Baseline Regression Results

Luke DiMartino

April 30, 2022

1 Summary

I am investigating gendered wage gaps in developing India. I am interested in two major aspects of the gaps. First, I am conducting a Machado-Mata decomposition of the wage gap. The MM decomposition builds on the traditional Blinder-Oaxaca decomposition, which splits the difference in means between an advantaged group and a disadvantaged group into a sum of two parts: the difference in endowed characteristics and the difference in returns to those characteristics. For example, men in the United States are on average more educated than women. Knowing only this fact about the two groups, a difference in mean wages is not economically unreasonable — men are better endowed with characteristics that make them more productive (whether this endowment is systematically unjust or unfair is outside of the scope of this paper). But, men also earn greater returns to these characteristics. For instance, a man experiences greater return for the same graduate degree than a woman. This is economically unreasonable, the so-called "unexplained" part of the gap.

The MM decomposition expands on the Blinder-Oaxaca decomposition by decomposing this gap in a similar manner but at different quantiles of the wage distribution.

I am also interested in the different returns women in particular experience at different quantiles of the wage distribution. As a policymaking tool, this has more obvious implications. Certain covariates may be correlated with higher wages, and causal inference might suggest that in general they raise wages significantly. But, quantile regression allows us to zoom in on certain parts of the distribution to determine whether that effect varies by wage level.

2 Baseline OLS Results

Appendix Table A1: **Pooled OLS Regression Results**

	White SE's	**Clustered SE's**	**Occupation FE's**	**District FE's**
female	-0.405***	-0.405***	-0.339***	-0.348***
	(0.00491)	(0.00526)	(0.00485)	(0.00465)
Literacy	-0.115***	-0.115***	-0.0500***	-0.0766***
	(0.00767)	(0.00796)	(0.00727)	(0.00682)
Years of Education	0.0838***	0.0838***	0.0479***	0.0368***
	(0.00151)	(0.00156)	(0.00146)	(0.00139)
Years of Pot. Experience	0.0335***	0.0335***	0.0239***	0.0158***
	(0.00115)	(0.00117)	(0.00108)	(0.00102)
Marital Status	0.0422***	0.0422***	0.0637***	0.0969***
	(0.00588)	(0.00611)	(0.00558)	(0.00525)
Constant	2.416***	2.416***	3.586***	4.001***
	(0.0160)	(0.0166)	(0.121)	(0.124)
Occupation FE's	No	No	Yes	Yes
District FE's	No	No	No	Yes
Observations	93382	93382	93382	93382
Adjusted R^2	0.364	0.364	0.444	0.520
AIC	178927.1	178927.1	166520.3	153192.9
BIC	179078.2	179078.2	167493.0	157782.9
F	3013.4	2639.6	673.4	220.3

Standard errors in parentheses

All models include caste fixed effects.

Additional controls include squared age, indicators for graduate degree holders program income, little english ability, and english fluency.

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

This table is a general ordinary least squares regression on the entire dataset. This table demonstrates a few key results.

1. The naive estimation of the mean wage gap is somewhere between 41% and 50%, using exact calculations from the coefficient on female.

2. There is not a dramatic difference between White and clustered standard errors. Comparing columns 1 and 2, the standard errors in the second are slightly larger, but not dramatically and not enough to affect results. This is important because although clustered standard errors are theoretically correct, computing them is infeasible for more complex models presented below.

3. Occupation is a massive source of wage inequality. Controlling for it between columns 2 and 3 absorbs a significant amount of the wage gap.

4. The coefficient on literacy is negative and statistically significant, suggesting that controlling for all other factors, being literate reduces wages by nearly 8%.

Appendix Table A2: **Pooled OLS Subsample Regression Results**

	Men	Women
Literacy	-0.0578***	-0.108***
	(0.00791)	(0.0145)
Years of Education	0.0387***	0.0316***
	(0.00163)	(0.00304)
Years of Pot. Experience	0.0184***	0.0115***
	(0.00128)	(0.00170)
Marital Status	0.119***	0.0172
	(0.00669)	(0.00952)
Constant	3.998***	3.360***
	(0.132)	(0.487)
Occupation FE's	Yes	Yes
District FE's	Yes	Yes
Observations	67177	26205
Adjusted R^2	0.479	0.467
AIC	110536.4	40562.5
BIC	114948.1	44477.7
F	131.2	.

