





Problem Set 7

Chang Gao

Q6

The rate of missing log-wages is 24.93%, I guess the `logwage` variable is MAR since the amount of missing value is a lot..

The following table is the output for Q6.

	Unique	Missing Pct.	Mean	SD	Min	Median	Max	Histogram
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	
		N			%			
college	college grad	530			23.8			
	not college grad	1699			76.2			
married	married	1431			64.2			
	single	798			35.8			

Q7

Multiple imputation provides estimates closer to the true value of 0.093 compared to mean imputation. The results from complete case analysis and imputing missing `logwages` with predicted values from the complete cases regression are almost identical. Compared to the true β_1 , due to approximately 25% of log wage values being missing, the complete case regression may have introduced selection bias.

Impute by predicted values from regression seems to be not reliable as it is heavily affected by missing data; multiple imputation seems to be good in this case.

Table 1: Comparison of Imputation Methods for Missing Logwage Data

	Complete Cases	Mean Imputation	Regression Imputation	Multiple Imputation
(Intercept)	0.534*** (0.146)	0.708*** (0.116)	0.534*** (0.112)	0.596** (0.174)
hgc (β_1)	0.062*** (0.005)	0.050*** (0.004)	0.062*** (0.004)	0.059*** (0.006)
collegenot college grad	0.145*** (0.034)	0.168*** (0.026)	0.145*** (0.025)	0.102* (0.037)
tenure	0.050*** (0.005)	0.038*** (0.004)	0.050*** (0.004)	0.049*** (0.006)
tenure_sq	-0.002*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
age	0.000 (0.003)	0.000 (0.002)	0.000 (0.002)	0.001 (0.003)
marriedsingle	-0.022 (0.018)	-0.027* (0.014)	-0.022+ (0.013)	-0.020 (0.018)
Num.Obs.	1669	2229	2229	2229
R2	0.208	0.147	0.277	0.232
R2 Adj.	0.206	0.145	0.275	0.230

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

Q8

For my ongoing China housing price study, I don't have a lot progress for traveling during spring break. Here are my progress:

(1) Analysis of policy dummies (housing purchasing restriction and government restrictions on real estate developers' borrowing). I created purchasing restriction dummies for each city that are not completely identical across cities, taking the value of 1 when purchasing restrictions are in effect. I also constructed similar dummies for provincial-level purchasing restriction policies. The preliminary results show that city-level purchasing restrictions have a significant positive impact of approximately 3% on commercial housing prices. The impact of provincial-level policies is not significant, which is understandable since it's generally lower-income provinces that issue purchasing restriction policies without sufficient implementation power. The government's lending restrictions on real estate developers have a significant impact of 0.05% on housing prices, which is very low, and I'm not sure if there's an economic intuition for this. Regarding the 3% positive impact of city-level purchasing restriction policies on housing prices, I need to further consider the causal relationship, because cities with rapidly rising prices are the ones that implement purchasing restriction policies, so the correlation coefficient between policy and housing prices can easily be positive. Do I need to decompose the suppressing effect of the policy on housing prices?

(2) After consulting with Dr. Yang, I plan to conduct a simple analysis of the spillover effect of purchasing restrictions in large cities on housing prices in small cities, and to construct a cross-sectional instrumental variable related to the one-child policy from about 25 years ago. This one-child policy was strictly enforced in large cities, but in small cities, it was not strictly enforced and was even relaxed. Considering China's household registration system and frictions in population mobility between cities, this variable would affect the current working population, but may not directly affect housing prices, so it might be an appropriate instrumental variable. I will try this approach.

(3) For the final project of this course, I plan to examine the distribution of household financial wealth from the CFPS (China Family Panel Survey) data, such as whether the distribution of household financial wealth has become more unequal from the past to the present across different cities? What are the differences in the rightward shift of these distributions among cities of different sizes?