Introductory Econometrics

Tutorial 9

PART A: To be done before you attend the tutorial. The solutions will be made available at the end of the week.

1. (a)

$$
\begin{aligned}
H_0 &: \delta_0 = \delta_1 = 0 \\
H_1 &: \text{ at least one of } \delta_0 \text{ or } \delta_1 \text{ is not zero} \\
& \quad \text{Regress } \log(wage) \text{ on a constant and } totuni \text{ to get } R_r^2 \text{ or } SSR_r \\
& \quad \text{Use } R^2 \text{ or } SSR \text{ of the unrestricted equation estimated above} \\
F &= \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(6763-4)} = \frac{\left(R_{ur}^2 - R_r^2\right)/2}{\left(1-R_{ur}^2\right)/(6763-4)} \sim F_{2,6759} \\
& \quad \text{Obtain } F_{crit} \text{ for the 5\% level of significance from the table or software} \\
& \quad \text{Reject the null if } F_{calc} \text{ is larger than } F_{crit}, \text{ do not reject otherwise}
\end{aligned}
$$

(b) The estimated model shows that with no university education, women on average get paid 30% lower than men. This is because $100 \times \left(e^{-0.360} - 1\right) = -30.23$. The estimated model also finds that the return to every year of university education is 5% for men, and it is 8% for women, and this difference is statistically significant. This is because the coefficient of $female_i \times totuni_i$ is 0.03 with a t-value of $\frac{0.03}{0.005} = 6$, which gives us more than 95% confidence that this is larger than zero. So, university education levels the playing field somewhat.

(c)

$$
\begin{aligned}
3.289 + 0.05 totuni^* &= (3.289 - 0.36) + (0.05 + 0.03) totuni^* \\
0.36 &= 0.03 totuni^* \\
totuni^* &= \frac{0.36}{0.03} = 12
\end{aligned}
$$

This predicts that women need 12 years of university education before they catch up with men. This is most probably a prediction outside of the range of data, because even a PhD takes less than 12 years of university education, and even though the sample is large, it is unlikely that it included many people with 12 years of university education. The prediction of regression models is only accurate within the normal range of data (it is most accurate for the sample average). The predictions of regression models for values of independent variable that is outside of the normal range for that variable should not be taken seriously.

(d) This means errors are heteroskedastic. While the OLS estimator is still an unbiased estimator, it is no longer BLUE. Also the standard errors reported above, which were calculated based on the usual formula for standard errors, are not correct and cannot be used as a basis for hypothesis testing.

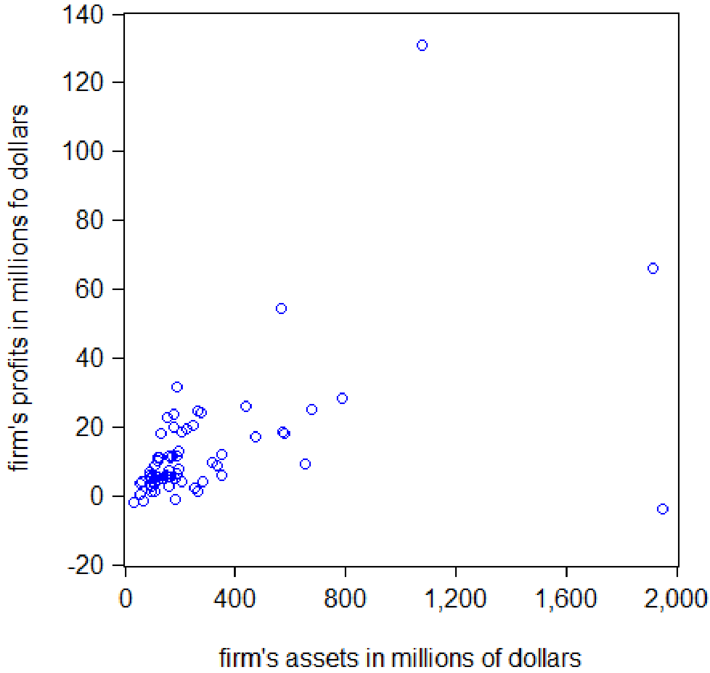
(e)

$$
\begin{aligned}
\text{Let } w_i &= \frac{1}{\sqrt{female_i + 1}} = \begin{cases} 1 & \text{if person } i \text{ is a man} \\ \frac{1}{\sqrt{2}} & \text{if person } i \text{ is a woman} \end{cases} \\
w_i \times \log(wage_i) &= \beta_0 w_i + \delta_0 w_i \times female_i + \beta_1 w_i \times totuni_i + \delta_1 w_i \times female_i \times totuni_i + w_i \times u_i
\end{aligned}
$$

