

Gendered Wage Discrimination in India: Sticky Floor or Glass Ceiling?*

Luke DiMartino[†]

May 2021

Preliminary and incomplete. Please do not cite or circulate.

Abstract

Prior analyses of the large *India Human Development Survey* have focused on mean differences of wages between men and women and corresponding decompositions such as the Blinder-Oaxaca method. To investigate the presence of a "sticky floor" or "glass ceiling," the magnitude of wage discrimination at the tails of the wage distribution, this paper uses quantile regression and the Machado-Mata-Melly decomposition to examine wage gaps at different quantiles of the wage distribution. *results go here*

*Thanks, Mom!

[†]Georgetown University. lad132@georgetown.edu

1 Introduction

Fake introduction to work with citations. Gender and caste are critical factors in Indian labor markets.

The seminal work, [Banerjee and Knight \(1985\)](#) detailed caste-based discrimination in labor markets in India. Also, [Agrawal \(2014\)](#) exists.

The remainder of this paper proceeds as follows: Section [2](#) reviews existing literature on wage discrimination in India. Next, Section [3](#) and Section [4](#) describe the empirical strategy and the data. I then present results in Section [5](#). Lastly, Section [6](#) concludes.

2 Literature Review

3 Methodology

3.1 Blinder-Oaxaca Decomposition

The Blinder-Oaxaca decomposition, first proposed by [Blinder \(1973\)](#) and [Oaxaca \(1973\)](#) considers the gap between mean wages of men and women as the sum of two factors. First, men are, on average, better endowed with covariates like education and experience that increase their productivity and by extension, their wages. Second, men experience greater returns to these covariates (e.g. a greater return for every year of education than women experience). The Blinder-Oaxaca decomposition terms the first the composition effect (the "explained" gap) and the second the wage structure effect (the "unexplained" gap).

Assume that the wages W can be naively and a vector of independent controlling variables X by:

$$W_i = \beta_0 + \beta_k \mathbf{X}_k + u_i. \quad (1)$$

If the sample can be divided into two groups, A and B , then the same equation could be used over subsamples.

$$W_{Ai} = \beta_{A0} + \beta_{A1}X_{1i} + \dots + \beta_{Ak}X_{ki} + u_{Ai}$$

$$W_{Bi} = \beta_{B0} + \beta_{B1}X_{1i} + \dots + \beta_{Bk}X_{ki} + u_{Bi}$$

.

Therefore, the mean outcomes are

$$\begin{aligned}\overline{W}_A &= \mathbf{X}'_A \hat{\beta}_A \\ \overline{W}_B &= \mathbf{X}'_B \hat{\beta}_B.\end{aligned}$$

Therefore, the mean difference is

$$\overline{W}_A - \overline{W}_B = \mathbf{X}'_A \hat{\beta}_A - \mathbf{X}'_B \hat{\beta}_B = \mathbf{X}'_A (\hat{\beta}_A - \hat{\beta}_B) + (\mathbf{X}'_A - \mathbf{X}'_B) \hat{\beta}_B. \quad (2)$$

The first term is the wage structure effect, describing how the difference in wages is an effect of different coefficient estimates, which are the effects of the control variables like education and experience. The second term is the composition effect, the difference in levels of the predictor variables, like amount of education and experience that determine the composition of each subsample.

3.2 Quantile Regression

Quantile regression was developed first by [Koenker and Bassett \(1978\)](#). While ordinary least squares estimates a conditional mean function (i.e. *conditional on the given values for \mathbf{X}' what is the expected value of Y ?*) quantile regression estimates conditional quantile functions (i.e. *conditional on the given values for \mathbf{X}' , what is the expected third quartile value of Y ?*). The insight of quantile regression is that it allows for coefficients to vary by quantile of the wage distribution. This matches economic intuition one might have: in low-wage jobs, employers may value experience more and formal education less, but in high-wage managerial jobs, employers might value formal education and experience equally. OLS would obscure the nuances of these effects.

The θ^{th} quantile of the conditional distribution is given by

$$Q_\theta(Y_i|X_i) = X_i\beta_\theta, \theta \in (0, 1). \quad (3)$$

For a given θ , the estimate of β_θ minimizes the sum of deviation,

$$\sum_{i=1}^n \rho_\theta(Y_i - X_i\beta_\theta) \quad (4)$$

where

$$\rho_{\theta} = \begin{cases} \theta(u) & \text{for } u > 0 \\ (1 - \theta)u & \text{for } u \leq 0 \end{cases} . \quad (5)$$

3.3 Machado-Mata Decomposition

The Machado-Mata decomposition, first developed by [Machado and Mata \(2005\)](#), is a generalization of the BO decomposition specifically for quantile regression. The MM method decomposes the gap at each quantile by constructing a counterfactual distribution, either using men’s characteristics and women’s returns, or vice-versa.

