**#Q1**

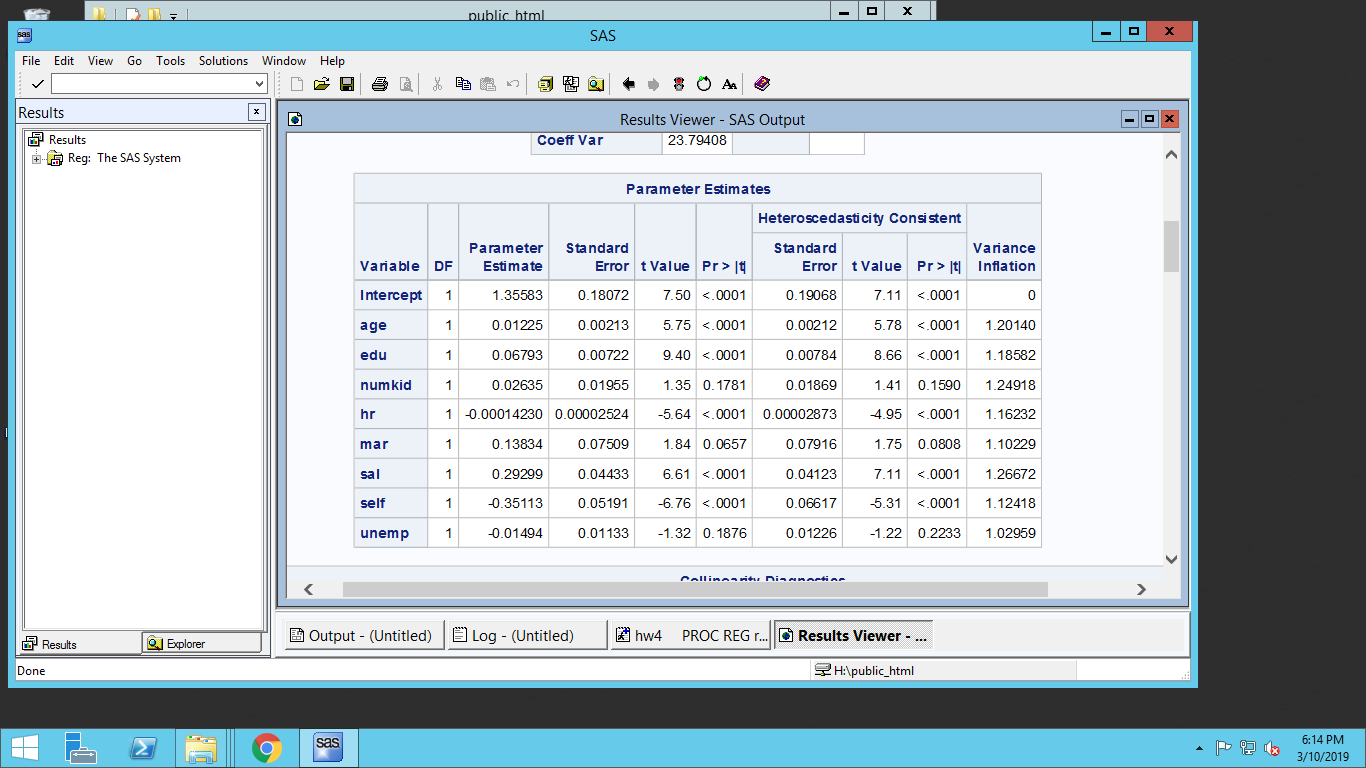
1.

Find the best linear regression model. Check for multicollinearity and take appropriate actions.

We ran the OLS regression with age, edu, numkid, hr, mar, sal, self, unemp as independent variable and performed White’s test and Breusch-Pagan test to detect heteroskedasticity. The P values on both tests were smaller than 0.05 so we rejected the null hypothesis and concluded that there’s heteroskedasticity.

