

PS7

Jack Wang

March 29, 2022

1 Data Summary

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
logwage	670	25	1.6	0.4	0.0	1.7	2.3
hgc	16	0	13.1	2.5	0.0	12.0	18.0
tenure	259	0	6.0	5.5	0.0	3.8	25.9
age	13	0	39.2	3.1	34.0	39.0	46.0

25% of logwage data is missing. It is most likely MNAR.

		N	%
college	college grad	530	23.8
	not college grad	1699	76.2
married	married	1431	64.2
	single	798	35.8

2 Model Summary

All the coefficients are smaller than the true value, because of the measurement error in the variable. The coefficient is biased towards zero. These imputation methods provide similar. Imputation with mean provides the worst results.

3 Project Progress

Previously I planned to use XBRL data from SEC, which can be get through API, to calculate accounting reporting complexity and examine the association with Non-GAAP disclosure. However, I just found there was a working paper on this idea, which have done almost exactly what I planned to do. I may need to think of another idea.

	MCAR	Mean Filling	MAR	mice
(Intercept)	0.534 (0.146)	0.708 (0.116)	0.534 (0.112)	0.651 (0.127)
hgc	0.062 (0.005)	0.050 (0.004)	0.062 (0.004)	0.057 (0.005)
collegenot college grad	0.145 (0.034)	0.168 (0.026)	0.145 (0.025)	0.121 (0.028)
tenure	0.050 (0.005)	0.038 (0.004)	0.050 (0.004)	0.039 (0.004)
I(tenure^2)	-0.002 (0.000)	-0.001 (0.000)	-0.002 (0.000)	-0.001 (0.000)
age	0.000 (0.003)	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)
marriedsingle	-0.022 (0.018)	-0.027 (0.014)	-0.022 (0.013)	-0.022 (0.015)
Num.Obs.	1669	2229	2229	2229
R2	0.208	0.147	0.277	0.218
R2 Adj.	0.206	0.145	0.275	0.216
AIC	1179.9	1091.2	925.5	1487.8
BIC	1223.2	1136.8	971.1	1533.5
Log.Lik.	-581.936	-537.580	-454.737	-735.918
F	72.917	63.973	141.686	103.234
RMSE	0.34	0.31	0.30	0.34