** (0.02)	0.040*** (0.03)	0.023* (0.03)
Coefficients	(0.354***) (0.02)	0.538*** (0.03)	0.441*** (0.02)	0.367*** (0.02)	0.271*** (0.03)	0.137*** (0.03)
			Casual workers			
Difference	0.457*** (0.01)	0.349*** (0.01)	0.455*** (0.01)	0.493*** (0.01)	0.471*** (0.01)	0.498*** (0.01)
Characteristics	(0.093***) (0.01)	0.034*** (0.02)	0.081*** (0.02)	0.138*** (0.03)	0.142*** (0.01)	0.145*** (0.02)
Coefficients	(0.364***) (0.01)	0.315*** (0.02)*	0.374*** (0.02)	0.355*** (0.01)	0.329*** (0.01)	0.353*** (0.02)

Notes: * for p<.05, ** for p<.01, *** for p<.001. Standard errors using bootstrapping are reported in parentheses.
Source: Author's calculation based on PLFS 2020–21 dataset using Stata v.13

The OB decomposition analysis revealed a noteworthy wage differential between male and female workers in both employment types (regular: 0.458, casual: 0.457). The raw wage gap, expressed as a percentage, was 58.1% for regular/salaried employees and 57.9% for casual workers. Worker characteristics explained a portion of these gaps (11% in regular, 9.7% in casual employment). However, discrimination emerged as the primary factor in both groups, accounting for 42.5% and 43.9% of the unexplained raw wage gap among regular/salaried and casual employees, respectively. Notably, the level of discrimination appeared similar across both worker types despite minor differences in explained factors. However, given the potential limitations of the OB decomposition, such as its possible inability to capture nuanced discrimination due to non-linear relationships and interactions between variables, further analysis was conducted using the MMM decomposition.

The MMM decomposition, employing QR methodology, investigated the gender wage disparity across different percentiles. The examination of regular/salaried workers showed a significant variation in the wage disparity across the wage range. The disparity decreased from 0.709 at the 10^{th} to 0.160 at the 90^{th} percentile. This translated into a significant advantage for males, with a raw gap of 103.2% at the 10^{th} percentile, narrowing to 17.4% at the 90^{th} percentile. The observed wage distribution pattern suggested a "sticky floor" phenomenon, leading to more pronounced wage gaps for female workers, particularly at lower income levels, in contrast to male workers. These findings are consistent with prior research indicating heightened discriminatory practices against females in low-wage occupations

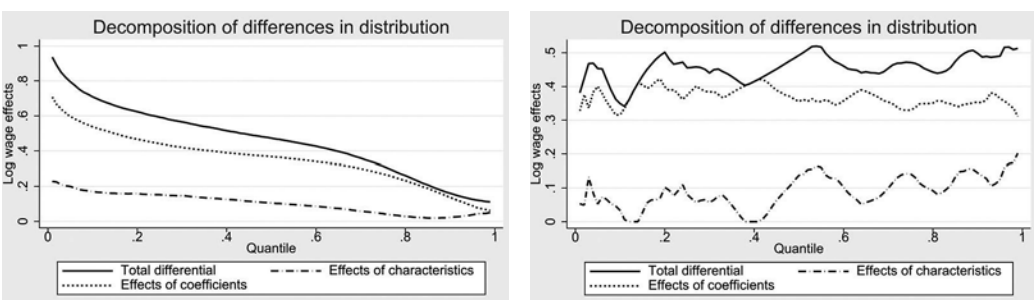
(Chakraborty and Mukherjee, 2014; Das, 2018). Discrimination was most pronounced at the 10th percentile, reaching 0.538, corresponding to a raw wage gap of 71.3%. Discrimination remained significant even at the median (0.473, 60.5% raw gap). Although the overall gap decreased toward the higher percentiles, discrimination persisted. High earners still encountered a 31.1% gap (75th percentile) and a 14.7% gap (90th percentile) due to discrimination in raw terms. The study also examined how characteristics contribute to the wage gap, revealing that their maximum contribution was observed at the 10th percentile, accounting for 18.6% of the overall gap. However, their influence diminished rapidly thereafter, reaching only 2.3% at the 90th percentile. This suggests that for high earners, discrimination alone could explain all the remaining wage disparity. Figure 1(a) replicated previous findings, illustrating a narrowing wage gap with increasing income. Lower-income quantiles exhibited larger disparities, influenced by both observed and unobserved characteristics, with the latter playing a more prominent role across all income strata.

