Review – OLS estimators

Given a random sample: \( y_i = \beta_0 + \beta_1 x_i + u_i \) \( i = 1, 2, \ldots, n \)

The OLS estimators of \( \beta_1 \) and \( \beta_0 \) are:

\[
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

Fitted value:

\( \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \)

Residual:

\( \hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \)

To assess the goodness-of-fit of our model we decomposed the total sum of squares \( SST \):

\[
SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} \hat{u}_i^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = SSR + SSE
\]

Computed the fraction of the total sum of squares (SST) that is explained by the model, denoted this as the R-squared of the regression

\[
R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}
\]

Can show this is the squared correlation between \( y \) and the fitted values
Review - Properties of OLS Estimators

Under assumptions SLR.1 to SLR.4 we have shown that the OLS estimators were **unbiased**: their “average” value over samples was equal to the true parameter values

\[ E(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad E(\hat{\beta}_1) = \beta_1 \]

To assess whether the distribution of the estimators was **concentrated** around the true parameters we added an additional assumption

**Assumption SLR.5 (Homoskedasticity):**

\[ \text{Var}(u|x) = \sigma^2 \]

With this extra assumption we concluded:

\[ \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\text{SST}_x} \quad \text{SST}_x = \sum_{i=1}^{n}(x_i - \bar{x})^2 \]

\[ \text{Var}(\hat{\beta}_0) = \frac{n^{-1} \sigma^2 \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \]

These are **conditional** variances: treat the x’s as known constants

Problem: don’t know \( \sigma^2 \)
Not such a big problem…

Estimate $\sigma^2$ by:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n - 2} = \frac{SSR}{n - 2}$$

It is unbiased for $\sigma^2$

The standard error of the regression is given by:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

Estimated standard error (se) of the parameters (Square root of estimated variances, substituting unknown parameters by estimates):

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\left[ \sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{1/2}}$$

And similarly for $\hat{\beta}_0$
Our lead example:

\[ \text{wage}_i = 245.073 + 52.513 \ \text{educ}_i \quad R^2 = 0.262 \quad n = 11064 \]

\[ \hat{\sigma}^2 = 155366 \quad \hat{\sigma} = 394.164 \]

\[ se(\hat{\beta}_0) = 7.575 \]

\[ se(\hat{\beta}_1) = 0.837 \]

Has pretty much everything we have learned so far…
Multiple Linear Regression Model
Multiple Linear Regression Model

Now suppose:

\[ Earnings = \beta_0 + \beta_1 \text{Education} + \beta_2 \text{Experience} + \beta_3 \text{Tenure} + u \]

Experience - measured as the number of years of labor market experience

Tenure - years with current employer

\( u \) contains other factors affecting earnings, it should be “unrelated” to the regressors

Want to interpret the coefficients as the “ceteribus paribus” effect of their variation on Earnings
Multiple Linear Regression: Assumptions

Assumption MLR.1 (Linearity in parameters)

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + u \]

Assumption MLR.2 (Random Sampling from the population)

We have a random sample:

\[ \{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : i = 1, 2, \ldots, n\} \]

satisfying the equation above

Assumption MLR.3 (No perfect Collinearity)

In the sample, none of the independent variables is a linear combination of the others.

This implies, among other things, that none of the independent variables is constant. Also, none of the independent variables is a multiple of another and none has perfect correlation with a linear combination of the others: this would mean that some variable was redundant \(\rightarrow\) could not “identify” the parameters

Assumption MLR.4 (Zero Conditional Mean)

\[ E(u|x_1, x_2, \ldots, x_k) = 0 \quad \text{this implies} \quad E(u) = 0 \]
OLS Estimators

\[ \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik})^2 \]

Let us minimise the Residuals Sum of Squares (SSR) given above. Take derivatives...

\[ \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} u_i = 0 \]
\[ \sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{i1} u_i = 0 \]
\[ \sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{i2} u_i = 0 \]
\[ \sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{ik} u_i = 0 \]

... We are imposing zero correlation between the regressors and the residuals: in the population, there is also zero correlation between the error term and the regressors...
Some definitions before solving...

Fitted value: \[
\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \ldots + \hat{\beta}_k x_{ik}
\]

Residual: \[
\hat{u}_i = (y_i - \hat{y}_i) = (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik})
\]
Decomposition of the Total Sum of Squares ($SST$)…again…

$$ SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} \hat{u}_i^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = SSR + SSE $$

- So, this decomposition is also valid for the multiple linear regression model
- Can compute the fraction of the total sum of squares ($SST$) that is explained by the model, denote this as the R-squared of the regression

$$ R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = \frac{\left(\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)\left(\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2\right)} $$

- $R^2$ can never decrease when another independent variable is added to a regression, and usually will increase
- Because $R^2$ will usually increase with the number of independent variables, it is not a good measure to compare models
Solving for OLS: Model in Matrix Form

