

Q1. Find the best linear regression model. Check for multicollinearity and take appropriate actions.

The best linear regression model is with the below mentioned independent variables:

age edu hr mar sal self Famearn

R² value is 0.6036

On testing for multicollinearity with VIF, COLLIN and STB functions, we found that the Condition Index was <100 for and the Variance Inflation Factor was less than 10(<10) for all the variables. Which helps us in concluding that there is no multicollinearity in the data.

Further, the R² value is 0.6078 to suggest that independent variables explain 60% of the variation in the dependent variable. There are still many variables that have not been observed and there might be heterogeneity present.

Q2. Develop a model to test if there are nonlinear effects for some variables. Which variables have non-linear effect on ln(wages)?

After studying we find that age variable has a non-linear effect on wage, suggesting wage will have a depreciation effect after reaching the saturation point for age. We observe the effect on the coefficient of Age_sq variable in the model to see it's effect.

Q3. Write a report on your findings. Discuss model fit, t-values, managerial meaning of coefficients, collinearity diagnostics, heteroscedasticity tests using White test, and Breusch-Pagan test. (use PROC Model here). There is no need to do Weighted Least Squares estimation. Just use White heteroscedastic corrected standard errors for your interpretation.

- **Model Fit:** We achieved an overall R² value of 0.6078, there could still be some heterogeneity in the model when looking at wage and the independent variables that help estimate that. The model fit is decent and there could be some unobserved factors still present in the model.
- **T-Value:** Age has a t-value of 3.72 and edu has a t-value of 2.61 which proves that these 2 independent variables are significant.

Coefficients: When we use **age edu hr mar sal self Famearn Kid1 Age_Sq**, as independent variables explaining wage we see the following effects:

Age and Age_sq increases the wage by 1.2%

Education increases the wage by 1.5%

Hr has an effect close to zero

Marriage causes the wage to an decrease by 16%, that is if a person is married he likely to have a wage 17% less.

A **Salaried** person is likely to have a wage 14% more than a non-salaried person.

A **Self-employed** person is likely to earn a wage 24% lower than a non-self-employed person.

Famearn is also close to zero

Kid1 if a family has 1 kid the wage goes down by 7%

All the variables included in the model are significant when we run Simple Linear Regression.

Using Fix-Effects Model we observe an R^2 value of 0.6098, which is 61% of the variation in the model is explained.

Being married is highly insignificant in this model as it is a slow changing variable over time and such slow variables generally can't be significantly used to estimate dependent variables in a fixed effects model.

In terms of **Heteroscedasticity, White test, and Breusch-Pagan test** help us conclude that there is evidence of heteroscedasticity in the model as we have rejected the null hypothesis in the test.

| Heteroscedasticity Test | | | | | |
|-------------------------|---------------|-----------|----|------------|--|
| Equation | Test | Statistic | DF | Pr > ChiSq | Variables |
| lw | White's Test | 221.6 | 59 | <.0001 | Cross of all vars |
| | Breusch-Pagan | 85.29 | 10 | <.0001 | 1, edu, age, self, sal, hr, mar, kid1, Edu_sq, Age_sq, Famearn |

Q4. Using the same model as above (Q1), run fixed effects models and random effects models

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

Create a table of coefficients side-by side with significant coefficients shown in bold (you may do this in Excel).

| Variable | FixOne | FixTwo | RanOne | RanTwo |
|-----------|---------|----------------|----------------|----------------|
| Intercept | -0.8708 | 1.25952 | 4.65515 | 1.25955 |

| | | | | |
|--------|---------|---------|---------|---------|
| edu | 0 | -0.091 | 0 | -0.091 |
| age | 0.12903 | 0.06978 | 0 | 0.06978 |
| self | -0.2155 | -0.2662 | -0.2166 | -0.2662 |
| sal | 0.117 | 0.1822 | 0.11868 | 0.18221 |
| hr | -0.0003 | -0.0003 | -0.0003 | -0.0003 |
| mar | 0.12531 | 0.1391 | 0.12693 | 0.13911 |
| kid1 | -0.1313 | -0.1195 | -0.1312 | -0.1195 |
| Edu_sq | 0 | 0.00713 | 0 | 0.00713 |
| Age_sq | -0.0014 | -0.0007 | -0.0013 | -0.0007 |

Q5. What is the effect of panel data models on the coefficients? What parameters have changed and by what percentage?

The observation after running panel data gives better insight as to how the parameters are changing after considering time series as well as cross-sectional heterogeneity.

After running the fixed effect model - fixone and fixtwo and random effect model- ranone and rantwo model on the data we find multicollinearity. The white test and Bresuch Pagan test tell there is multicollinearity and same can be seen in the Hausman test also. So random effect model cannot be used on the data.

Interpreting the fixed effect model, the changes in coefficient values from fixone to fixtwo are mentioned below:

Edu: The value is zero for both the models

Age: Drop from 12.9% to 0

Self: The value decrease by from -0.2155 to -0.2166 so there is a change of 0.11 %

Sal: Increase from 0.117 to 0.1186 there is a change of 0.16 %

Hr: There is no change

Mar: changed from 0.12531 to 0.12693. There is a change of 0.16%

Kid1: There is a negligible change.

Edu_Sq: Has no change

Age_sq: There is negligible change 0.0014 to 0.0013

The effect of proc panels models is that the from fix one to fix two the value of coefficient Sal, Mar has increased and for self, it has decreased.

Q6. We are especially interested in the effect of education on wages. How much (%) has this coefficient changed across the different models? What is the correct estimate of the effect of education on wages?

With only Linear Regression, can Education variable be estimated. In the Fixed Effects model, there is no change and will be no change as it is a time invariant variable. This is demonstrated in the fixed effects model.

The changes that we see from fixed effects model are not too far apart.