Statistics B2: Econometrics
Solution to Econometrics exam
2016-03-23

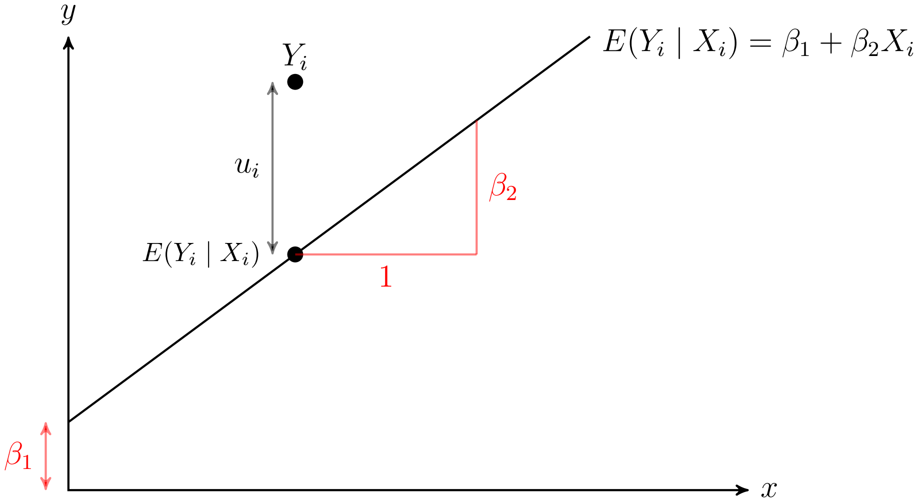
Department of Statistics, Uppsala University

Spring 2016

Task 1

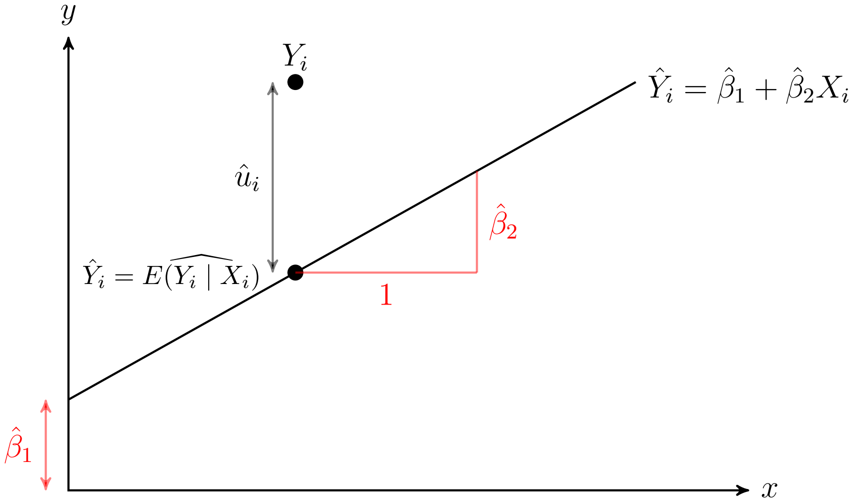
A)

PRF:



B)

The corresponding figure for the SRF is:



C)

A log-lin model is given by

$$\ln y_i = \beta_1 + \beta_2 x_i + e_i,$$

where $100\beta_2$ is the expected percentage change in Y given a one unit absolute increase in X. The model is linear in the parameters and nonlinear in the variables.

D)

A lin-log model is given by

$$y_i = \beta_1 + \beta_2 \ln x_i + e_i$$

where $\frac{\beta_2}{100}$ is the expected absolute change in Y given a one percentage increase in X. The model is linear in the parameters and nonlinear in the variables.

Task 2

A)

In Model (1), $\hat{\beta}_2$ is the expected change in miles per gallon (MPG) given a one unit increase in age, holding other variables fixed.

B)

In Model (1), $R^2 = 0.809$ is the proportion of variation in MPG explained by the regressors.

