

Microeconomics of Development

Data Exercise Two

Amit Basole

December 11, 2021

Earnings gaps and returns to education

In this exercise we will calculate earnings gaps by gender, caste and religion, as well as estimate the returns to an additional year of education using a simple Mincerian wage equation. We will use India Human Development Survey (IHDS) data. Here are the relevant variables.

Variables

1. ID11 - Religion of household
2. ID13 - caste group of household
3. RO5: HQ4 2.5 Age
4. RO3: HQ4 2.3 Sex
5. RO6: HQ4 2.6 Marital Status
6. RO7: HQ4 2.7 Primary Activity Status
7. ED3: HQ19 11.3 Education: English ability
8. EDUC7: HQ19 11.4,6 Education: Completed Years, 7cats
9. ED6: HQ19 11.6 Education: Completed Years, never, 1=0
10. URBAN4_2011: 4-cat urban/rural from 2011 census & IHDS-II
11. INCNONAG: HQ13 7.3 income: non-agricultural wage
12. INCSALARY: HQ13 7.3 income: salary position
13. WSEARN: HQ13 7.10 annual earnings -person total
14. WT - survey weight
15. DISTID - district code
16. STATEID - state code

Social identities and earnings gaps

1. Tabulate the mean of WSEARN (annual earnings per person) by gender, caste, religion, and type of primary occupation.
2. Calculate the following earnings ratios:
 - Male to female
 - Hindu to Muslim
 - Brahmin to SC

- Salaried to casual wage
3. Regress log earnings on a gender dummy with survey weights to find the aggregate earnings gap. Add age, age squared, and education. What happens to the coefficient on the gender dummy?
 4. Add religion, caste, and primary activity status one at a time to see which one makes the biggest difference to the gender coefficient. How would you interpret your results?
 5. If the coefficient on the gender dummy decreases over time, can we interpret this as a sign of converging male and female earnings? (Hint - think selection bias).

Estimating returns to education using a Mincerian earnings equation

The basic specification is: $y = \alpha + \beta * education + \gamma * age + \delta * age^2 + \epsilon$, where y is wage or salary income.

1. Estimate the return to an additional year of education separately for casual wage and salaried workers. Interpret both the coefficients on education.
2. Repeat 1 by stratifying the regression by gender (i.e. separate regressions for men and women). What do you find? How is this different from including a dummy for gender in the regression in 1?
3. Repeat 1, this time, stratifying the regression by level of education - calculate returns to an additional year of schooling for those who have more than 10 years of education. What is the difference between your results for 1 versus 3?