The examination of the gender wage differentials among casual workers using the MMM decomposition showed a clear contrast with regular/salaried employees. The overall gap was lowest at the 10th percentile (0.349) but increased steadily towards the 90th percentile (0.498), with a slight dip at the 75th percentile (0.471). This translated to a raw wage gap of 41.8% at the 10th percentile, peaking at 64.5% at the 90th percentile before dropping back to 60.2% at the 75th percentile. Like regular workers, the gap for casual workers is statistically significant at all percentiles (10th to 90th). However, unlike the "sticky floor" effect or the "glass ceiling" barrier observed for regular workers, the casual worker gap exhibited a unique upward trend, indicating a potentially different form of discrimination females face in casual work. The unexplained wage differential component, which may indicate potential bias, was the dominant factor across all income percentiles. This unexplained component increased from 0.315 at the 10th percentile (37% of the raw gap) to 0.353 at the 90th percentile (42.3% of the raw gap). In contrast, the explained portion, attributed to differences in worker characteristics, was significantly smaller, ranging from 0.034 at the 10th percentile (3.5% of the raw gap) to 0.145 at the 90th percentile (15.6% of the raw gap). This suggests that discrimination plays a dominant role in the wage gap, particularly for higher-wage earners. Further research is necessary to comprehensively understand the specific forms of discrimination experienced by females in casual work, alongside the systemic and cultural mechanisms that perpetuate such inequities. Figure 1(b) supports this interpretation, revealing nuanced differences in the decomposition. While unexplained factors remain significant, their relative contribution remains higher than observed for regular workers. This finding suggests that workers' characteristics have a more significant impact on elucidating wage disparity in casual employment.

Figure 1 Plot of the MMM decomposition results for the male and female wage differential

a) Regular/salaried employees

b) Casual workers



Source: Author's calculation based on PLFS 2020–21 dataset using Stata v.13

The utilisation of the MMM decomposition, based on the QR framework and implemented across various percentiles, demonstrated statistically significant reductions in standard errors for both coefficient estimates and raw wage differences compared to those derived from the OLS-based OB decomposition. This increased accuracy significantly strengthened the robustness of the results, thereby enhancing the credibility and trustworthiness of the conclusions regarding the extent and factors influencing the identified gender pay disparity.

CONCLUSION

The QR modelling approach unveiled key factors influencing wage structures for male and female workers, distinguished by employment type (regular vs casual). Socioeconomic indicators, such as social group, marital status, and education, significantly impacted wages across income levels. Urban workers, irrespective of gender or employment type, earned more than their rural counterparts. Lower social groups faced wage disadvantages compared to higher ones, particularly SCs in regular employment and STs in casual work. Marriage positively impacted regular employees' wages, with widowed females even experiencing an unexpected advantage, although this effect was less pronounced for casual workers. Education strongly improves wages, especially for regular employees, with higher qualifications leading to more significant gains than casual workers. Gender interacted with education, benefiting regular female employees more from higher qualifications than males, a trend not observed for casual female workers. White-collar jobs offered the highest wages for regular males, while blue-collar jobs also positively influenced their earnings. Female regular employees showed no significant differences across occupational categories. Agriculture was the most lucrative sector for casual workers, while white-collar work offered no advantage, and blue-collar jobs were associated with lower wages than agriculture. Specific industries further complicated the picture, with some offering wage premiums and others leading to lower earnings for both employment types. These findings underscore the complex interrelationship of social, economic, and sectoral factors in shaping wage structures for diverse worker segments.

The MMM analysis revealed significant gender wage gaps among regular/salaried and casual workers across all percentiles. For regular/salaried employees, the largest gaps were observed at lower percentiles, ranging from 103.2% to 17.4%, indicating a "sticky floor" effect. This suggests that females in lower wage brackets face obstacles to career advancement due to discriminatory practices. Worker characteristics played a minor role, with discrimination being the main driver. Among casual workers, the wage gap was significant across all percentiles, ranging from 41.8% to 64.5%. However, no clear "sticky floor" or "glass ceiling" pattern indicated pervasive discrimination at all wage levels. The analysis attributes the wage gap among casual workers primarily to gender-based discrimination rather than differences in characteristics.

