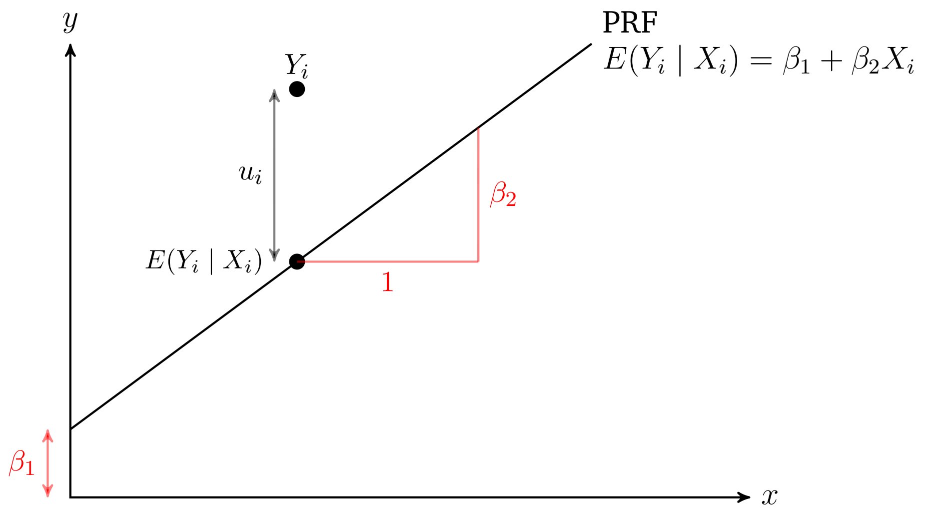
# Statistics B2: Econometrics
# Solution to Supplementary Econometrics exam 2017-04-19

Department of Statistics, Uppsala University

Spring 2017

## Task 1

**A)**

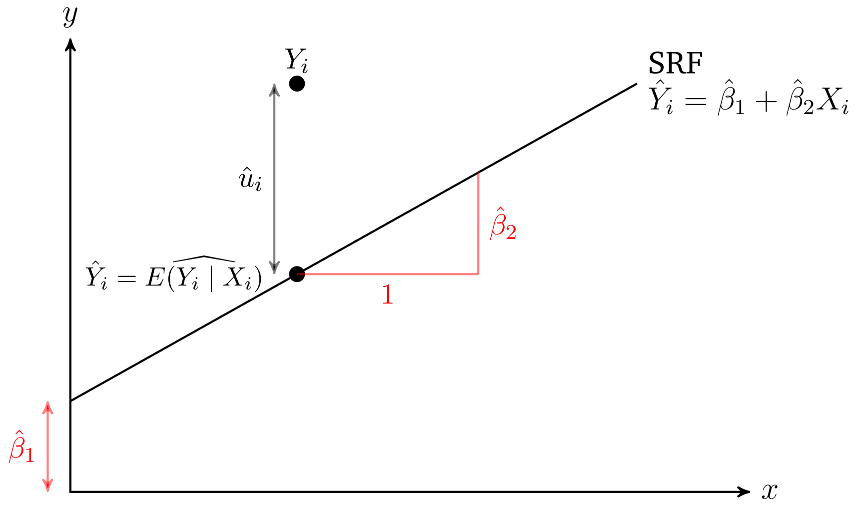



**B)**

The *Sample* Regression Function (SRF) is given by

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i.$$

The corresponding figure is:




**C)**

Given a unit change in $X$ we expect a $100\beta_2$ per cent change in Y
OR:
An absolute increase in $X$ by one unit increases the conditional expectation of Y by $100\beta_2$ per cent

**D)**

Given a one per cent increase in $X$ we expect a $\frac{\beta_2}{100}$ unit change in Y.
OR:
A one per cent increase in $X$ gives a $\frac{\beta_2}{100}$ unit change in the conditional expectation of Y.