C)

Two-sided t-test

1. Hypothesis

$$H_0 : \beta_5 = 0$$
$$H_1 : \beta_5 \neq 0$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$\hat{\beta}_5$$

4. Assumptions

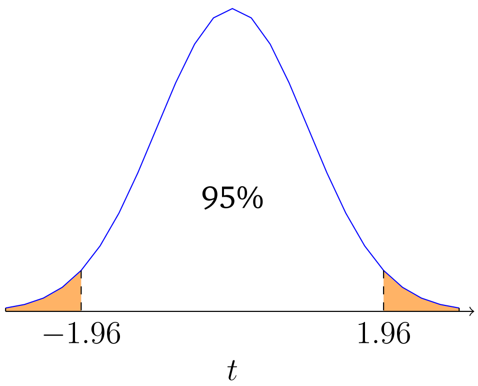
$u_i \sim NID(0, \sigma^2)$ or n large and no heteroscedasticity
$Cov(u_i, X_i) = 0$
No exact multicollinearity.

5. Test statistic

$$t_{obs} = \frac{\hat{\beta}_5 - \beta_5^{H_0}}{\hat{\sigma}_{\hat{\beta}_5}} \sim t_{n-k}$$

6. Rejection rule and figure

$$\text{Reject if:} \quad t_{obs} > t_{386,0.025} \approx 1.96$$
$$\text{Or:} \quad t_{obs} < -t_{386,0.025} \approx -1.96$$



7. Calculations and results

$$t_{obs} = \frac{\hat{\beta}_5 - \beta_5^{H_0}}{\hat{\sigma}_{\hat{\beta}_5}} = \frac{0.0028 - 0}{0.1019} = 0.8863$$
$$|t_{obs}| < |t_{crit}| = 1.96$$

$\Rightarrow H_0$ is not rejected.

8. Conclusion

We cannot reject the null hypothesis at the 5% significance level. We do not find evidence that displacement has an effect on MPG.

D)

Considering the correlation matrix in Figure 2.4, the result in previous subtask is not very surprising. While displacement is negatively correlated with MPG, which is what we expect from (engineering) theory, it is also highly correlated with some of the other regressors (weight, horsepower, cylinders). When regressors are highly correlated we have multicollinearity which inflates the variances of the parameter estimates (VIF). This is probably the reason we could not reject the null in the previous subtask. More intuitively, what happens is that we cannot distinguish between, for example, the effect of displacement versus the effect of weight.

E)

Model (1) has an adjusted R^2 equal to 0.806, while Model (3) has an adjusted R^2 of 0.807. We can conclude that Model (3) has higher explanatory power when taking the number of variables into account.

F)

One-sided t-test

1. Hypothesis

$$H_0 : \gamma_2 \geq 0$$
$$H_1 : \gamma_2 < 0$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$\hat{\gamma}_2$$

4. Assumptions

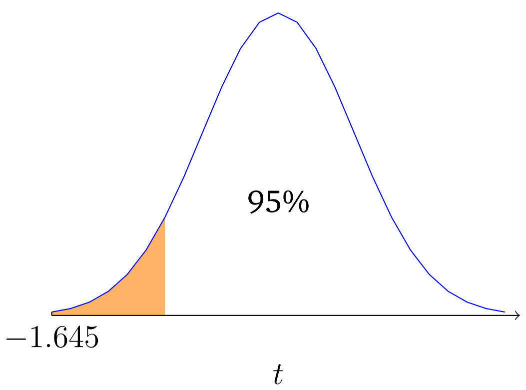
$u_i \sim NID(0, \sigma^2)$ or n large and no heteroscedasticity
$Cov(u_i, X_i) = 0$
No exact multicollinearity.

5. Test statistic

$$t_{obs} = \frac{\hat{\gamma}_2 - \gamma_2^{H_0}}{\hat{\sigma}_{\hat{\gamma}_2}} \sim t_{n-k}$$

6. Rejection rule and figure

$$\text{Reject if:} \quad t_{obs} < -t_{389,0.05} \approx -1.645$$



7. Calculations and results

$$t_{obs} = \frac{\hat{\gamma}_2 - \gamma_2^{H_0}}{\hat{\sigma}_{\hat{\gamma}_2}} = \frac{-0.757 - 0}{0.049} = -15.308$$
$$t_{obs} < t_{crit} = -1.645$$

$\Rightarrow H_0$ is rejected.

8. Conclusion

We reject the null hypothesis at the 5 % significance level. We reject the hypothesis that age has no effect on MPG in favour of the alternative that it has a negative effect.

