Problem Set 5 Figures

Luke DiMartino

April 13, 2022

I am analyzing data from the India Human Development Survey. My focus is on gender disparities and the effects of education on them.

My baseline hypothesis, of course, is that there is a difference in wealth and income between men and women. The variables are intuitively named, except RO3, which is the indicator for sex.

This is a panel dataset, so first I'll report basic statistics about the balance of the panel, with panel survival by state and by sex.

```
. tab STATEID PWAVES, row nofreq
```

State code	which surveys p has been in only 2012	only 2005	both 11	Total
Jammu & Kashmir 01	10.90	12.21	76.89	100.00
Himachal Pradesh 02	10.48	14.50	75.02	100.00
Punjab 03	10.16	13.51	76.33	100.00
Chandigarh 04	20.78	24.16	55.06	100.00
Uttarakhand 05	10.82	13.99	75.19	100.00
Haryana 06	13.16	12.29	74.56	100.00
Delhi 07	29.06	28.89	42.05	100.00
Rajasthan 08	13.12	13.57	73.31	100.00
Uttar Pradesh 09	13.20	13.01	73.79	100.00
Bihar 10	13.27	15.07	71.66	100.00
Sikkim 11	16.57	16.77	66.67	100.00
Arunachal Pradesh 12	11.00	21.93	67.07	100.00
Nagaland 13	26.96	32.18	40.86	100.00
Manipur 14	7.42	18.18	74.40	100.00
Mizoram 15	9.19	27.44	63.37	100.00
Tripura 16	21.28	25.90	52.83	100.00
Meghalaya 17	9.92	14.71	75.36	100.00
Assam 18	23.92	24.33	51.75	100.00
West Bengal 19	10.62	12.42	76.97	100.00
Jharkhand 20	13.23	19.22	67.56	100.00
Orissa 21	10.26	13.04	76.70	100.00
Chhattisgarh 22	12.88	11.98	75.15	100.00
Madhya Pradesh 23	11.99	14.05	73.97	100.00
Gujarat 24	13.62	17.70	68.67	100.00
Daman & Diu 25	9.76	11.39	78.84	100.00
Dadra+Nagar Haveli 26	16.85	17.01	66.14	100.00
Maharashtra 27	9.74	11.64	78.61	100.00
Andhra Pradesh 28	13.34	21.22	65.44	100.00
Karnataka 29	14.15	18.53	67.32	100.00
Goa 30	5.91	7.54	86.55	100.00
Kerala 32	9.80	17.94	72.26	100.00
Tamil Nadu 33	10.28	15.92	73.79	100.00
Pondicherry 34	7.63	13.80	78.56	100.00
Total	12.75	15.41	71.84	100.00

```
.
. tab RO3 PWAVES, row nofreq
```

HQ4 2.3 Sex	which surveys p has been in only 2012	only 2005	both 11	Total
Male 1	10.83	14.50	74.67	100.00
Female 2	14.69	16.33	68.98	100.00
Total	12.75	15.41	71.84	100.00

```
.
. ttest INCOME, by(RO3)
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Male 1	211,889	128778.6	466.3794	214681	127864.5	129692.7
Female 2	208,422	124881.3	455.824	208098.5	123987.9	125774.7
Combined	420,311	126846	326.1556	211451.2	126206.8	127485.3
diff		3897.264	652.3065		2618.764	5175.765

```
    diff = mean(Male 1) - mean(Female 2)                          t =   5.9746
H0: diff = 0                                     Degrees of freedom =   420309

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

This t-test shows that there is a highly statistically significant naive difference in income between men and women.

```
. ttest WSHOURLY, by(RO3)
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Male 1	70,170	29.1133	.1263209	33.46194	28.86571	29.36088
Female 2	27,243	17.07203	.145522	24.01907	16.78679	17.35726
Combined	97,413	25.74577	.1011724	31.57697	25.54748	25.94407
diff		12.04127	.2220863		11.60598	12.47656

```
    diff = mean(Male 1) - mean(Female 2)                          t =  54.2189
H0: diff = 0                                     Degrees of freedom =    97411

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

This t-test shows that there is a highly statistically significant naive difference in hourly wage between men and women — on average, men earn about 12 rupees more per hour than women.

```
. ttest POOR, by(RO3)
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Male 1	211,718	.2147715	.0008925	.4106647	.2130223	.2165208
Female 2	208,262	.2254948	.0009157	.4179088	.2237	.2272897
Combined	419,980	.2200891	.0006393	.414307	.218836	.2213421
diff		-.0107233	.0012785		-.0132292	-.0082174

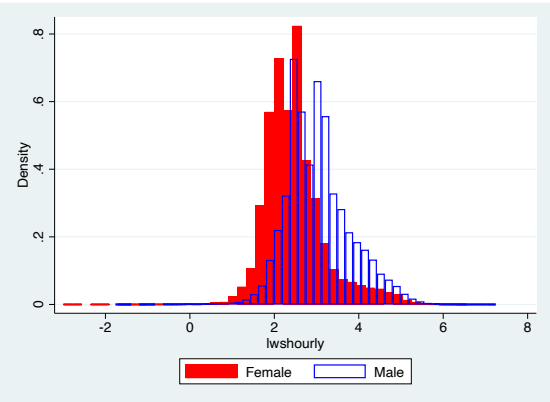
```
    diff = mean(Male 1) - mean(Female 2)                          t =  -8.3871
H0: diff = 0                                     Degrees of freedom =   419978

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

This t-test shows that there is a highly statistically significant naive difference in poverty between men and women — women are about 1.1 percentage points more likely to be poor.

```
. gen lwshourly = log(WSHOURLY)
(322,898 missing values generated)
. pctile income_dec = INCOME, nq(10)
. pctile wshourly_dec = WSHOURLY, nq(10)
. hist wshourly_dec, by(RO3)
```

The histogram of hourly wages has tails far too long to be meaningful, but this histogram grouped by decile shows the difference in the wage distribution between men and women.

```
. twoway (histogram lwshourly if RO3==2, color(red)) ///
>        (histogram lwshourly if RO3==1, ///
>            fcolor(none) lcolor(blue)), legend(order(1 "Female" 2 "Male" ))
```



This histogram shows the difference in log hourly wage distributions between men and women.

Going forward, my plan is to clean the income and wage data to get more accurate measures, including determining what to do with outliers and negative results. Then, I will construct control variables from the relevant variables in the survey. From there, I should be able to develop a fundamental OLS model with broad controls.

Then, I will develop a household fixed effects model and investigate potential instruments for gender-equality-related independent variables.