- To solve for those messy first order conditions, write first the model in matrix form.
- The model for the $n$ observations of $y$ and the regressors is:

\[
y = X\beta + u
\]

\[
y = \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}, \quad X = \begin{bmatrix}
1 & x_{11} & x_{12} & \ldots & x_{1k} \\
1 & x_{21} & x_{22} & \ldots & x_{2k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \ldots & x_{nk}
\end{bmatrix}, \quad u = \begin{bmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{bmatrix}, \text{ and } \beta = \begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_k
\end{bmatrix}
\]
First Order Conditions in Matrix Form

\[ \sum_{i=1}^{n} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik} \right) = \sum_{i=1}^{n} \hat{u}_i = 0 \]

\[ \sum_{i=1}^{n} x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{i1} \hat{u}_i = 0 \]

\[ \sum_{i=1}^{n} x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{i2} \hat{u}_i = 0 \]

\[ \ldots \]

\[ \sum_{i=1}^{n} x_{ik} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{ik} \hat{u}_i = 0 \]

\[ X' \hat{u} = 0 \Leftrightarrow X'(y - X\hat{\beta}) = 0 \]

or \[ (X'X)\hat{\beta} = X'y \]

which leads to \[ \hat{\beta} = (X'X)^{-1}X'y \]
OLS Estimator

\[ \hat{\beta} = (X'X)^{-1} X'y \]

where:

\[ y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \]

- Assumption MLR.3 (No perfect Collinearity) implies that \( X \) has full column rank \((=k+1)\)
- This means: No column of \( X \) is a linear combination of the other columns
- This in turn implies that \((X'X)^{-1}\) exists
Properties of OLS: Unbiasedness

Under MLR.1 to MLR.4, the OLS estimator is unbiased for \( \beta \):

\[
E(\hat{\beta}_j) = \beta_j \quad j = 0, 1, \ldots, k
\]

Proof:

Notice that \( y = X\beta + u \)

Then,

\[
\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u)
\]

\[
= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u
\]

So, conditional on \( X \), and using the fact that \( E(u|x_1, x_2, \ldots, x_k) = 0 \) and that the sample is random, we can conclude that:

\[
E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(u|X) = \beta + (X'X)^{-1}X'0 = \beta
\]
Properties of OLS: Variance

Assumption MLR.5 (Homoskedasticity) \[ V(u|x_1, x_2, \ldots, x_k) = \sigma^2 \]

With matrices, \( \text{Var}(AX) = A \text{Var}(X)A' \), where \( X \) is a random variable

\[
\text{Var}(\hat{\beta}|X) = \text{Var}[(X'X)^{-1}X'u|X] = \\
= (X'X)^{-1}X' \text{Var}(u|X)X(X'X)^{-1}
\]

Since \( X'X \) is symmetric (equal to its transpose) \( \rightarrow \) inverse is also symmetric

Random sampling and homoskedasticity imply: \[
\text{Var}(u|X) = \sigma^2 I_n
\]

Therefore: \[
\text{Var}(\hat{\beta}|X) = (X'X)^{-1}X'(\sigma^2 I_n)X(X'X)^{-1} = \\
= \sigma^2(X'X)^{-1}X'X(X'X)^{-1}
\]

\[
\text{Var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}
\]
Estimating the Error Variance

• In practice, the error variance $\sigma^2$ is unknown, must estimate it...

$$\hat{\sigma}^2 = \left( \sum_{i=1}^{n} \hat{u}_{i}^2 \right) / (n - k - 1) \equiv SSR / df$$

$df$ (degrees of freedom) is the (number of observations) – (number of estimated parameters)

• Can show that this is an unbiased estimator of the error variance

• The standard error of the regression is given by:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$
Review – Multiple Regression Model

Assumption MLR.1 (Linearity in parameters)

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + u \]

E.g.: \( \text{Earnings} = \beta_0 + \beta_1 \text{Education} + \beta_2 \text{Experience} + \beta_3 \text{Tenure} + u \)

Assumption MLR.2 (Random Sampling from the population)

We have a random sample: \( \{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : i = 1, 2, \ldots n\} \)

satisfying the equation in MLR.1

Can write: \( \mathbf{y} = \mathbf{X} \beta + \mathbf{u} \) where,

\[
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}, \quad
\begin{bmatrix}
1 & x_{11} & x_{12} & \ldots & x_{1k} \\
1 & x_{21} & x_{22} & \ldots & x_{2k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \ldots & x_{nk}
\end{bmatrix}, \quad
\begin{bmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{bmatrix}, \quad \text{and } \beta = \begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_k
\end{bmatrix}
Review – Multiple Regression Model

**Assumption MLR.3 (No perfect Collinearity)**

- In the sample, none of the independent variables is a linear combination of the others
- This means that each variable has additional information
- If this assumption is violated, we cannot identify where the variations in $y$ are coming from

**Assumption MLR.4 (Zero Conditional Mean)**

$$E(u|x_1, x_2, \ldots, x_k) = 0$$  
this implies  
$$E(u) = 0$$
OLS Estimators

\[ \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik} \right)^2 \]