G)

Based on the scatterplot in Figure 2.6 we clearly have some problems with heteroscedasticity. The consequences for OLS estimation is that it is no longer BLUE, although it is still unbiased, it is not efficient. In this case we should use HAC-estimators for the standard errors.

Task 3

A)

F-test of model restrictions

1. Hypothesis

$$\begin{aligned} H_0: &\quad \beta_4 = \beta_5 = \beta_6 = 0 \\ H_1: &\quad \text{At least one } \beta_j \neq 0, j = 4, 5, 6 \end{aligned}$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$\hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$$

R^2_{UR} Unrestricted model (1)

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.

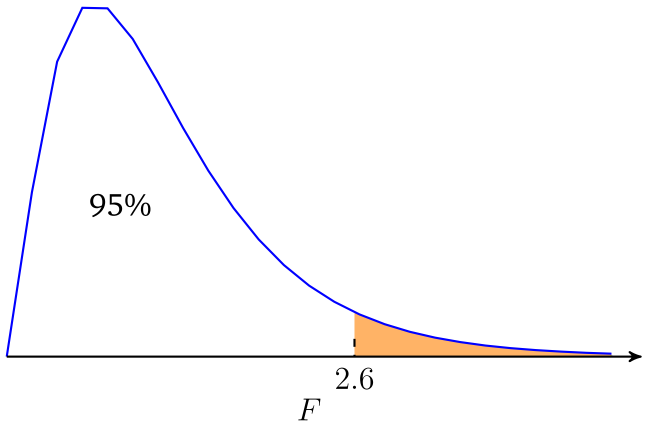
5. Test statistic

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

$r:$ Restrictions
$k:$ Parameters in UR model

6. Rejection rule and figure

$$\text{Reject if:} \quad F_{obs} > F_{3,386,0.05} \approx F_{3,\infty,0.05} = 2.6$$



7. Calculations and results

$$F_{obs} = \frac{(0.809 - 0.808)/3}{(1 - 0.809)/(392 - 6)} = 0.674 < F_{crit} = 2.6$$

8. Conclusion

We cannot reject the null hypothesis at the 5% significance level. That is, we cannot reject the hypothesis that the parameters corresponding to acceleration, displacement and horsepower are simultaneously zero.

B)

F-test of model

1. Hypothesis

$$H_0: \quad \gamma_2 = \gamma_3 = 0$$
$$H_1: \quad \text{At least one } \gamma_j \neq 0, j = 2, 3$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

R^2_{UR} Unrestricted model

R^2_{R} Restricted model

Note that the restricted model is a constant, γ_1, which does not have any variance, so it can't explain any of the variation in MPG, 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.

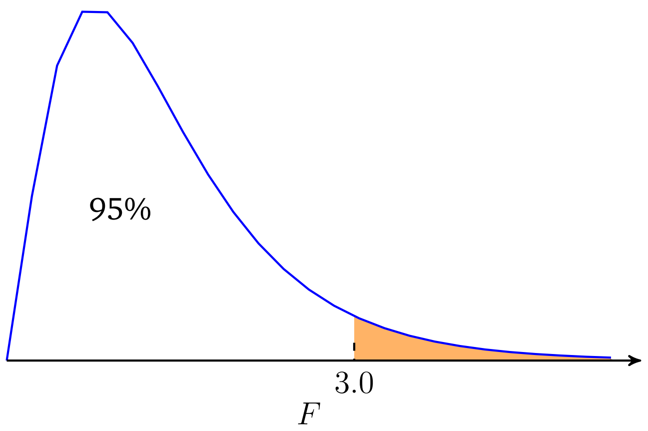
5. Test statistic

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

r : Restrictions
k : Parameters in UR model

6. Rejection rule and figure

Reject if: $F_{obs} > F_{2,389,0.05} \approx F_{2,\infty,0.05} = 3.0$



7. Calculations and results

$$F_{obs} = \frac{(0.808 - 0)/2}{(1 - 0.808)/(392 - 3)} = 819.473 > F_{crit} = 3.0$$

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 MPG.

