Statistics B2: Econometrics
Solution to Supplementary Econometrics exam
2016-04-20

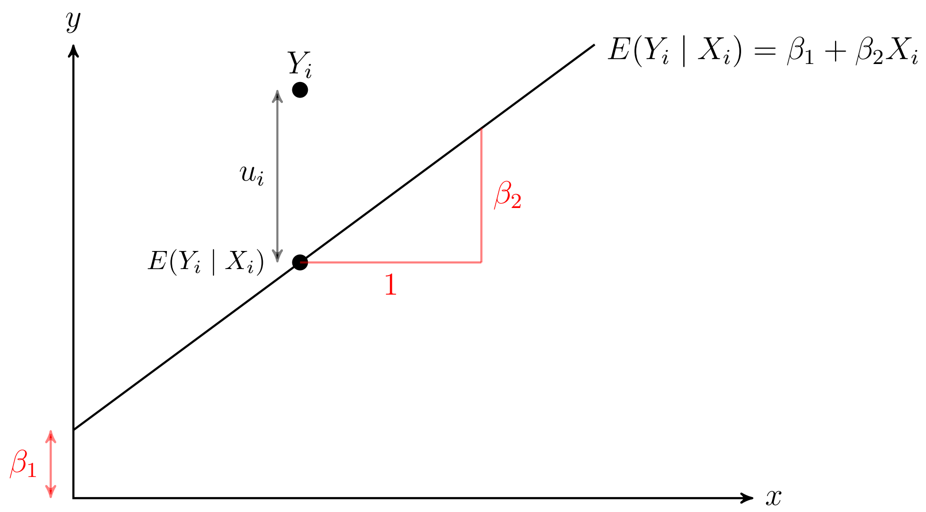
Department of Statistics, Uppsala University

Spring 2016

Task 1

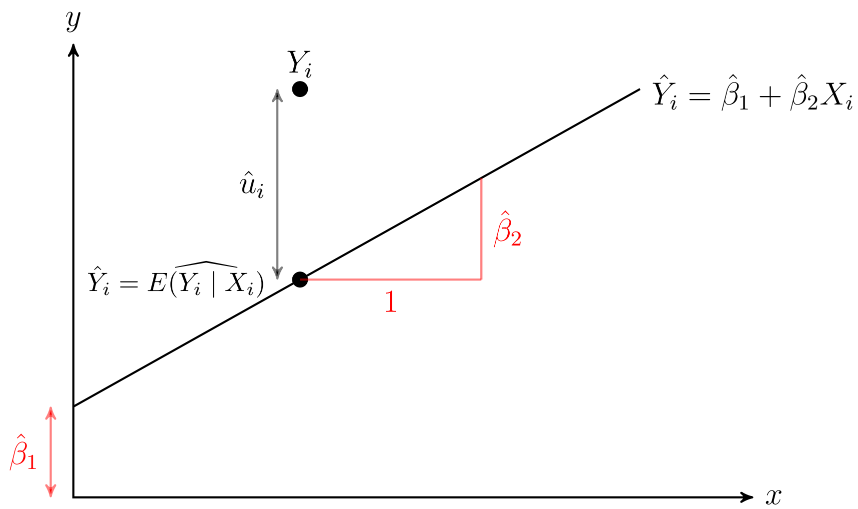
A)

PRF:



B)

The corresponding figure for the SRF is:



C)

The assumption of $cov(u_i, u_j) = 0,\ j \neq i$ means that in time series data the model should capture all temporal dependence in the data. If the model is purely cross sectional, there should be no temporal dependence in the data.

D)

The assumption of $cov(X_i, u_i) = 0$ can also be called *correctly specified model*. If some regressor is left out which should in fact be in the model, and that regressor is correlated with another regressor that is included in the model, the assumption will be violated.

Task 2

A)

In Model (1), β_2 is the expected change in salary given a one unit increase in age, holding other variables fixed.

B)

In Model (1), $\hat{\beta}_2$ is the estimated expected change in salary given a one unit increase in tenure, holding other variables fixed. From Figure 2.1 the estimated coefficient is 33.42, so the estimated expected change in salary given a unit increase in tenure is 33.42, holding other variables fixed.