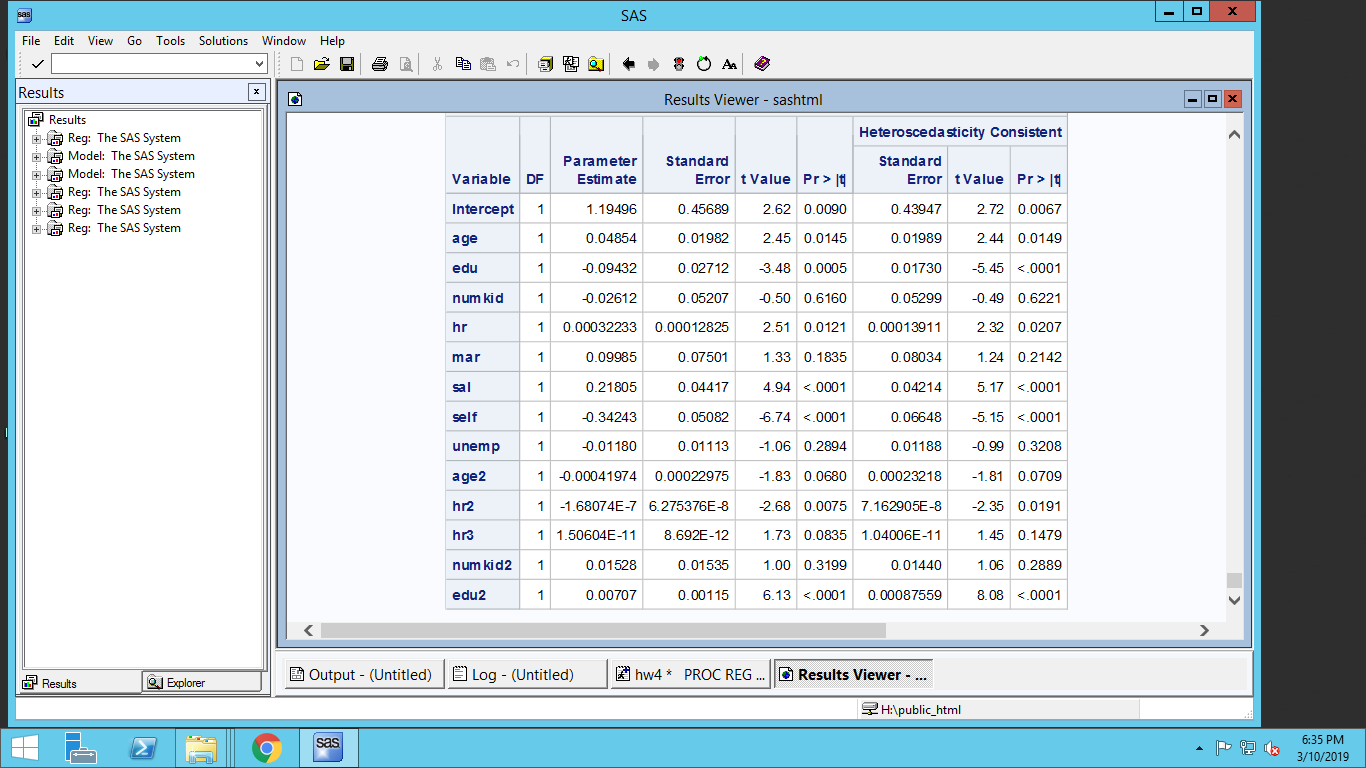
Therefore, we used heteroskedasticity consistent standard error and found that the coefficients on age, edu, hr, sal and self were significantly different from 0. The coefficients on numkid, mar and unemp are not significantly different from 0, suggesting that these variables do not affect ln(wage).

The VIF, Eigenvalue and condition index shows that multicollinearity is not a problem in this regression.