**E)**

$\beta_2$ is the expected percentage change in Y given a relative (one per cent) change in X.
OR:
$\beta_2$ is the percentage change in conditional expectation of Y given a relative (one per cent) change in X.

## Task 2

**A)**

$\beta_2$ is the expected increase in years of education given an increase in the education of ones father by one year, holding education of ones mother fixed.
OR:
$\beta_2$ is the increase in conditional expectation of Y (years of education) given an absolute increase in Feduc (fathers years of education), holding Meduc (mothers years of education) fixed.

**B)**

$\hat{\beta}_5$ is the estimated expected increase in years of education given an increase in IQ by 1, holding education of parents and gender fixed. In this case we estimate that the average effect of an additional point in IQ is 0.063 years of education, holding the other variables fixed.

**C)**

The coefficient of determination of model two is 0.22. This means that the education of ones father and mother, together with gender, explains 22 per cent of the variation in years of education.

**D)**

### Two-sided t-test

#### 1. Hypothesis

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

#### 2. Significance level

$$\alpha = 0.05$$

#### 3. Estimator

$$\hat{\beta}_2$$

#### 4. Assumptions

$$u_i \sim NID(0, \sigma^2)$$
$$Cov(u_i, X_i) = 0$$

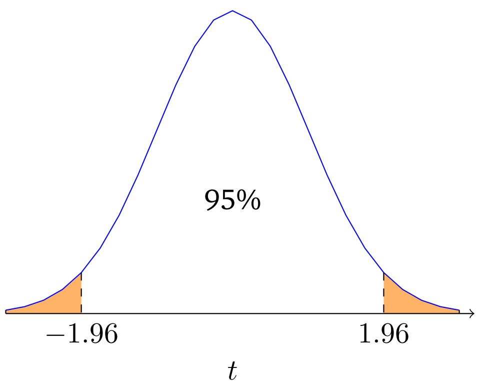
#### 5. Test statistic

$$t_{obs} = \frac{\hat{\beta}_2 - \beta_{H_0}}{\hat{\sigma}_{\hat{\beta}_3}} \sim t_{n-k}$$
$$t_{obs} = \frac{\hat{\beta}_2 - 0}{\hat{\sigma}_{\hat{\beta}_2}} \sim t_{719}$$

#### 6. Rejection rule and figure

Reject if: $t_{obs} > t_{719,0.025} \approx t_{\infty,0.025} = 1.96$

Or: $t_{obs} < -t_{719,0.025} \approx -t_{\infty,0.025} = -1.96$



**7. Calculations and results**

$$t_{obs} = \frac{0.215635 - 0}{0.027518} = 7.836229 > t_{crit}$$
$$\Rightarrow H_0 \text{rejected.}$$

**8. Conclusion**

We reject the null hypothesis at the 5% significance level. That is, reject the hypothesis that fathers education explains nothing in the variation of education, in favour of the alternative.

**E)**

## One-sided t-test

**1. Hypothesis**

$$H_0 : \beta_4 \leq 0.5$$
$$H_1 : \beta_4 < 0.5$$

**2. Significance level**

$$\alpha = 0.05$$

**3. Estimator**

$$\hat{\beta}_4$$

**4. Assumptions**

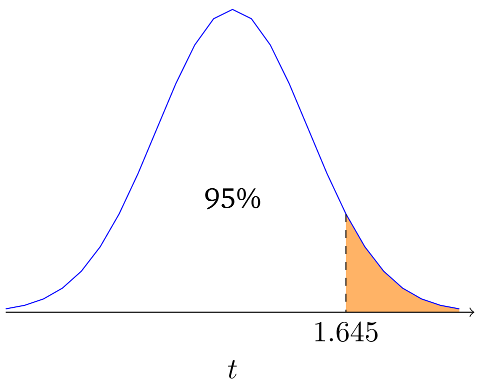
$$u_i \sim NID(0, \sigma^2)$$
$$Cov(u_i, X_i) = 0$$