C)

Two-sided t-test

1. Hypothesis

$$H_0 : \beta_4 = 0$$
$$H_1 : \beta_4 \neq 0$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$\hat{\beta}_4$$

4. Assumptions

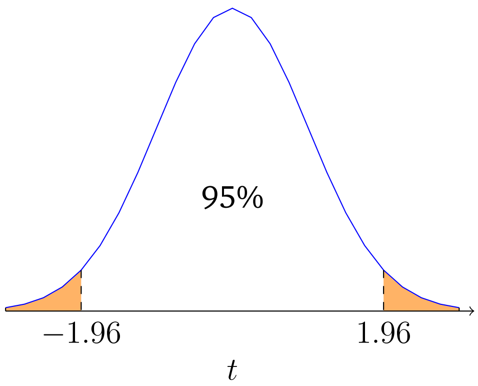
$u_i \sim NID(0, \sigma^2)$ or n large and no heteroscedasticity
$Cov(u_i, X_i) = 0$
No exact multicollinearity.

5. Test statistic

$$t_{obs} = \frac{\hat{\beta}_4 - \beta_4^{H_0}}{\hat{\sigma}_{\hat{\beta}_4}} \sim t_{n-k}$$

6. Rejection rule and figure

$$\text{Reject if:} \quad t_{obs} > t_{443,0.025} \approx 1.96$$
$$\text{Or:} \quad t_{obs} < -t_{443,0.025} \approx -1.96$$



7. Calculations and results

$$t_{obs} = \frac{\hat{\beta}_4 - \beta_4^{H_0}}{\hat{\sigma}_{\hat{\beta}_4}} = \frac{0.0399 - 0}{0.0046} = 8.63$$
$$|t_{obs}| > |t_{crit}| = 1.96$$

$\Rightarrow H_0$ is not rejected.

8. Conclusion

We reject the null hypothesis at the 5% significance level. We do find evidence that sales has an effect on salary.

D)

One-sided t-test

1. Hypothesis

$$H_0 : \beta_4 \leq 0$$
$$H_1 : \beta_4 > 0$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$\hat{\beta}_4$$

4. Assumptions

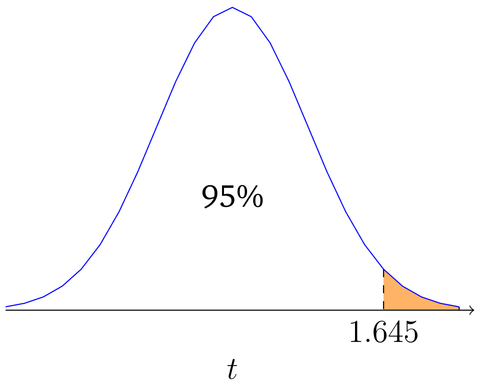
$u_i \sim NID(0, \sigma^2)$ or n large and no heteroscedasticity
$Cov(u_i, X_i) = 0$
No exact multicollinearity.

5. Test statistic

$$t_{obs} = \frac{\hat{\beta}_4 - \beta_4^{H_0}}{\hat{\sigma}_{\hat{\beta}_4}} \sim t_{n-k}$$

6. Rejection rule and figure

$$\text{Reject if:} \quad t_{obs} > t_{443,0.05} \approx 1.645$$



7. Calculations and results

$$t_{obs} = \frac{\hat{\beta}_4 - \beta_4^{H_0}}{\hat{\sigma}_{\hat{\beta}_4}} = \frac{0.4127 - 0}{0.0485} = 8.5171$$
$$t_{obs} > t_{crit} = 1.645$$

$\Rightarrow H_0$ is rejected.

8. Conclusion

We reject the null hypothesis at the 5 % significance level. We reject the hypothesis that profits has no effect on salary in favour of the alternative that it has a positive effect.

E)

Model (1) has an R^2 equal to 0.169, while Model (2) has an R^2 of 0.172. We can conclude that Model (2) has higher explanatory power. Model (2) explains 17.2 percent of the variation in salary while model (1) explains 16.9 percent of the variation.

F)

Because models (2) and (3) have a different number of explanatory variables the adjusted R^2 is appropriate for a comparison. This measure penalizes models with a greater amount of regressors. Model (2) has an adjusted R^2 of 0.166 while model (3) has an adjusted R^2 of 0.135. Model (2) explains 16.6 percent of the variation in salary when taking the number of regressors into account.