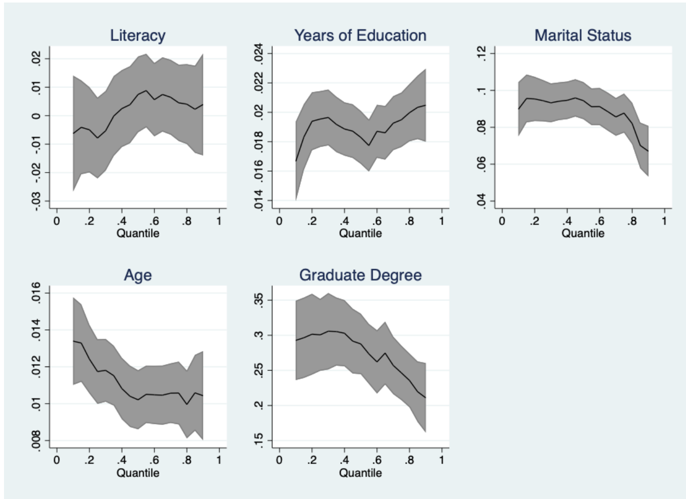
Standard errors in parentheses

All models include caste fixed effects.

Additional controls include squared age, indicators for graduate degree holders, program income, little english ability, and english fluency.

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

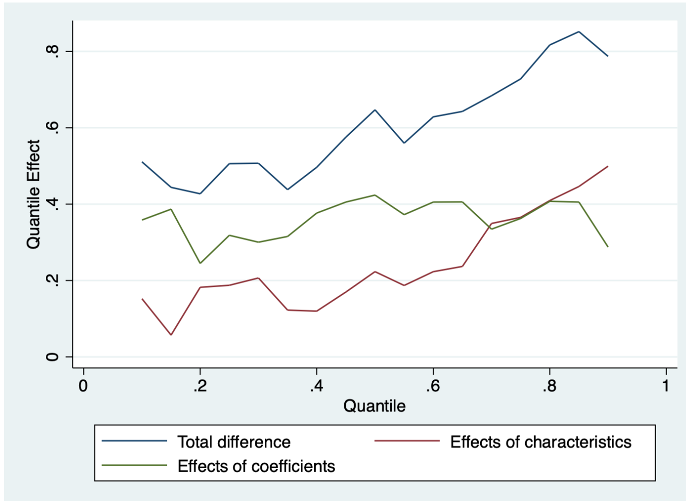
This table is a OLS regression on two subsamples, men and women. This is equivalent to using interaction effects with female for every variable, but makes the results a bit more clear. The difference in returns between men and women is considered economically unreasonable, and there are a few variables where this difference is significant. Marital status, in particular, affects men's wages dramatically more than women's, while a graduate degree provides far greater returns to women than to men.



Appendix Figure A1: Coefficients from entire sample across quantiles

3 Quantile Regression Results

This figure is a model for the main figure I will present in the paper. It shows the estimates of coefficients at different conditional quantiles. However, since this is pretty computationally intensive, this image is from an arbitrary 10% subsample of the dataset, not from women, which is the figure I will present in the final paper. However, it still is indicative of how varied coefficients can be across the wage distribution. A graduate degree seems to be worth substantially less at the high end of the distribution where it is more precisely estimated. Likewise, marriage boosts wages substantially less at the top of the distribution.



Appendix Figure A2: Decomposition of wage gap across quantiles

This figure shows the decomposition of the wage gap by half-decile. The ”effects of coefficients” are the effects of greater returns, while the ”effects of characteristics” are the effect on the wage gape cause by differing endowments. While the mean difference grows as wage increases, that appears to be primarily a growth in characteristics differences — the ”unexplained discrimination” is constant across the distribution.

4 Moving Forward

Moving forward, I would like to potentially add more control variables and fine-tune the more complex models. I am reading papers using this data and examining what choices they make when cleaning variables as well. The quantile regression models are highly computationally intensive, and so I would like to use more robust methods to get more precise estimates when I have finalized the specifications.