Econ 621 Assignment 2

Nathan Mather

September 25, 2018

Part a

Replication of Table III

variable	Full Sample Mean	Full Sample SD	Working Women Mean	Working Women SD
age	42.54	8.07	41.97	7.72
educ	12.29	2.28	12.66	2.29
kidslt6	0.24	0.52	0.14	0.39
kidsge6	1.35	1.32	1.35	1.32
husage	45.12	8.06	44.61	7.95
huseduc	12.49	3.02	12.61	3.04
wage	2.37	3.24	4.18	3.31
huswage	7.48	4.23	7.23	3.57
nonlab_i	3.76	5.90	3.39	6.07
hours	740.58	871.31	1302.93	776.27
hushrs	2267.27	595.57	2233.46	582.91

Part b

The full sample is not used in this regression because wage is only observed for people who are working.

The Problem with this is that selection into the workforce is not random. In fact, being in the workforce is positively correlated with the error term in the regression. If we consider a latent variable model for desired hours with

$$latent\_hours = lwage + nwifeinc + kidslt6 + kidsge6 + age + educ + \epsilon$$

$$hours = \begin{cases} latent\_hours & latent\_hours > 0 \\ 0 & latent\_hours \leq 0 \end{cases}$$

The sample excludes observations with zero hours and so observations where $latent\_hour$ is negative. This is of course more likely if ϵ is negative. So, our sample around zero house consists of observations that are more likely to have positive ϵ.

Another reason would be if an unobservable trait like work ethic, is correlated both with the decision to work and the error term.

term	estimate	std.error	statistic	p.value
(Intercept)	2114.70	347.44	6.09	0.00
lwage	-17.41	80.71	-0.22	0.83
nwifeinc	-4.25	3.20	-1.33	0.18
kidslt6	-342.50	130.69	-2.62	0.01
kidsge6	-115.02	29.27	-3.93	0.00
age	-7.73	5.80	-1.33	0.18
educ	-14.44	18.06	-0.80	0.42

Part C

The coefficient on wage is positive under this specification rather than negative like in part b. Division bias may have been an issue in part B, since wage is income/hours worked, which would bias the result towards -1. So it is not surprising the estimate using IV is higher.

term	estimate	std.error	statistic	p.value
(Intercept)	2127.53	351.88	6.05	0.00
lwage	45.74	220.64	0.21	0.84
nwifeinc	-4.45	3.27	-1.36	0.17
kidslt6	-337.18	131.44	-2.57	0.01
kidsge6	-112.29	30.40	-3.69	0.00
age	-7.85	5.81	-1.35	0.18
educ	-21.03	30.16	-0.70	0.49

Part d

This estimate is not significantly different than the one in part B.

term	estimate	std.error	statistic	p.value
(Intercept)	2041.92	315.37	6.47	0.00
lwage	-32.35	126.13	-0.26	0.80
nwifeinc	-4.43	3.41	-1.30	0.20
kidslt6	-275.51	168.56	-1.63	0.10
kidsge6	-98.24	30.94	-3.17	0.00
age	-8.64	5.09	-1.70	0.09
educ	2.43	20.23	0.12	0.90

Part e

This estimate is much larger and statistically significant. This is not surprising. using reported wages corrects the division bias and so corrects for the thing biasing the estimate in part b towards negative one. However, using reported wages as an instrument probably does not solve the endogeneity issue. If effort is correlated with higher wages and higher work effort than this estimate will continue to be biased upwards.

term	estimate	std.error	statistic	p.value
(Intercept)	2135.87	330.87	6.46	0.00
lwage	328.52	156.53	2.10	0.04
nwifeinc	-5.87	3.58	-1.64	0.10
kidslt6	-300.67	168.71	-1.78	0.08
kidsge6	-88.90	33.31	-2.67	0.01
age	-9.45	5.36	-1.76	0.08
educ	-38.24	23.10	-1.66	0.10

Part f

This is my estimate:

term	estimate	std.error	statistic	p.value
(Intercept)	-14.07	15.95	-0.88	0.38
nwifeinc	-0.02	0.00	-4.55	0.00
kidslt6	-0.86	0.12	-7.18	0.00
kidsge6	-0.06	0.04	-1.36	0.17
age	0.75	0.77	0.98	0.33
educ	1.10	1.92	0.57	0.57
age_sq	-0.02	0.01	-1.05	0.29
age_cu	0.00	0.00	1.06	0.29
educ_sq	-0.06	0.11	-0.60	0.55
educ_cu	0.00	0.00	0.54	0.59
age_educ	-0.02	0.04	-0.35	0.73
age_sq_educ	-0.00	0.00	-0.01	1.00
age_educ_sq	0.00	0.00	0.77	0.44
unem	-0.01	0.02	-0.64	0.52
city	0.04	0.11	0.38	0.70
motheduc	0.00	0.02	0.14	0.89
fatheduc	-0.01	0.02	-0.71	0.48

Part g

I calculated the IMR as directed.

Part h

The coefficient on IMR is not statistically significant. This suggests sampel selection is not important.

term	estimate	std.error	statistic	p.value
(Intercept)	4.85	10.85	0.45	0.66
age	-0.08	0.50	-0.17	0.87
educ	-0.76	1.36	-0.56	0.58
age_sq	-0.00	0.01	-0.13	0.90
age_cu	-0.00	0.00	-0.01	0.99
educ_sq	0.01	0.08	0.10	0.92
educ_cu	0.00	0.00	1.11	0.27
age_educ	0.03	0.03	0.99	0.32
age_sq_educ	0.00	0.00	0.31	0.76
age_educ_sq	-0.00	0.00	-2.41	0.02
unem	-0.00	0.01	-0.12	0.90
city	0.09	0.07	1.29	0.20
motheduc	-0.01	0.01	-0.76	0.45
fatheduc	-0.02	0.01	-1.27	0.21
IMR	-0.15	0.15	-1.04	0.30

Part i

My result is significantly different.

term	estimate	std.error	statistic	p.value
(Intercept)	2304.15	456.12	5.05	0.00
lwage_hat	-25.46	236.95	-0.11	0.91
nwifeinc	-0.68	6.63	-0.10	0.92
age	-2.04	10.40	-0.20	0.84
educ	-40.77	49.08	-0.83	0.41
kidslt6	-182.77	261.51	-0.70	0.49
kidsge6	-108.16	32.20	-3.36	0.00
IMR	-304.09	466.11	-0.65	0.51

Part j

These results are exactly the same

term	estimate	std.error	statistic	p.value
(Intercept)	2319.42	435.03	5.33	0.00
lwage	64.43	215.83	0.30	0.77
nwifeinc	-1.04	5.94	-0.18	0.86
kidslt6	-183.79	274.29	-0.67	0.50
kidsge6	-105.48	31.76	-3.32	0.00
age	-2.57	9.91	-0.26	0.80
educ	-49.07	48.43	-1.01	0.31
IMR	-294.11	432.36	-0.68	0.50