G)

Derivation of $(1-\alpha)100\%$ confidence interval:
We know that if $u_i \sim NID(0, \sigma^2)$ then

$$\frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}} \sim t_{n-k}.$$

We have that

$$P\big(t_{(n-k),1-\alpha/2} < \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}} < t_{(n-k),\alpha/2}\big) = 1 - \alpha,$$

By symmetry of t-distribution, $t_{(n-k),1-\alpha/2} = -t_{(n-k),\alpha/2}$ and without changing the probability we can operate as follows

$$= P\big(- t_{(n-k),\alpha/2} < \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}} < t_{(n-k),\alpha/2}\big) = P\big(- t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}} < \hat{\beta} - \beta < t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}}\big)$$

$$= P\big(t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}} > \beta - \hat{\beta} > -t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}}\big) = P\big(\hat{\beta} + t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}} > \beta > \hat{\beta} - t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}}\big)$$

$$= P\big(\hat{\beta} - t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}} < \beta < \hat{\beta} + t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}}\big) = 1 - \alpha.$$

Therefore a $(1-\alpha)100\%$C.I. is given by

$$\hat{\beta} \pm t_{(n-k),\alpha/2}\hat{\sigma}_{\hat{\beta}}.$$

H)

A 95 % C.I is given by

$$35.51 \pm 1.96 \cdot 9.89$$
$$= [16.13 \;\;:\;\; 54.89].$$

In repeated sampling we expect 95 per cent of intervals calculated like this to cover the true value of β_3.

Task 3

A)

F-test of model restrictions

1. Hypothesis

$$H_0: \quad \beta_2 = \beta_4 = 0$$
$$H_1: \quad \text{At least one } \beta_j \neq 0, j = 2,\ 4$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$\hat{\beta}_2, \hat{\beta}_4$$

R^2_{UR} Unrestricted model (4)

R^2_R Restricted model (3)

4. Assumptions

$u_i \sim NID(0, \sigma^2)$or n large and no heteroscedasticity
$Cov(u_i, X_i) = 0$
No exact multicollinearity.

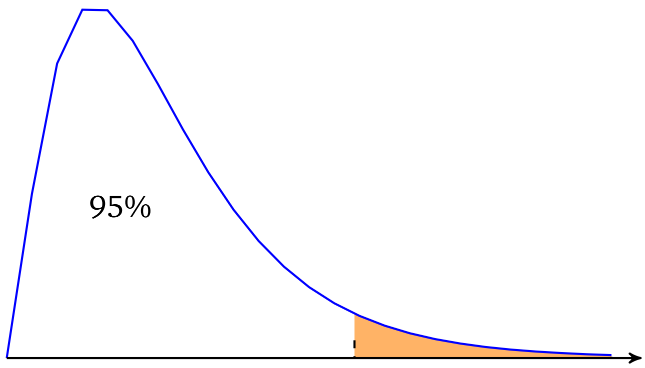
5. Test statistic

$$F_{obs} = \frac{(R^2_{UR} - R^2_R)/r}{(1 - R^2_{UR})/(n-k)} \sim F_{r,(n-k)}$$

r : Restrictions
k : Parameters in UR model

6. Rejection rule and figure

$$\text{Reject if:} \quad F_{obs} > F_{2,443,0.05} \approx F_{2,\infty,0.05} \approx 3.0$$



7. Calculations and results

$$F_{obs} = \frac{(0.192 - 0.137)/2}{(1 - 0.192)/(447 - 4)} = 15.08 > F_{crit} = 3.0$$

8. Conclusion

We reject the null hypothesis at the 5% significance level. That is, we reject the hypothesis that the parameters corresponding to tenure and sales are simultaneously zero.

B)

F-test of model

1. Hypothesis

$$H_0: \quad \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_1: \quad \text{At least one } \beta_j \neq 0, j = 2,\ 3,\ 4$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$R^2_{UR} \quad \text{Unrestricted model}$$
$$R^2_{R} \quad \text{Restricted model}$$

Note that the restricted model is a constant, which does not have any variance, so it can't explain any of the variation in $salary$*, hence, the* $R^2_R = 0$.

4. Assumptions

$u_i \sim NID(0, \sigma^2)$ or n large and no heteroscedasticity
$Cov(u_i, X_i) = 0$
No exact multicollinearity.

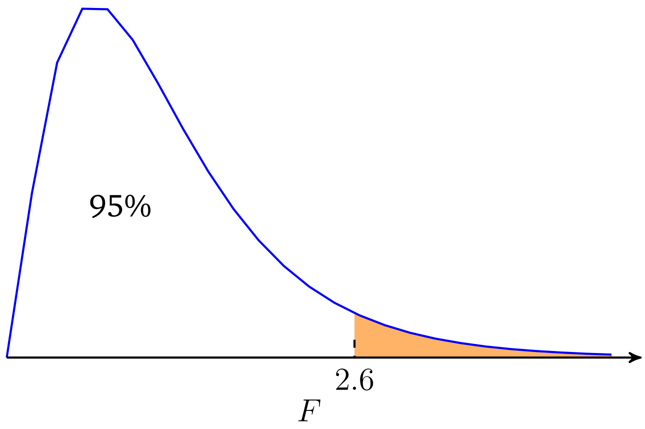
5. Test statistic

$$F_{obs} = \frac{(R^2_{UR} - R^2_R)/r}{(1 - R^2_{UR})/(n-k)} \sim F_{r,(n-k)}$$

r : Restrictions
k : Parameters in UR model

6. Rejection rule and figure

Reject if: $F_{obs} > F_{3,443,0.05} \approx F_{3,\infty,0.05} = 2.6$



7. Calculations and results

$$F_{obs} = \frac{(0.172 - 0)/3}{(1 - 0.172)/(447 - 4)} = 30.69 > F_{crit} = 2.6$$

