PS12 Smith

April 27, 2024

1 Question 6 Response

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.5839	0.0420	37.75	0.0000
college1	0.1554	0.1003	1.55	0.1215
exper	0.0141	0.0063	2.24	0.0253
married1	-0.0538	0.0360	-1.49	0.1357
kids	0.0812	0.0352	2.31	0.0212
union1	-0.0622	0.0708	-0.88	0.3798

Experience and kids are statistically significant, but the other variables are not.

The logwage is missing 30.69 percent of the time. This data is likely MNAR. The other variables have data but income just might not have been reported for that individual (either by choice, or perhaps that individual did not have an income to report) in the data.

2 Question 7 Response

See Model summary at the end of the PDF. Within hgc, the heckman selection is the worst predictor compared to the actual .091. I suppose this isn't surprising since the Heckman selection which sets the missing values equat to zero instead of trying to estimate a value. In the Complete Cases model, all the missing values are removed, so the number of observation is smallest. However, the coefficient is the closest to actual at a .059 and measures the best significance level at .009. The Mean imputation estimates the missing value, so it has the same number of observations as the Heckman Selection. It's less accurate in prediction compared to the Complete Cases at a .036 value, but it is still significant. At least in this instance, dropping the missing variables seems to be the most appropriate estimator.

3 Question 9 Response

Average predicted probability (original model) is: 0.2373394 Average predicted probability (counterfactual policy): 1.616075

The results of the counterfactual policy do not make sense, since a probability should be between 0 and 1. Setting the other coefficients to zero does not work in this analysis. This suggests that the other variables do have some type of relationship to the union variable.

	Complete Cases	Mean Imputation	Heckman Selection
(Intercept)	0.834***	1.149***	0.446***
	0.834***	1.149***	20.553***
	(0.113)	(0.078)	(0.122)
	(0.113)	(0.078)	(1.111)
hgc	0.059***	0.036***	-1.104***
	0.059***	0.036***	0.091***
	(0.009)	(0.006)	(0.010)
	(0.009)	(0.006)	(0.066)
union1	0.222*	0.068	-1.113***
	0.222*	0.068	0.186*
	(0.087)	(0.047)	(0.084)
	(0.087)	(0.047)	(0.213)
college1	-0.065	-0.126**	-0.565*
	-0.065	-0.126**	0.092
	(0.106)	(0.048)	(0.100)
	(0.106)	(0.048)	(0.227)
exper	0.050***	0.021**	-0.506***
	0.050***	0.021**	0.054***
	(0.013)	(0.007)	(0.012)
	(0.013)	(0.007)	(0.030)

	(1)	(2)	(3)
I(exper^2)	-0.004**	-0.001**	-0.002+
	(0.001)	(0.000)	(0.001)
married1			-2.275***
			(0.162)
kids			0.495***
			(0.114)
invMillsRatio			-0.695***
			(0.060)
sigma			0.696
rho			-0.998
Num.Obs.	1545	2229	2229
R2	0.038	0.020	0.092
R2 Adj.	0.035	0.018	0.088
AIC		3808.4	
BIC		3848.4	
Log.Lik.		-1897.193	
F	12.106	9.207	
RMSE	0.67	0.57	0.66

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$