**5. Test statistic**

$$t_{obs} = \frac{\hat{\beta}_4 - \beta_4^{H_0}}{\hat{\sigma}_{\hat{\beta}_4}} \sim t_{n-k}$$
$$t_{obs} = \frac{\hat{\beta}_4 - 0.5}{\hat{\sigma}_{\hat{\beta}_3}} \sim t_{722-4}$$

**6. Rejection rule and figure**

$$\text{Reject if:} \quad t_{obs} > t_{718,0.05} \approx t_{\infty,0.05} = 1.645$$




**7. Calculations and results**

$$t_{obs} = \frac{0.536655 - 0.5}{0.151911} = 0.241 < t_{crit} = 1.645 \Rightarrow H_0 \text{ not rejected.}$$

**8. Conclusion**

We cannot reject the null hypothesis, that males on average have less than or equal to half a year longer education, at the 5% significance level.

**F)**

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

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

**G)**

A 95 % C.I is given by

$$0.215635 \pm 1.96 \cdot 0.027518$$
$$= [0.1617 \ : \ 0.2696].$$

In repeated sampling we expect 95 per cent of intervals calculated like this to cover the true value of $\beta_2$.

# Task 3

**A)**

## F-test of model

### 1. Hypothesis

$$\begin{aligned} H_0 &: \quad \beta_2 = \beta_3 = 0 \\ H_1 &: \quad \text{At least one } \beta_j \neq 0, j = 2, 3 \end{aligned}$$

### 2. Significance level

$$\alpha = 0.05$$

### 3. Estimator(s) & Statistics

$$\begin{array}{ll} & \hat{\beta}_2,\ \hat{\beta}_3 \\ R^2_{UR} & \text{Unrestricted model} \\ R^2_R(=0) & \text{Restricted model} \end{array}$$

### 4. Assumptions

$$u_i \overset{iid}{\sim} N(0, \sigma^2)$$
$$Cov(u_i, X_i) = 0$$
No perfect multicollinearity

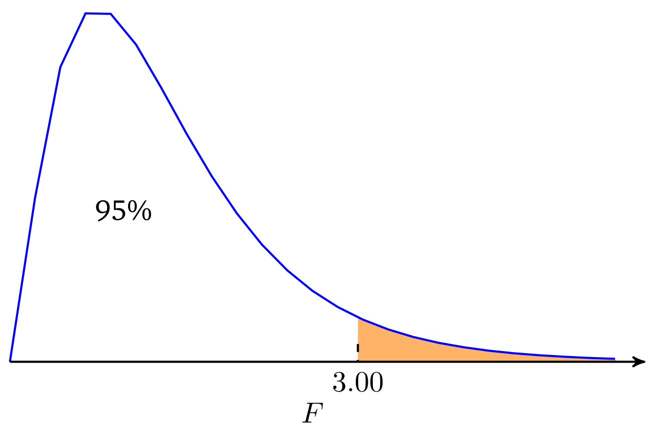
### 5. Test statistic

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

$r$ : No. of restrictions
$k$ : No. of parameters in UR model

### 6. Rejection rule and figure

Reject if: $F_{obs} > F_{2,722-3,0.05} \approx F_{2,\infty,0.05} = 3.00$



**7. Calculations and results**

$$F_{obs} = \frac{(0.206017 - 0)/2}{(1 - 0.206017)/(722 - 3)} = 93.28 > F_{crit} = 3.00$$
$$\Rightarrow H_0 \text{rejected.}$$

**8. Conclusion**

We reject the null hypothesis at the 5% significance level. That is, we cannot find evidence to reject the hypothesis that the model explains nothing of the variation in the dependent variable.

**B)**

If the sample only contains individuals that are males or females then being male is perfectly predicted by *not female*. This is a classic case of the dummy variable trap causing perfect multicollinearity.

**C)**

If we want to keep both variables in the model we can specify it without an intercept,

$$Educ_i = \beta_2 Feduc_i + \beta_3 Meduc_i + \beta_4 Male_i + \beta_5 Female_i + u_i$$

which can be estimated.