4 Data

4.1 Data

To investigate gendered wage gaps, I use individual-level data from the India Human Development Survey (IHDS). The IHDS is a survey jointly conducted by the National Council of Applied Economic Research (NCAER) in Delhi; University of Maryland, College Park; Indiana University; and the University of Michigan. It is a nationally-representative panel survey of more than 42,000 households and 420,311 individuals with an 85% reinterview rate. The first set of interviews, IHDS 1, was conducted in 2004 and 2005. The second set of interviews, IHDS 2, was conducted in 2011 and 2012.

4.2 Variables

Of the 420,311 individuals, 97,413 observations report hourly wages. Another 1,557 are missing data for relevant regressors.

The dependent variable of interest is *female*, a dummy variable for sex.

4.3 Descriptive Statistics

5 Results

6 Conclusion

References

- Agrawal, Tushar**, “Gender and caste-based wage discrimination in India: some recent evidence,” *Journal for Labour Market Research*, 12 2014, 47.
- Banerjee, Biswajit and J. B. Knight**, “Caste discrimination in the Indian urban labour market,” *Journal of Development Economics*, April 1985, 17 (3), 277–307.
- Blinder, Alan S.**, “Wage Discrimination: Reduced Form and Structural Estimates,” *The Journal of Human Resources*, 1973, 8 (4), 436–455.
- Koenker, Roger and Gilbert Bassett**, “Regression Quantiles,” *Econometrica*, 1978, 46 (1), 33–50.
- Machado, José A. F. and José Mata**, “Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression,” *Journal of Applied Econometrics*, 2005, 20 (4), 445–465.
- Oaxaca, Ronald**, “Male-Female Wage Differentials in Urban Labor Markets,” *International Economic Review*, 1973, 14 (3), 693–709.

A Appendix Tables and Figures

A.1 Descriptive Statistics

Appendix Table A1: Sample Descriptive Statistics			
	(1)	(2)	(3)
	Full sample	Men	Women
HQ13 7.10 hourly wage	bonuses	25.53	28.87
17.01	(29.19)	(30.57)	(23.22)
lwshourly	2.88	3.04	2.48
	(0.786)	(0.753)	(0.724)
female	0.28	0.00	1.00
	(0.450)	(0)	(0)
Literacy	0.66	0.75	0.42
	(0.474)	(0.432)	(0.494)
Years of Education	5.68	6.55	3.47
	(5.032)	(4.860)	(4.776)
Marital Status	0.76	0.77	0.73
	(0.426)	(0.418)	(0.445)
Age	36.87	36.83	37.00
	(12.87)	(12.91)	(12.76)
Observations	95764	68810	26954

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix Table A2: Inflation-Adjusted Sample Descriptive Statistics

	(1)	(2)	(3)
	Full sample	Men	Women
Inflation-adjusted hourly wages	33.63 (40.36)	38.27 (42.83)	21.79 (30.14)
Log inflation-adjusted hourly wages	3.14 (0.792)	3.30 (0.763)	2.73 (0.710)
female	0.28 (0.450)	0.00 (0)	1.00 (0)
Literacy	0.66 (0.474)	0.75 (0.432)	0.42 (0.494)
Years of Education	5.68 (5.032)	6.55 (4.860)	3.47 (4.776)
Marital Status	0.76 (0.426)	0.77 (0.418)	0.73 (0.445)
Age	36.87 (12.87)	36.83 (12.91)	37.00 (12.76)
Observations	95764	68810	26954

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix Table A3: Subsample Regressions with Men and Women, $\tau = .9$

	(1)	(2)
	Men	Women
q90		
married	0.444*** (22.04)	0.277*** (7.05)
HQ19 11.2 Educ: Literacy	0.742 (1.50)	1.641*** (4.91)
HQ19 11.6 Educ: Completed Years, never,i1=0	0.104*** (21.75)	0.131*** (10.03)
HQ19 11.1 Educ: Post 2nd subj	0.0357*** (5.54)	0.0408*** (3.40)
HQ12 7.4 Occupation -job1	-0.00807*** (-19.20)	-0.00974*** (-10.11)
Constant	2.445*** (4.98)	1.162** (3.02)
Observations	14112	2942

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B Appendix One