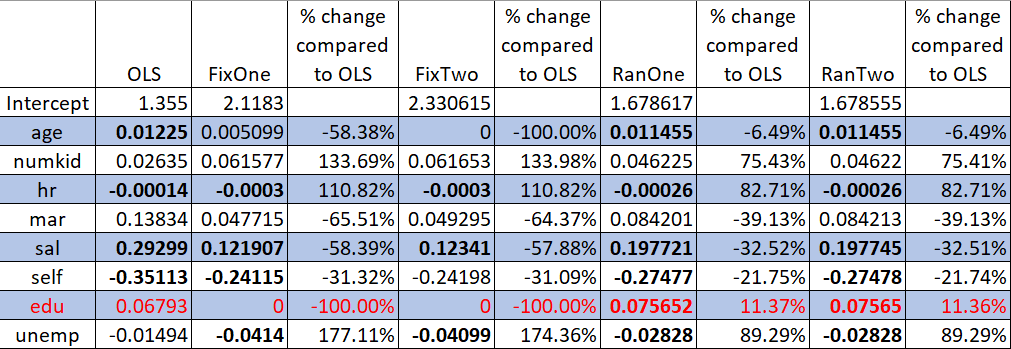
2.

We added age^2 hr^2 hr^3 numkid^2 edu^2 into our regression and found that the coefficient on hr2 and edu2 are significantly different from 0 (P value<0.05). The R-square of this model is 0.2939. Therefore, we conclude that hr and edu have nonlinear effect on ln(wage).



3. & 4.

We ran the OLS model, Fixone, Fixtwi, Ranone, Rantwo models with age, edu, numkid, hr, mar, sal, self, unemp as independent variables:



R-square of OLS model is 0.2498, which means that the independent variables can explain 24.98% the variations in the dependent variable which is ln(wage). R-square of the model with non-linear effects improved slightly to 0.2939.

In OLS model, the significant parameters are age, education, work hours per year, salary dummy variable, and self-employed dummy variable. While age, salary have positive effect on wages, work hours per year and self-employed have negative effect on wages. When we add the non-linear effect of work hours per year and education, age, education, work hours per year, salary dummy variable, self-employed dummy variable, hour-squared, and education-squared are the significant parameters.

For the OLS regression model, we perform various options such as VIF, COLLIN, White test and Breusch-Pagan test. Based on the results we get, VIF, Eigenvalue and condition index shows that multicollinearity is not a problem in the OLS model. However, because p-value from the White test and Breusch-Pagan test are less than 0.05, there is heteroskedasticity problem existing in the model. We therefore used heteroskedasticity-robust standard errors to solve this problem.

Coefficient interpretation according to the OLS regression:

Ceteris paribus, when a person is 1 year older, the wage per hour is expected to increase by 1.23%

Ceteris paribus, a person who work 1 more hour/year is expected to have a 0.014% lower wage/hour

Ceteris paribus, a person who receive salary is expected to have wage/hour that is 29.299% higher than otherwise.

Ceteris paribus, a self-employed person is expected to have wage/hour that is 35.11% lower than otherwise.

Ceteris paribus, a person with 1 more year of education is expected to earn 6.79% more wage/hour.

Numkid, mar and unemp do not have significant impact on ln(wage)

5.

The table above shows the coefficients of parameters among OLS, FixOne, FixTwo, RanOne, and RanTwo models, where:

* FIXONE - One way fixed effects – cross section heterogeneity
* FIXTWO - Two way fixed effects – cross section and time series heterogeneity
* RANONE - One way random effects - cross section heterogeneity
* RANTWO – Two way random effects cross section and time series heterogeneity

The table in part 3 and 4 also shows the changes in coefficients of parameters across four different panel models as well as the percentage changes in coefficients in panel models compared to the OLS model. The coefficients don’t change much between FixOne and FixTwo models, same for coefficients in RanOne and RanTwo models.