C)

In order to include the categorical variable "wheel drive", we can code it as dummy variables. Important to remember is to either drop one of the categories (and use it as a reference) or drop the intercept. The model could look like this:

$$MPG = \gamma_1 + \gamma_2 Age + \gamma_3 Weight + \gamma_4 RearWD + \gamma_5 FrontWD + e.$$

D)

Jarque-Bera Test

1. Hypothesis

$$\begin{aligned} H_0 &: \text{ The error term is normally distributed} \\ H_1 &: \text{ The error term is not normally distributed} \end{aligned}$$

2. Significance level

$$\alpha = 0.05$$

3. Estimator(s)/Statistics

$$S = \widehat{Skewness} = \frac{\frac{1}{n}\sum(Y_i - \bar{Y})^3}{\hat{\sigma}_Y^3}$$
$$K = \widehat{Kurtosis} = \frac{\frac{1}{n}\sum(Y_i - \bar{Y})^4}{\hat{\sigma}_Y^4}$$

4. Assumptions

n large

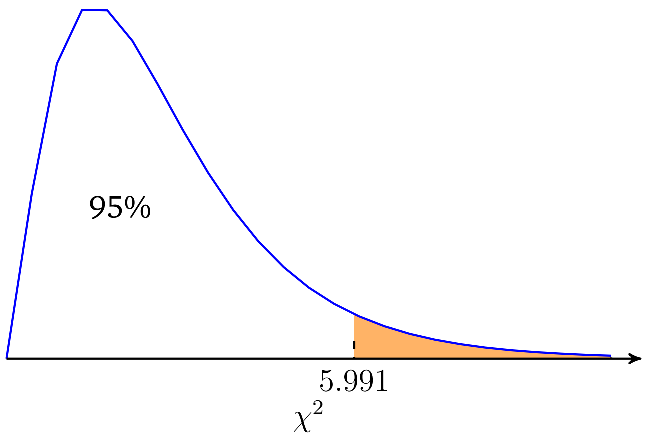
5. Test statistic

$$JB_{obs} = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

which under the null (asymptotically) follows a χ^2-distribution with 2 degrees of freedom.

6. Rejection rule and figure

Reject if: $JB_{obs} > \chi^2_{0.05}(2) = 5.991$



7. Calculations and results

$$JB_{obs} = 392\left[\frac{0.575^2}{6} + \frac{(4.101-3)^2}{24}\right] = 41.381 > \chi^2_{crit} = 5.991$$

8. Conclusion

We reject the null hypothesis at the 5% significance level. That is, we reject the hypothesis that the residuals are normally distributed in favour of the alternative.

E)

The important thing for inference is that estimators $\hat{\beta}_j$ are normally distributed. In the case where residuals are normal themselves, normality of OLS-estimators follows because they are linear estimators and linear combinations of normally distributed random variables are also normal. The second case is where residuals are not normal, and we need to rely on CLT for normality of estimators and test statistics (which are inverted to obtain confidence intervals).

F)

The problem with doing such a point prediction is that it is outside of our data range and we don't know how the model performs there, even if it performs well inside the data range. Consider making a point prediction for a car that is brand new and weighs *nothing*, it is fully possible to make such a point prediction using our model, but the model is not estimated on weightless cars and thus not guaranteed to give a good or even reasonable point prediction for such a car. Similarly, we don't know anything about the gas consumption of cars that weigh 500 pounds because we haven't observed such a car in our sample.

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

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

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

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

$$= \left(\frac{1}{\sum_{i=1}^{n} X_i^2}\right)^2 E\left[(u_1 X_1 + u_2 X_2 + \ldots + u_n X_n) \times (u_1 X_1 + u_2 X_2 + \ldots + u_n X_n)\right]$$

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

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

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

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

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

C)

1.

This is typically the case when a relevant variable has been omitted.

2.

In this case the OLS estimator is not BLUE, in fact it is biased.

3.

Try to include the omitted variable. If we do that it is no longer contained in the error term and the error term in turn no longer covaries with X_2.

4.

In this case we can use X_4 and X_5 to run the regression

$$X_2 = \alpha_1 + \alpha_2 X_4 + \alpha_3 X_5 + e$$

where we obtain fitted values $\hat{X}_2$ to include in the first regression instead of X_2.

5.

This is not a problem. We are not interested in the estimates of α_j, just the fitted values $\hat{X}_2$. As long as X_4 and X_5 explain variation in X_2 it is fine.