This means that the model will stay as is for men, and it will be multiplied by $\frac{1}{\sqrt{2}}$ for women. This makes the error term for women $\frac{u}{\sqrt{2}}$, which has variance $Var\left(\frac{u}{\sqrt{2}} \mid totuni, female = 1\right) =$

$\frac{1}{2}Var(u \mid totuni, female = 1) = \frac{1}{2} \times 2\sigma^2 = \sigma^2$, which is now the same as the variance for men. Hence the heteroskedasticity problem is solved. The significance of this is that the OLS estimator of the parameters in this transformed equation will be BLUE, and the standard errors will provide reliable inference.

2. (a)



It suggests a positive slope.

(b) It suggests that the assumption of homoskedasticity does not apply here because the dispersion of profits around the regression line is larger for firms with larger assets. Therefore, it suggests that the OLS estimator will be unbiased but not the best linear unbiased estimator of the intercept and slope parameters.

Do not forget to bring your answers to PART A and a copy of the tutorial questions to your tutorial.

<u>Part B:</u> This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.

1. The general model is:

$$profits_i = \beta_0 + \delta_0 mno_i + \beta_1 assets_i + \delta_1 (mno_i \times assets_i) + u_i \tag{1}$$

(a)

$$\begin{aligned} H_0 &: \quad \delta_0 = \delta_1 = 0 \\ H_1 &: \quad \text{at least one of } \delta_0 \text{ or } \delta_1 \text{ is not zero} \end{aligned}$$

(b)

$$\begin{aligned} \widehat{profits} &= \underset{(2.78)}{1.56} + \underset{(4.09)}{8.23} mno + \underset{(0.006)}{0.05} \, assets - \underset{(0.009)}{0.05} \, mno * assets \\ n &= 69, \quad R^2 = 0.52 \end{aligned}$$

i. Breusch-Pagan test when the alternative hypothesis is $Var\left(u_i \mid mno_i, assets_i\right) = \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i$

$$\begin{aligned} H_0 &: Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \text{ for all } i \\ H_1 &: Var\left(u_i \mid mno_i, assets_i\right) = \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i \end{aligned}$$

Estimated auxiliary regression:

$$\begin{aligned} \hat{u}^2 &= 12.84 + 0.81 assets - 157.47 mno \\ n &= 69, \quad R^2_{\hat{u}} = 0.183 \\ BP &= nR^2_{\hat{u}} \sim \chi^2_2 \quad \text{under } H_0, \ BP_{crit} = 5.99 \text{ at the 5\% l.o.s.} \\ BP_{calc} &= 69 \times 0.183 = 12.63 \text{ (may not be exactly the same as EViews due to rounding)} \\ BP_{calc} &> BP_{crit} \Rightarrow \text{We reject the null and conclude that errors are HTSK} \end{aligned}$$

Notation sometimes bugs students. For example, here under the null I have denoted the variance by σ^2, but under the alternative, if α_1 and α_2 are zero, the variance will be the same for all observations, but it is denoted by α_0. Of course notation is arbitrary, but it can confuse some. If that is the case, you may want to denote the variance under the null α_0.

ii. White test.

$$\begin{aligned} H_0 &: Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \text{ for all } i \\ H_1 &: Var\left(u_i \mid mno_i, assets_i\right) \text{ is a smooth function of } mno_i \text{ and } assets_i \end{aligned}$$

