**Homework assignment 4**

Q1. I have given you a panel data on wages (Wage data) in which N=334, T=3 years (1984-1986).

For each ID, the data is sorted by year. You need to create the ID and year variables.

|  |  |  |
| --- | --- | --- |
| **Columns** | **Variable name** | **Description** |
| C1 | Edu | Education in years |
| C2 | Hr | Work hours per year |
| C3 | Wage | Dollar wage per hour |
| C4 | Famearn | Family earnings in dollars per year |
| C5 | Self | Dummy for self-employed |
| C6 | Sal | Dummy for salaried |
| C7 | Mar | Dummy for married |
| C8 | Numkid | Number of children |
| C9 | Age |  |
| C10 | unemp | Local unemployment percentage |

We need to do a regression to understand the determinants of “natural log (wages)” that is {ln(wage)}.

We need to understand the effect of the following variables: age, edu, numkid, hr, mar, sal, self, unemp.

1. Find the best linear regression model. Check for multicollinearity and take appropriate actions.
2. Develop a model to test if there are nonlinear effects for some variables. Which variables have non-linear effect on ln(wages).
3. Using the same model, run fixed effects models and random effects models

i.e., FIXEDONE, FIXEDTWO, RANONE, RANTWO.

Create a table of coeffic

ients side-by side with significant coefficients shown in bold (you may do this in Excel).

1. Write a report on your findings. Interpret model fit, t-values, meaning of coefficients, collinearity diagnostics, White test, Breusch-Pagan test etc.
2. What is the effect of panel data models on the coefficients. What parameters have changed and by what percentage?
3. We are especially interested in the effect of education on wages. Notice how much (%) has this coefficient changed across the different models?

Q2. I have provided a dataset PIMS.dat which has data on industrial goods manufacturers. The variables in the data are in the following order. These variables and definitions are given in the paper by Robinson and Fornell (1985) on pioneering advantages (see Tables 1, 2 and 3). As in the paper by Robinson and Fornell (1985), we will estimate a simultaneous system of five equations. While the paper considered consumer goods industries, we are interested in replicating the analysis for industrial goods industries.

|  |  |
| --- | --- |
| MS | Relative market share |
| QUAL | Relative quality |
| PRICE | Relative price |
| PLB | Product line width |
| DC | Relative direct costs |
| PION | Whether a firm is a pioneer (1) or not (0) |
| EF | Whether a firm is an early follower (1) or not (0) |
| PHPF | Pioneer \*high purchase frequency |
| PLPF | Pioneer \*low purchase frequency |
| PSC | Pioneer \*seasonal product change |
| PAPC | Pioneer \*annual/periodic product change |
| NCOMP | Number of competitors |
| MKTEXP | Relative marketing expenditures (similar to ‘relative advertising and promotion’) |
| TYRP | Twenty year pioneer |
| PNP | Percentage of new products |
| CUSTTYP | Relative customer type |
| NCUST | Relative Number of customers |
| CUSTSIZE | Relative customer size |
| PENEW | Plant and equipment newness |
| CAP | Capacity utilization |
| RBVI | Relative backward vertical integration |
| EMPRODY | Employee productivity |
| UNION | Percentage of employees unionized |

Please estimate a 2SLS model with the following five equations.

model MS=qual plb price pion ef phpf plpf psc papc ncomp mktexp

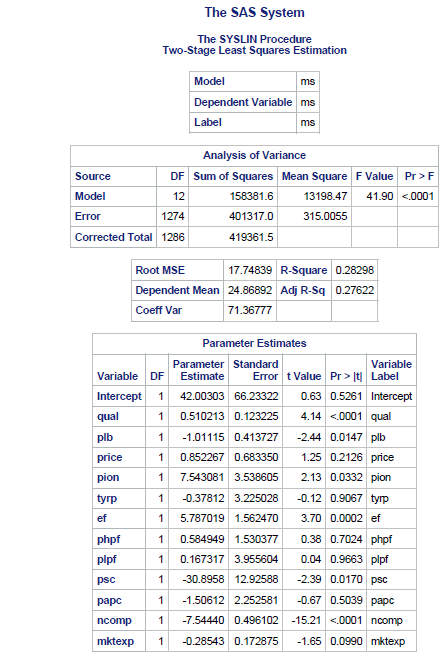
model Qual=price dc pion ef tyrp mktexp pnp

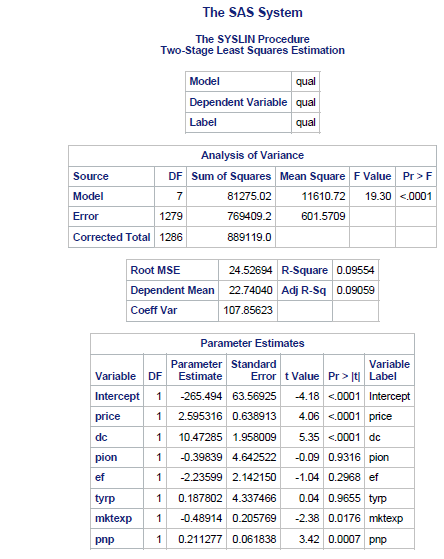
model PLB=dc pion tyrp ef pnp custtyp ncust custsize