Our study explored the potential of QR methodology to unveil non-linear wage patterns across income distribution for both regular/salaried and casual workers, irrespective of gender. Compared to traditional linear regression, QR models in the study have performed in several key areas. First, QR demonstrates greater adaptability to workforce heterogeneity, accommodating non-linear relationships between wages and explanatory variables. Second, it demonstrates robustness to outliers inherent in wage data, yielding more precise and reliable estimates. These findings suggest QR's superior applicability for labour market analysis. QR can inform targeted policy development by enhancing our understanding of wage structures. Further insights into the gender wage gap can be gained through the MMM decomposition within the QR framework. In contrast to OB decomposition, our study found that the MMM resulted in smaller standard errors and enabled the examination of wage distribution across different percentiles. This allowed for a more detailed understanding of the wage gap, highlighting possible "glass-ceiling" or "sticky-floor" effects within the distribution. This solidifies its appeal for investigating wage discrimination in future research as well.

This study's findings emphasise the importance of accurately selecting appropriate methodologies to measure and analyse wage disparities in the labour market. Applying QR and MMM decomposition techniques demonstrates their potential as a robust framework for future research in this domain. These methods offer powerful tools to dissect the multifaceted nature of wage inequality, including the gender wage gap. QR and the MMM decomposition pave the way for developing more informed policy interventions to achieve greater labour market equity by facilitating a deeper understanding of these complex phenomena.

ACKNOWLEDGMENTS

Samapriya Trivedi is a recipient of the Indian Council of Social Science Research Doctoral Fellowship. His article is largely an outcome of his doctoral work sponsored by ICSSR. However, the responsibility for the facts stated, opinions expressed and the conclusions drawn are entirely that of the authors.

References

ADIREKSOMBAT, K., FANG, Z., SAKELLARIOU, C. (2010). The Evolution of gender wage Differentials and Discrimination in Thailand: 1991–2007 An application of Unconditional Quantile Regression [online]. *MPRA Paper*. <https://ideas.repec.org/p/pra/mprapa/27516.html>.

AGRAWAL, T. (2013). Gender and caste–based wage discrimination in India: Some recent evidence [online]. *Journal for Labour Market Research*, 47(4): 329–340. <https://doi.org/10.1007/s12651–013–0152-z>.

ALKHARUSI, H. (2012). Categorical variables in regression analysis: a comparison of dummy and effect coding [online]. *International Journal of Education*, 4(2): 202. <https://doi.org/10.5296/ije.v4i2.1962>.

ARELLANO, M., BONHOMME, S. (2017). Quantile selection models with an application to understanding changes in wage inequality [online]. *Econometrica*, 85(1): 1–28. <https://doi.org/10.3982/ecta14030>.

ARMSTRONG, R. D., FROME, E. L., KUNG, D. S. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem [online]. *Communications in Statistics. Simulation and Computation*, 8(2): 175–190. <https://doi.org/10.1080/03610917908812113>.

AZAM, M. (2012). Changes in wage structure in urban India, 1983–2004: a quantile regression decomposition [online]. *World Development*, 40(6): 1135–1150. <https://doi.org/10.1016/j.worlddev.2012.02.002>.

AZAM, M., PRAKASH, N. (2015). A distributional analysis of public-private wage differential in India [online]. *LABOUR*, 29(4): 394–414. <https://doi.org/10.1111/labr.12068>.

BAI, Y., VEALL, M. R. (2023). Minimum wages and mental health: Evidence from Canada [online]. *SSM. Mental Health*, 3: 100187. <https://doi.org/10.1016/j.ssmmh.2023.100187>.

BIEWEN, M., ERHARDT, P. (2021). Arhomme: An implementation of the Arellano and Bonhomme (2017) estimator for quantile regression with selection correction [online]. *The Stata Journal*, 21(3): 602–625. <https://doi.org/10.1177/1536867x211045516>.

CHAKRABORTY, S., MUKHERJEE, S. (2014). Gender wage gap in the Indian labour market: Evidence from the NSS 66th round data. *The Indian Journal of Labour Economics*, 57(2): 259–280.

CHEN, F., CHALHOUB-DEVILLE, M. (2014). Principles of quantile regression and an application [online]. *Language Testing*, 31(1): 63–87. <https://doi.org/10.1177/0265532213493623>.

DAS, P. (2018). Wage gap and employment status in Indian labour market: Quantile based counterfactual analysis. *World Journal of Applied Economics*, 4(2): 117–134. <https://doi.org/10.22440/wjae.4.2.4>.

DESHPANDE, A., GOEL, D., KHANNA, S. (2017). Bad Karma or Discrimination? Male–Female Wage Gaps Among Salaried Workers in India [online]. *World Development*, 102: 331–344. <https://doi.org/10.1016/j.worlddev.2017.07.012>.