In the panel models, age, married status, salary, self-employment have less effect on wages, compared to the OLS model. On the other hand, number of children, work hours per year, and local unemployment percentage in the panel models has greater impact on wages than it does in the OLS model.

6.

Based on the table above, we can see that because education and age are time-invariant variables, the FixOne effect model does not capture the effect of education on wages, and the FixTwo model does not capture the effects of education and age on wages. Therefore, the coefficient of education in FixOne model is 0, and that of education and age in FixTwo model are 0. On the other hand, random effect models can estimate the effect of time-invariant variables (education) on wages. Therefore, compared to the OLS, the coefficient of education in random effect models increases by 11.37% from 0.068 to 0.076. In other words, when we control heterogeneity by using random effects, the education positively has more effect on wages compared to the OLS model.

**#Q2**

1. Run the 2SLS model using SAS (PROC SYSLIN) and estimate the effect of pioneering on market share. Be sure to consider the direct effects as well as the indirect effects. (read the paper on pioneering advantages for this interpretation).

2SLS model:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Coefficient table | Equations | | | | |
| Variable | MS | Qual | PLB | Price | DC |
| **Intercept** | a1= 39.52133 | b1=-265.494 | c1=109.0706 | d1=100.314 | g1=1.140754 |
| **qual** | **a2=0.508689** |  |  | **d2=0.141663** | g2=0.035383 |
| **plb** | **a3=-1.00943** |  |  |  |  |
| **price** | a4=0.875453 | **b2=2.595316** |  |  |  |
| **pion** | **a5=7.17536** | b3=-0.39839 | c2=1.715145 | d3=1.661022 | g3=-0.07697 |
| **ef** | **5.791929** | -2.23599 | -0.12958 | 0.070665 | 0.140887 |
| **phpf** | 0.572868 |  |  |  |  |
| **plpf** | 0.166099 |  |  |  |  |
| **psc** | **-30.8958** |  |  |  |  |
| **papc** | -1.46455 |  |  |  |  |
| **ncomp** | **-7.55417** |  |  |  |  |
| **mktexp** | -0.29136 | **-0.48914** |  | **0.224952** |  |
| **dc** |  | **10.47285** | **-8.73302** | -0.45759 |  |
| **tyrp** |  | 0.187802 | -0.29136 | 0.979397 | 0.221522 |
| **pnp** |  | **0.211277** | **0.054686** | -0.02127 |  |
| **custtyp** |  |  | **3.940911** |  |  |
| **ncust** |  |  | 0.225838 |  |  |
| **custsize** |  |  | 0.520397 |  |  |
| **ms** |  |  |  | -0.01764 | 0.004963 |
| **penew** |  |  |  |  | -0.0034 |
| **cap** |  |  |  |  | 0.000041 |
| **rbvi** |  |  |  |  | -0.04885 |
| **emprody** |  |  |  |  | **0.002523** |
| **union** |  |  |  |  | 0.00146 |

***Note: a(i); b(i); c(i); d(i) are defined in the table above.***

Effect of PION on MS: X = a5 + a2\*b3 + a3\*c2 +a4\*d3+ 0.57\*phpf + 0.166\*plpf - 30.90\*psc -1.46\*papc

→ X1 = 6.6955 +0.57\*phpf + 0.166\*plpf - 30.90\*psc -1.46\*papc

This suggests that when a firm is a pioneer, that firm is expected to have market share of X1 percentage higher than non-pioneer firms.

2. Run a simple regression model of market share as given in the first equation. What is the effect of pioneering on market share using this simple model? How does this effect change across different models.

In a simple OLS model, the effect of PION on MS is:

X2= 9.85 + 1.50\*phpf + 1.15\*plpf - 20.90\*psc -1.11\*papc

This suggests that when a firm is a pioneer, that firm is expected to have market share of X2 percentage higher than non-pioneer firms. After we run the 2SLS model with instrumental variables to solve endogeneity, the effect of PION on MS is much less positive.