PROBLEM SET № 12

Grant Naberhaus, University of Oklahoma   04/30/2020

Answers to Questions Asked

Log wage is most likely MAR as some people may not feel comfortable answering when asked for their income, and of all the observations in the data set, only 69.3 percent have a log wage observation.

Based on the beta value for hgc, the heckman model was the best. Omitting NA observations was the second most accurate, while the least accurate was mean imputation.

Mean for first probit model: 0.2373
Mean for second probit model: 0.20226

The second estimated probit is not realistic, as wives and mothers can hold union jobs. I thought the first probit model was interesting, as it shows that those who are married and/or have children have a higher probability of being in a union. Union jobs are generally considered more secure than non union jobs, so I thought it was interesting that those who may have to financially provide for others are more likely to be in a more secure job, but this also makes sense. Below are the tables displaying information requested:

Before Factor Variables Created

Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
logwage	1,545	1.652	0.688	−0.956	1.201	2.120	4.166
hgc	2,229	12.455	2.444	5	11	14	18
college	2,229	0.105	0.306	0	0	0	1
exper	2,229	6.435	4.867	0.000	2.452	9.778	25.000
married	2,229	0.635	0.482	0	0	1	1
kids	2,229	0.429	0.495	0	0	1	1
union	2,229	0.237	0.426	0	0	0	1

After Factors Created

Table 2:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
logwage	1,545	1.652	0.688	−0.956	1.201	2.120	4.166
hgc	2,229	12.455	2.444	5	11	14	18
exper	2,229	6.435	4.867	0.000	2.452	9.778	25.000
kids	2,229	0.429	0.495	0	0	1	1

Table 3:

	OLS		selection
	Dependent variable:		
	logwage		
	(1)	(2)	(3)
hgc	0.059***	0.036***	0.091***
	(0.009)	(0.006)	(0.010)
union1	0.222**	0.068	0.186**
	(0.087)	(0.047)	(0.084)
college1	−0.065	−0.126***	0.092
	(0.106)	(0.048)	(0.100)
exper	0.050***	0.021***	0.054***
	(0.013)	(0.007)	(0.012)
expersq	−0.004***	−0.001***	
	(0.001)	(0.0004)	
I(expersq)			−0.002*
			(0.001)
Constant	0.834***	1.149***	0.446***
	(0.113)	(0.078)	(0.122)
Observations	1,545	2,229	2,229
R^2	0.038	0.020	
Adjusted R^2	0.035	0.018	
ρ			−0.998
Inverse Mills Ratio			−0.695*** (0.060)
Residual Std. Error	0.676 (df = 1539)	0.568 (df = 2223)	
F Statistic	12.106*** (df = 5; 1539)	9.207*** (df = 5; 2223)	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4:

	union	
	Dependent variable:	
	(1)	(2)
hgc	−1.009***	−1.009***
	(0.098)	(0.098)
college1	0.397	0.397
	(0.427)	(0.427)
exper	1.849***	1.849***
	(0.156)	(0.156)
married1	0.588***	
	(0.206)	
married		0.000
		(0.206)
kids	0.799***	0.000
	(0.202)	(0.202)
Constant	−6.743***	−7.330***
	(0.804)	(0.881)
Observations	2,229	2,229
Log Likelihood	−114.289	−114.289
Akaike Inf. Crit.	240.578	240.578

Note: *p<0.1; **p<0.05; ***p<0.01