8. Conclusion

We reject the null hypothesis at the 5% significance level. That is, we reject the hypothesis that the model does not explain any of the variation in salary.

C)

Perfect/exact multicollinearity occurs when two or more variables are perfect linear combination of each other. Consider a model for wage where both age and date of birth are included. We can convert between the two using age = todays date − date of birth, and thus the two variables carry exactly the same information. This will cause exact multicollinearity and it is not possible to estimate the model.

D)

When we have a high degree of multicollinearity two or more regressors are highly correlated and it is not possible to distinguish from which of the regressors the explanatory power comes.

E)

Before we estimate a model it always a good idea to calculate a correlation matrix where high correlations can be identified and appropriate measures taken.

F)

If we estimate a model with a high degree of multicollinearity we can expect to see non-significant parameter estimates but a high R^2. For model (4) there is a potential problem considering that profits and sales are likely correlated.

Task 4

A)

Derivation of OLS estimator (without intercept)

$$\text{arg min}\left(\sum_{i=1}^{n} u_i^2\right) = \text{arg min}\left(\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\right) = \text{arg min}\left(\sum_{i=1}^{n}(Y_i - \hat{\beta}X_i)^2\right)$$

$$\frac{\partial \sum_{i=1}^{n}(Y_i - \hat{\beta}X_i)^2}{\partial \hat{\beta}} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}X_i)(-X_i) = 0 \quad F.O.C$$

$$\Rightarrow \sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} \hat{\beta}X_i^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} X_iY_i = \hat{\beta}\sum_{i=1}^{n} X_i^2$$

$$\Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} X_i^2}$$

B)

Derivation of the variance of the OLS-estimator

$$V\left(\hat{\beta}\right) = V\left(\frac{\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} X_i^2}\right) = E\left[\left(\hat{\beta} - E(\hat{\beta})\right)^2\right] = E\left[\left(\frac{\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} X_i^2} - \beta\right)^2\right]$$

$$= E\left[\left(\frac{\sum_{i=1}^{n} X_i(\beta X_i + u_i)}{\sum_{i=1}^{n} X_i^2} - \beta\right)^2\right]) = E\left[\left(\frac{\beta\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=1}^{n} X_iu_i}{\sum_{i=1}^{n} X_i^2} - \beta\right)^2\right]$$

$$= E\left[\left(\beta + \frac{\sum_{i=1}^{n} X_iu_i}{\sum_{i=1}^{n} X_i^2} - \beta\right)^2\right] = \left(\frac{1}{\sum_{i=1}^{n} X_i^2}\right)^2 E\left[\left(\sum_{i=1}^{n} X_iu_i\right)^2\right]$$

$$= \left(\frac{1}{\sum_{i=1}^{n} X_i^2}\right)^2 E\left[(u_1X_1 + u_2X_2 + \ldots + u_nX_n) \times (u_1X_1 + u_2X_2 + \ldots + u_nX_n)\right]$$

$$= \frac{1}{\left(\sum_{i=1}^{n}(X_i^2)\right)^2} E\left[\left(\sum_{i=1}^{n} u_i^2 X_i^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} u_i u_j X_i X_j\right)\right]$$

$$= \frac{1}{\left(\sum_{i=1}^{n} X_i^2\right)^2} \left(\sum_{i=1}^{n} E(u_i^2) X_i^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} E(u_i u_j) X_i X_j\right)$$

$$= \frac{\sum_{i=1}^{n} E[u_i^2] X_i^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2} + \frac{2\sum_{i=2}^{n}\sum_{j=1}^{i-1} Cov(u_i, u_j) X_i X_j}{\left(\sum_{i=1}^{n} X_i^2\right)^2}$$

Assuming that $Cov(u_i, u_j) = 0$ for $i \neq j$, and $E[u_i^2] = V(u_i) = \sigma_i^2$

$$V(\hat{\beta}) = \frac{\sum_{i=1}^{n} \sigma_i^2 X_i}{(\sum_{i=1}^{n} X_i^2)^2}.$$