Estimated auxiliary regression:

$$\begin{aligned} \hat{u}^2 &= -470.11 + 536.17 mno + 3.33 assets - 0.001 assets^2 - 3.29 mno * assets + 0.001 mno * as \\ n &= 69, \quad R^2_{\hat{u}} = 0.372 \\ W &= nR^2_{\hat{u}} \sim \chi^2_5 \quad \text{under } H_0, \ W_{crit} = 11.07 \text{ at the 5\% l.o.s.} \\ W_{calc} &= 69 \times 0.372 = 25.67 \text{ (may not be exactly the same as EViews due to rounding)} \\ W_{calc} &> W_{crit} \Rightarrow \text{We reject the null and conclude that errors are HTSK} \end{aligned}$$

Note that mno^2 is the same as mno, so we do not include it twice (otherwise we get exact multicollinearity).

iii. The special form of the White test that uses the predicted value of $profits$ and its square as predictors of variance.

$$\begin{aligned} H_0 &: Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \text{ for all } i \\ H_1 &: Var\left(u_i \mid mno_i, assets_i\right) \text{ is a smooth function of } mno_i \text{ and } assets_i \end{aligned}$$

Note that the null and the alternative is the same as what they were for the regular White test. Estimated auxiliary regression:

$$\begin{aligned} \hat{u}^2 &= -529.70 + 65.45 \widehat{profits} - 0.43 \widehat{profits}^2 \\ \widehat{profits} &= profits - \hat{u} \text{ is an easy way to compute } \widehat{profits} \\ n &= 69, \quad R^2_{\hat{u}} = 0.369 \\ W &= nR^2_{\hat{u}} \sim \chi^2_2 \quad \text{under } H_0, \ W_{crit} = 5.99 \text{ at the 5\% l.o.s.} \\ W_{calc} &= 69 \times 0.369 = 25.46 \\ W_{calc} &> W_{crit} \Rightarrow \text{We reject the null and conclude that errors are HTSK} \end{aligned}$$

(c) No. $profits$ can be negative, so logarithmic transformation is not an option.

(d) It is good to go over why these waiting cures HTSK at least for one of them.

i. $Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \times assets_i \rightarrow w_i = \frac{1}{\sqrt{assets_i}}$

ii. $Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \times assets_i^2 \rightarrow w_i = \frac{1}{assets_i}$

iii. $Var\left(u_i \mid mno_i, assets_i\right) = \sigma^2 \log(assets_i) \rightarrow w_i = \frac{1}{\sqrt{\log(assets_i)}}$

(e) Generate $w = \frac{1}{\sqrt{assets}}$. Use that to estimate the unrestricted weighted model:

$$\begin{aligned} \widehat{w * profits} &= 0.123w + 2.50w * mno + 0.056w * assets - 0.031w * mno * assets \\ n &= 69,\ SSR_{ur} = 25.32 \end{aligned}$$

And the restricted weighted model

$$\begin{aligned} \widehat{w * profits} &= 0.123w + 0.056w * assets \\ n &= 69,\ SSR_r = 27.94 \end{aligned}$$

$$\begin{aligned} H_0 &: \delta_0 = \delta_1 = 0 \\ H_1 &: \text{ at least one of } \delta_0 \text{ or } \delta_1 \text{ is not zero} \\ F &= \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(69-4)} \sim F_{2,65} \qquad \text{under } H_0 \\ F_{crit} &= 3.15 \text{ for the 5\% level of significance using } F_{2,60} \text{ from the table} \\ F_{calc} &= \frac{(27.94 - 25.32)/2}{25.32/65} = 3.36 \\ F_{calc} &> F_{crit}, \text{ so we reject the null, and we conclude that the structure of ownership} \\ & \quad \text{has some effect the relationship between profits and assets} \end{aligned}$$

It is interesting to compare WLS results with OLS results because even though at the end we do conclude that there is some difference between owner managed and non-owner managed firms, the difference is not as dramatic as what the OLS estimates tell us.

