# 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 Please do not cite or circulate.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*

[†]Georgetown University. lad132@georgetown.edu

# 1 Introduction

Gender wage discrimination has long been the subject of political and scholarly debate because the removal of discriminatory barriers has the power to enhance economic development. While economists explored the paradoxes of wage discrimination primarily in developing countries (Blau and Kahn, 1992), wage discrimination wage largely ignored elsewhere.

The body of empirical literature examining wage discrimination in India has primarily focused on estimations of the wage gap in general and on the returns to additional years of education. In their seminal work, Banerjee and Knight (1985) examined caste-based discrimination. Research on gender wage discrimination has relied upon the Indian National Sample Survey Bhaumik and Chakrabarty (2008) Agrawal (2014) exists.

While mean gaps are important, newer methods for decomposing differences go beyond single-statistic measures because mean differences may not be consistent across the wage distribution. Jenkins (1994) goes so far as to criticize the single-statistic measure as inadequate for understanding the reality of discrimination.

New research like attempts to measure and decompose gaps and estimate coefficients across the wage distribution. This method has been applied for public-private wage differences (Azam and Prakash, 2015) and for wage di

In this paper, I conduct a distributional analysis of gender wage gaps using wage data from the India Human Development Survey. The methodology, for the most part, mirrors the implementation of the Machado-Mata decomposition in **?**, which analyzes the Indian National Sample Survey.

The remainder of this paper proceeds as follows: Section 2 and Section 3 describe the empirical strategy and the data. I then present results in Section 4. Lastly, Section 5 concludes.

# 2 Methodology

## 2.1 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 indepedent controlling variables $X$ by:

$$W_i = \beta_0 + \beta_k \mathbf{X}_k + u_i. \tag{1}$$

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

$$W_{Ai} = \beta_{A0} + \beta_{A1}X_{1i} + \ldots + \beta_{Ak}X_{ki} + u_{Ai}$$
$$W_{Bi} = \beta_{B0} + \beta_{B1}X_{1i} + \ldots + \beta_{Bk}X_{ki} + u_{Bi}$$

.

Therefore, the mean outcomes are

$$\overline{W}_A = \mathbf{X'}_A \hat{\beta}_A$$
$$\overline{W}_B = \mathbf{X'}_B \hat{\beta}_B.$$

Therefore, the mean difference is

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

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.

## 2.2 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 $\theta^{\text{th}}$ quantile of the conditional distribution is given by

$$Q_\theta\left(Y_i|X_i\right) = X_i\beta_\theta,\ \theta \in (0,1). \tag{3}$$

For a given $\theta$, the estimate of $\beta_\theta$ minimizes the sum of deviation,

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

where

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

### 2.3   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.

## 3   Data

### 3.1   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; Univerity 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.

The sample is

## 3.2    Variables

The independent variable of interest is hourly wages. XXX outliers were trimmed, defined as those observations either below the $25^{\text{th}}$ income percentile less three times the interquartile range or greater than the $75^{\text{th}}$ income percentile plus three times the interquartile range, to reduce influential points and ensure misreported data did not affect results.

The grouping variable of interest is a *female*, an indicator for women. Independent variables of interest include *age*, *years of education*, *marital status*, and *literacy* — all self-explanatory.

## 3.3    Descriptive Statistics

Summary statistics of the sample, and subsamples of men and women are presented in Table 1.

# 4    Results

# 5    Conclusion

# References

**Agrawal, Tushar**, "Gender and caste-based wage discrimination in India: some recent evidence," *Journal for Labour Market Research*, 12 2014, *47*.

**Azam, Mehtabul and Nishith Prakash**, "A Distributional Analysis of Public–Private Wage Differential in India," *LABOUR*, 2015, *29* (4), 394–414.

**Banerjee, Biswajit and J. B. Knight**, "Caste discrimination in the Indian urban labour market," *Journal of Development Economics*, April 1985, *17* (3), 277–307.

**Bhaumik, Sumon Kumar and Manisha Chakrabarty**, "Does move to market have an impact on earnings gap across gender? Some evidence from India," *Applied Economics Letters*, 2008, *15* (8), 601–605.

**Blau, FD and LM Kahn**, "The Gender Earnings Gap - Learning from International Comparisons," *American Economic Review*, May 1992, *82* (2), 533–538. 104th Annual Meeting of the American Economic Assoc., New Orleans, LA, Jan 3-5, 1992.

**Blinder, Alan S.**, "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources*, 1973, *8* (4), 436–455.

**Jenkins, Stephen P**, "Earnings discrimination measurement: A distributional approach," *Journal of Econometrics*, 1994, *61* (1), 81–102.

**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: Inflation-Adjusted Sample Descriptive Statistics

|  | (1)<br>Full sample | (2)<br>Men | (3)<br>Women |
|---|---|---|---|
| Inflation-adjusted hourly wages | 33.50 | 38.11 | 21.71 |
|  | (39.94) | (42.32) | (29.98) |
| Log hourly wages | 3.14 | 3.30 | 2.72 |
|  | (0.791) | (0.760) | (0.712) |
| female | 0.28 | 0.00 | 1.00 |
|  | (0.449) | (0) | (0) |
| Literacy | 0.66 | 0.76 | 0.42 |
|  | (0.472) | (0.429) | (0.494) |
| Years of Education | 5.84 | 6.72 | 3.58 |
|  | (5.211) | (5.022) | (5.002) |
| Marital Status | 0.77 | 0.78 | 0.74 |
|  | (0.421) | (0.415) | (0.437) |
| Age | 36.52 | 36.42 | 36.77 |
|  | (12.05) | (12.10) | (11.92) |
| Years of Potential Experience | 25.68 | 24.70 | 28.19 |
|  | (13.81) | (13.61) | (14.02) |
| Observations | 93382 | 67177 | 26205 |

mean coefficients; sd in parentheses

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

Appendix Table A2: Subsample Regressions with Men and Women, $\tau = .9$

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

$t$ statistics in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

# B    Appendix One