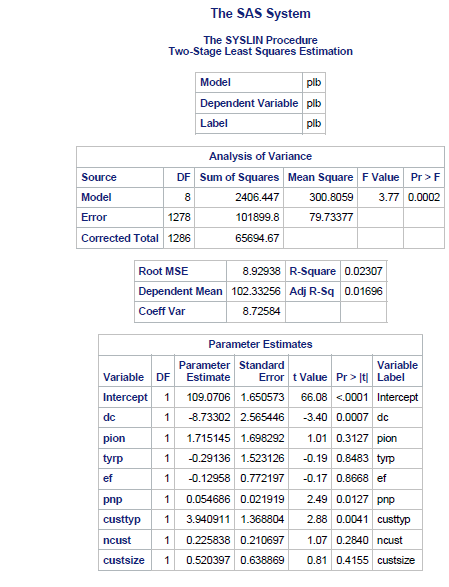
model Price=ms qual dc pion ef tyrp mktexp pnp

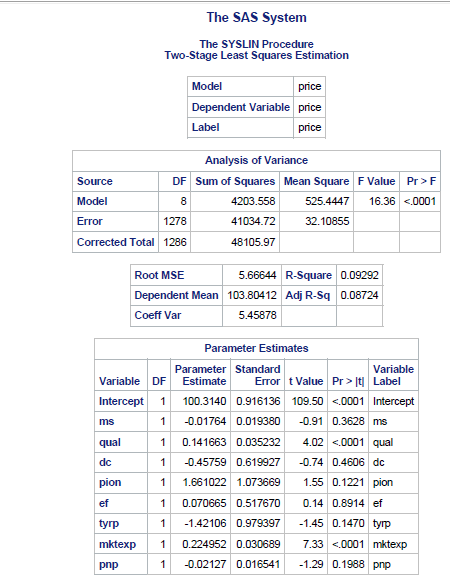
model DC=ms qual pion ef tyrp penew cap rbvi emprody union

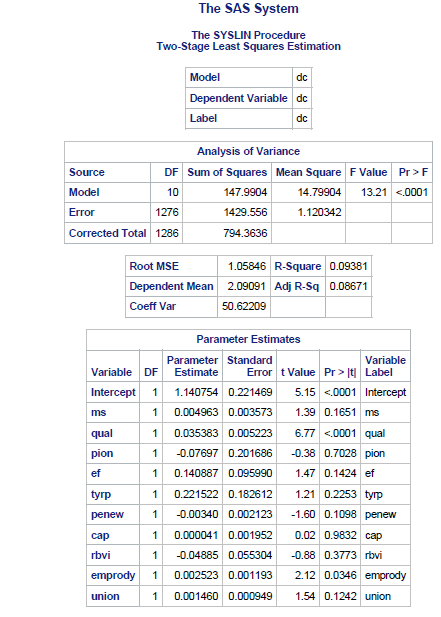
1. **Run the 2SLS model using SAS (PROC SYSLIN) and estimate the effect of pioneering on market share. Be sure to consider the direct effects as well as the indirect effects. (read the paper on pioneering advantages for this interpretation).**











Endogenous Variables: MS, QUAL, PLB, PRICE, DC; hence 5 simultaneous equations.

Instrument Variables: PION, EF, PHPF, PLPF, PSC, PAPC, NCOMP, MKTEXP, TYRP, PNP, CUSTTYP, NCUST,

Product quality, Product Line Breadth, If a business is a market pioneer, Whether a firm is an early follower, Pioneer selling seasonally changed goods, and Number of competitors are significant (i.e. |t-value|>1.96 and p-value<0.05) at 95% confidence level.

Direct Effect: Pioneering is a significant variable. (t value = 2.13 and p-value = 0.0332) so there is an effect of pioneering on market share. If the firm is a pioneer, the market share will increase by 7.54 units. Also, if a firm is a pioneer and sells goods that are changed seasonally then the market share will decrease by 30.89 units.

Indirect Effect: (where pion = 1)

PION-QUAL-MS: -0.39839\* 0.510213= -0.20

PION-PLB-MS: 1.715145\*(-1.01115) = -1.73

PION-PRICE-MS: 1.661022\*0.852267 = 1.41

PION-DC-PLB-MS: -0.07697\*(-8.73302)\*(-1.01115)= -0.68

PION-DC-PRICE-MS: -0.07697\*(-0.45759)\*(0.852267)= 0.03

PION-DC-QUAL-MS: -0.07697\*(10.47285)\*(0.510213)= -0.41

Total Effects = 5.9

1. **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.**

2. In Linear Regression, pioneering has positive and significant effect (t value = 4.88 and p-value < 0.0001) at 95% significance level. If a firm is a pioneer, then market share increases by 12.602 units.

Difference between Two Stage Least Squares and Linear Regression:

a. Pioneering is significant in both the models, but its coefficient is high in simple regression (12.6) as compared to 2SLS (7.54 direct and 5.9 indirect). This is because there is endogeneity due to simultaneity occurring in MS, QUAL, PLB, PRICE, DC in simple regression which is corrected through 2SLS. Simple Regression is violating the unbiased assumption.

b. Price is significant in the simple regression model since it is correlated with Quality. This endogeneity is removed by 2SLS and so it becomes insignificant.

c. Considering 95% significance level, Product Quality, Product Line Breadth, Whether firm is an early follower, Pioneer Seasonal Product Change, Number of competitors are significant in both the models but have increased coefficients in simple regression confirming endogeneity as variables are showing effects of relation with each other. The remaining variables remain insignificant in both 2SLS and simple regression model.

2 SLS model has resolved endogeneity and unobserved heterogeneity and is giving unbiased results.