DESHPANDE, A., SHARMA, S. (2015). Disadvantage and discrimination in self-employment: Caste gaps in earnings in Indian small businesses [online]. *Small Business Economics*, 46(2): 325–346. <https://doi.org/10.1007/s11187–015–9687–4>.

DURAISAMY, M., DURAISAMY, P. (2016). Gender wage gap across the wage distribution in different segments of the Indian Labour Market, 1983–2012: Exploring the glass ceiling or sticky floor phenomenon [online]. *Applied Economics*, 48(43): 4098–4111. <https://doi.org/10.1080/00036846.2016.1150955>.

DUTTA, P. V. (2006). Returns to education: New evidence for India, 1983–1999 [online]. *Education Economics*, 14(4): 431–451. <https://doi.org/10.1080/09645290600854128>.

FITZENBERGER, B., KOENKER, R., MACHADO, J., MELLY, B. (2021). Economic applications of quantile regression 2.0 [online]. *Empirical Economics*, 62(1): 1–6. <https://doi.org/10.1007/s00181–021–02186–1>.

GAZELEY, I., HOLMES, R., NEWELL, A., REYNOLDS, K. P., RUFRANCOS, H. G. (2018). Inequality among European working households, 1890–1960. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.3137485>.

HNATKOVSKA, V., LAHIRI, A., PAUL, S. (2012). Castes and labour mobility [online]. *American Economic Journal: Applied Economics*, 4(2): 274–307. <https://doi.org/10.1257/app.4.2.274>.
HUANG, Q., ZHANG, H., CHEN, J., HE, M. (2017). Quantile Regression Models and their Applications: a Review [online]. *Journal of Biometrics & Biostatistics*, 8(3). <https://doi.org/10.4172/2155–6180.1000354>.
INDIA WAGE REPORT (2018). *International Labour Organisation* [online]. [cit. 25.12.2023]. <https://www.ilo.org/wcmsp5/groups/public/–––asia/–––ro–bangkok/–––sro–new_delhi/documents/publication/wcms_638305.pdf>.
JENKINS, S. P. (1999). INEQDECO: Stata module to calculate inequality indices with decomposition by subgroup [online]. *Statistical Software Components*. <https://ideas.repec.org/c/boc/bocode/s366002.html>.
KHANNA, S. (2012). Gender wage discrimination in India: Glass ceiling or sticky floor? [online]. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2115074>.
KINGDON, G. G., UNNI, J. (2001). Education and Women's Labour Market Outcomes in India [online]. *Education Economics*, 9(2): 173–195. <https://doi.org/10.1080/09645290110056994>.
KOENKER, R. (2005). *Quantile regression* [online]. New York: Cambridge University Press. <https://doi.org/10.1017/cbo9780511754098>.
KOENKER, R., BASSETT, G. (1978). Regression quantiles [online]. *Econometrica*, 46(1): 33–50. <https://doi.org/10.2307/1913643>.
KOENKER, R., MACHADO, J. A. (1999). Goodness of fit and related inference processes for quantile regression [online]. *Journal of the American Statistical Association*, 94(448): 1296–1310. <https://doi.org/10.1080/01621459.1999.10473882>.
MACHADO, J. A., MATA, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression [online]. *Journal of Applied Econometrics*, 20(4): 445–465. <https://doi.org/10.1002/jae.788>.
MADAN, S., MOR, S. (2022). Is Gender Earnings Gap a Reality? Signals from Indian Labour Market [online]. *Statistika: Statistics ans Economy Journal*, 102(2): 168–183. <https://doi.org/10.54694/stat.2021.19>.
MADHESWARAN, S., ATTEWELL, P. (2007). Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey [online]. *Economic and Political Weekly*, 42(41): 4146–4153. <http://www.jstor.org/stable/40276549>.
MELLY, B. (2006). *Estimation of Counterfactual Distributions Using Quantile Regression* [online]. University of St.Gallen Discussion Paper. <https://www.alexandria.unisg.ch/handle/20.500.14171/83172>.
MITRA, A. (2016). Education and earning linkages of regular and casual workers in India: a quantile regression approach [online]. *Journal of Social and Economic Development*, 18(1–2): 147–174. <https://doi.org/10.1007/s40847–016–0029–4>.
ÑOPO, H. (2008). Matching as a Tool to Decompose Wage Gaps [online]. *The Review of Economics and Statistics*, 90(2): 290–299. <https://doi.org/10.1162/rest.90.2.290>.
ÑOPO, H., DAZA, N., RAMOS, J. (2012). Gender earning gaps around the world: a study of 64 countries [online]. *International Journal of Manpower*, 33(5): 464–513. <https://doi.org/10.1108/01437721211253164>.
OAXACA, R. (1973). Male-female wage differentials in urban labour markets [online]. *International Economic Review*, 14(3): 693–709. <https://doi.org/10.2307/2525981>.
PALA, V., NONGSPUNG, A. (2022). Trends in wages of regular wage employees and casual workers in Indian states [online]. *The Indian Journal of Labour Economics*, 65(3): 667–687. <https://doi.org/10.1007/s41027-022-00386-6>.
PATIDAR, V. K., WADHVANI, R., SHUKLA, S., GUPTA, M., GYANCHANDANI, M. (2023). Quantile Regression Comprehensive in Machine Learning: a Review [online]. In: *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS),* 1–6. <https://doi.org/10.1109/SCEECS57921.2023.10063026>.
PORTER, S. R. (2014). Quantile regression: Analyzing changes in distributions instead of means [online]. *Higher Education: Handbook of Theory and Research*, 335–381. <https://doi.org/10.1007/978–3–319–12835-1_8>.
SENGUPTA, A., DAS, P. (2014). Gender Wage Discrimination across Social and Religious Groups in India: Estimates with Unit Level Data [online]. *Economic and Political Weekly*, 49(21): 71–76. <http://www.jstor.org/stable/24479554>.
SENGUPTA, P., PURI, R. (2021). Gender pay gap in India: a reality and the way forward – an empirical approach using quantile regression technique [online]. *Studies in Microeconomics*, 10(1): 50–81. <https://doi.org/10.1177/2321022221995674>.
SI, X., LI, M. (2023). Impact of the internet use on informal workers' wages: Evidence from China [online]. *PloS One*, 18(5): e0285973. <https://doi.org/10.1371/journal.pone.0285973>.
SPENCE, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3): 355–374. <https://doi.org/10.2307/1882010>.
STATACORP (2013). *Stata Statistical Software: Release 13.1*. College Station TX, StataCorp LLC.
Unit Level data of Periodic Labour Force Survey (PLFS) July 2020–June 2021 [online]. Ministry of Statistics and Program Implementation, Government of India. <https://mospi.gov.in/PLFS-July20-June21>.
WALDMANN, E. (2018). Quantile regression: A short story on how and why [online]. *Statistical Modeling*, 18(3–4): 203–218. <https://doi.org/10.1177/1471082x18759142>.