2. The wisdom of WLS:

(a) If data for all 10 individuals were reported, then we knew that the best linear unbiased estimator for the population mean would be the sample average of these 10 numbers, i.e. $\frac{1}{10}\sum_{i=1}^{10} w_i$. Even though the raw data for the last 8 individuals are not reported, we still know that $wbar_1 = \frac{1}{4}\sum_{i=3}^{6} w_i$ and $wbar_2 = \frac{1}{4}\sum_{i=7}^{10} w_i$, so we can get $\sum_{i=3}^{6} w_i = 4 \times wbar_1$ and $\sum_{i=7}^{10} w_i = 4 \times wbar_2$. So from the reported data $\{w_1, w_2, wbar_1, wbar_2\}$ we can still compute the best linear unbiased estimator consider to be the best estimate for μ? *(Tutors will manage the discussion and lead you to the desired answer!)*

$$\frac{1}{10}\sum_{i=1}^{10} w_i = \frac{1}{10}\left(w_1 + w_2 + 4 \times wbar_1 + 4 \times wbar_2\right)$$

and we know that this is the best linear unbiased estimator of μ (because it is the OLS estimator in the model $w_i = \mu + u_i$).

(b) In

$$\begin{pmatrix} w_1 \\ w_2 \\ wbar_1 \\ wbar_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}$$

the OLS estimator of β_0 is the mean of the dependent variable, so

$$\begin{aligned}
\hat{\beta}_0 &= \frac{1}{4}\left(w_1 + w_2 + wbar_1 + wbar_2\right) \\
&= \frac{1}{4}\left(w_1 + w_2 + \frac{1}{4}\sum_{i=3}^{6} w_i + \frac{1}{4}\sum_{i=7}^{10} w_i\right) \\
\Rightarrow \; & E\left(\hat{\beta}_0\right) = \frac{1}{4}\left(E\left(w_1\right) + E\left(w_2\right) + \frac{1}{4}\sum_{i=3}^{6} E\left(w_i\right) + \frac{1}{4}\sum_{i=7}^{10} E\left(w_i\right)\right) \\
&= \frac{1}{4}\left(\mu + \mu + \frac{1}{4}\sum_{i=3}^{6} \mu + \frac{1}{4}\sum_{i=7}^{10} \mu\right) \\
&= \frac{1}{4}\left(\mu + \mu + \frac{1}{4} \times 4\mu + \frac{1}{4} \times 4\mu\right) = \frac{1}{4}\left(\mu + \mu + \mu + \mu\right) = \mu
\end{aligned}$$

Therefore $\hat{\beta}_0$ is an unbiased estimator of μ. However, the variances of u_1 and u_2 are equal to σ^2, but the variances of $wbar_1$ and $wbar_2$ are $\frac{\sigma^2}{4}$. Hence the errors are heteroskedastic, so this estimator is not BLUE. Its variance is

$$\begin{aligned}
Var\left(\hat{\beta}_0\right) &= \frac{1}{16}\left(Var\left(w_1\right) + Var\left(w_2\right) + Var\left(wbar_1\right) + Var\left(wbar_2\right)\right) \\
&= \frac{1}{16}\left(\sigma^2 + \sigma^2 + \frac{\sigma^2}{4} + \frac{\sigma^2}{4}\right) = \frac{1}{16}\left(\frac{10\sigma^2}{4}\right) = \frac{5\sigma^2}{32} = 0.15625\sigma^2
\end{aligned}$$

In comparison with the variance of $\frac{1}{10}\left(w_1 + w_2 + 4 \times wbar_1 + 4 \times wbar_2\right)$, which is $\frac{\sigma^2}{10} = 0.1\sigma^2$, this highlights that $\hat{\beta}_0$ is not BLUE.

(c) We need to multiply the model for the last to observation by $\frac{1}{\sqrt{\frac{1}{4}}} = 2$ to equalise the variances of the errors. This transforms the model to:

$$\begin{pmatrix} w_1 \\ w_2 \\ 2wbar_1 \\ 2wbar_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} \beta_0 + \begin{pmatrix} u_1 \\ u_2 \\ 2u_3 \\ 2u_4 \end{pmatrix}$$

in which all errors have the same variance. The OLS estimator of β_0 in this model is:

$$\begin{aligned}
\hat{\beta}_0 &= \left(X'X\right)^{-1} X'Y \\
&= \left[\begin{pmatrix} 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix}\right]^{-1} \begin{pmatrix} 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ 2wbar_1 \\ 2wbar_2 \end{pmatrix} \\
&= \frac{1}{10}\left(w_1 + w_2 + 4 \times wbar_1 + 4 \times wbar_2\right)
\end{aligned}$$

Voilà! WLS gives us the correct answer!