**D)**

Because $R^2$ is a non-decreasing function of the number of explanatory variables. A model with additional parameters will therefore always have higher $R^2$ than a more simple model.

**E)**

Instead, he should use adjusted $R^2$, which punishes models that have more regressors. $\bar{R}^2$ for model 5 is $0.217489$, meaning that the models explains 21.75 per cent of the variation in education, when taking the number of regressors into account.

**F)**

## F-test: Restrictions of parameters

**1. Hypothesis**

$$\begin{aligned} H_0 &: \quad \beta_5 = \beta_6 = \beta_7 = 0 \\ H_1 &: \quad \text{At least one } \beta_j \neq 0, j = 5, 6, 7 \end{aligned}$$

**2. Significance level**

$$\alpha = 0.05$$

**3. Estimator(s) & Statistics**

$$\hat{\beta}_5,\ \hat{\beta}_6,\ \hat{\beta}_7$$

$R^2_{UR}$ Unrestricted model

$R^2_R$ Restricted model

**4. Assumptions**

$$u_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$Cov(u_i, X_i) = 0$$

No perfect multicollinearity

**5. Test statistic**

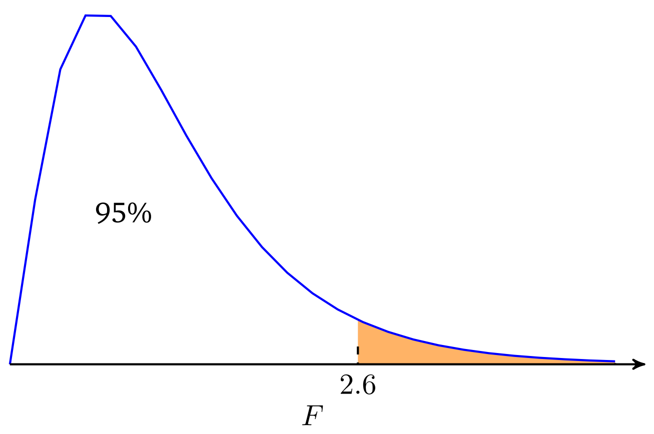
$$F_{obs} = \frac{(R^2_{UR} - R^2_R)/r}{(1 - R^2_{UR})/(n-k)} \sim F_{r,(n-k)}$$

$r$ : No. of restrictions

$k$ : No. of parameters in UR model

**6. Rejection rule and figure**

Reject if: $F_{obs} > F_{3,715,0.05} \approx F_{3,\infty,0.05} = 2.6$




**7. Calculations and results**

$$F_{obs} = \frac{(0.224 - 0.219582)/3}{(1 - 0.224)/(722 - 7)} = 1.357 < F_{crit} = 2.6 \Rightarrow H_0 \text{not rejected.}$$

**8. Conclusion**

We cannot find evidence to reject the null that urban, married and age explain nothing of the variation in education on the 5 per cent significance level.

## Task 4

### A)

Consider the model

$$Y_i = \beta X_i + u_i,$$

$i = 1, \dots, N$, and $X_i$ is considered fixed in repeated sampling, and $u_i \sim NID(0, \sigma^2)$

$$\min_{\hat{\beta}} \left( \sum_{i=1}^{n} \hat{u}_i^2 \right) = \min_{\hat{\beta}} \left( \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \right) = \min_{\hat{\beta}} \left( \sum_{i=1}^{n} (Y_i - (\hat{\beta} X_i))^2 \right)$$

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

Derivative w.r.t $\hat{\beta}$, using that the derivative of a sum is equal to the sum of the derivatives:

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

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

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

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

### B)

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

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

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

$$= \beta + \frac{\sum_{i=1}^{n} E[u_i X_i]}{\sum_{i=1}^{n} X_i^2}$$

Assuming that $E[u_i X_i] = Cov(u_i, X_i) = 0$,

$$\Rightarrow E[\hat{\beta}] = \beta.$$