ANNEX

Table A1 Distribution of male and female workers (in percent) based on different characteristics

Variables/characteristics	Regular/salaried employees		Casual workers	
	Male	Female	Male	Female
	(N = 29 691)	(N = 9 727)	(N = 18 773)	(N = 5 513)
Residential location				
Rural	76.5	23.5	72.8	27.3
Urban	75.8	24.2	83.5	16.5
Social group				
STs	67.6	32.4	67.4	32.6
SCs	73.4	26.6	75.2	24.9
OBCs	77.2	22.8	73.9	26.1
Others	78.1	21.9	82.3	17.7
Marital status				
Never married	81.4	18.6	91.2	8.8
Married currently	78.2	21.8	75.4	24.6
Widowed	14.8	85.2	20.4	79.6
Divorced/separated	32.1	67.9	38.3	61.7
General education				
Illiterate	52.0	48.0	59.0	41.0
No formal schooling	77.6	22.4	76.0	24.0
Up to primary school	72.6	27.4	75.6	24.4
Middle school	82.7	17.3	84.9	15.1
Secondary school	81.5	18.5	86.6	13.4
Higher secondary	81.1	18.9	88.8	11.2
Graduate	77.8	22.2	91.3	8.7
Postgraduate or above	67.4	32.6	89.5	10.6
Technical education				
Not received	76.4	23.6	74.4	25.6
Have technical education	74.3	25.7	94.5	5.5
Occupations				
White-collar	68.9	31.2	75.8	24.2
Blue-collar	80.9	19.1	74.5	25.5
Agricultural	80.8	19.2	73.6	26.4
Industries				
Production and extraction	85.2	14.8	60.3	39.7
Infrastructure & utilities	94.1	5.9	89.0	11.0
Goods & service distribution	91.6	8.5	93.8	6.2
Knowledge & service-based	80.9	19.1	79.5	20.5
Public and social services	44.8	55.2	60.3	39.8

Source: Author's calculation based on PLFS 2020–21 dataset using Stata v.13