OLS obtained by minimising the Residuals Sum of Squares (SSR). First order conditions are:

\[ \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} u_i = 0 \]
\[ \sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{i1}u_i = 0 \]
\[ \sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{i2}u_i = 0 \]
\[ \ldots \]
\[ \sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{ik}u_i = 0 \]

\[ X'\hat{u} = 0 \iff X'(y - X\hat{\beta}) = 0 \]

or \[ (X'X)\hat{\beta} = X'y \] which leads to \[ \hat{\beta} = (X'X)^{-1}X'y \]

Assumption MLR.3 (No perfect Collinearity) implies that \((X'X)^{-1}\) exists.
Review - Properties of OLS: Unbiasedness

Under MLR.1 to MLR.4, the OLS estimator is unbiased for $\beta$:

$$E(\hat{\beta}_j) = \beta_j \quad j = 0, 1, \ldots, k$$

**Properties of OLS: Variance**

Needed an additional assumption:

**Assumption MLR.5 (Homoskedasticity)** $\text{Var}(u \mid x_1, x_2, \ldots, x_k) = \sigma^2$

which implies, together with random sampling ($\rightarrow u_i, u_j \ i \neq j$ are independent), that:

$$\text{Var}(u \mid X) = \sigma^2 I_n \quad \text{leading to} \quad \text{Var}(\hat{\beta} \mid X) = \sigma^2 (X'X)^{-1}$$

Note:

$$\text{Var}(\hat{\beta} \mid X) = \begin{bmatrix}
\text{Var}(\hat{\beta}_0 \mid X) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1 \mid X) & \cdots & \text{Cov}(\hat{\beta}_0, \hat{\beta}_k \mid X) \\
\text{Cov}(\hat{\beta}_1, \hat{\beta}_0 \mid X) & \text{Var}(\hat{\beta}_1 \mid X) & \cdots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_k \mid X) \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}(\hat{\beta}_k, \hat{\beta}_0 \mid X) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_1 \mid X) & \cdots & \text{Var}(\hat{\beta}_k \mid X)
\end{bmatrix}$$

$$\text{Cov}(\hat{\beta}_i, \hat{\beta}_j \mid X) = \text{Cov}(\hat{\beta}_j, \hat{\beta}_i \mid X)$$
Estimating the Error Variance

• In practice, the error variance $\sigma^2$ is unknown, must estimate it...

$$\hat{\sigma}^2 = \left( \sum_{i=1}^{n} \hat{u}_i^2 \right) / (n - k - 1) \equiv SSR/df$$

$df$ (degrees of freedom) is the (number of observations) – (number of estimated parameters)

• Can show that this is an unbiased estimator of the error variance

• The standard error of the regression is given by:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$
Decomposition of the Total Sum of Squares (SST)

\[ SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} \hat{u}_i^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = S SR + SSE \]

\[
R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = \frac{\left(\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)\left(\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2\right)}
\]

- \( R^2 \) can never decrease when another independent variable is added to a regression, and usually will increase.
- Because \( R^2 \) will usually increase with the number of independent variables, it is not a good measure to compare models.
Interpreting the Regression coefficients

- Let’s focus on the case of two explanatory variables (besides the constant), so \( k=2 \)

- We estimate: \( \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \),

- It can be shown that: \( \hat{\beta}_1 = \frac{\left( \sum_{i=1}^{n} \hat{r}_{i1} y_i \right)}{\sum_{i=1}^{n} \hat{r}_{i1}^2} \)

  where \( \hat{r}_{i1} \) are the residuals obtained when we estimate the regression

  \( \hat{x}_1 = \hat{y}_0 + \gamma_2 \hat{x}_2 \)

- Notice that \( \hat{\beta}_1 \) is the simple linear regression estimator, with \( \hat{r}_{i1} \) instead of the original regressor \( x_1 \) (note that the average of the residuals is always 0)

- Therefore, the estimated effect of \( x_1 \) on \( y \) equals the (simple regression) estimated effect of the “part” of \( x_1 \) that is not explained by \( x_2 \)
Interpreting the Regression coefficients

- Let us estimate the effect of Education on Wages, taking also into account the effect of Experience.

- Start by regressing education on experience (even if it seems silly...), storing the residuals of this regression.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>110.916</td>
<td>0.075</td>
</tr>
<tr>
<td>Labor Market Experience (in years)</td>
<td>-0.114</td>
<td>0.003</td>
</tr>
<tr>
<td>n</td>
<td>11064</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.172</td>
<td></td>
</tr>
</tbody>
</table>
Interpreting the Regression coefficients

• Now regress wages on the previously stored residuals

Dependent Variable: Wages

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>657.893</td>
<td>3.524</td>
</tr>
<tr>
<td>Residual from the Education regression</td>
<td>66.404</td>
<td>0.866</td>
</tr>
<tr>
<td>n</td>
<td>11064</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.347</td>
<td></td>
</tr>
</tbody>
</table>
Interpreting the Regression coefficients

• What if we regress wages on education and experience directly?

Dependent Variable: Wages

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-111.128</td>
<td>11.906</td>
</tr>
<tr>
<td>Education (in years)</td>
<td>66.4041</td>
<td>0.863</td>
</tr>
<tr>
<td>Labor Market Experience (in years)</td>
<td>11.672</td>
<td>0.301</td>
</tr>
</tbody>
</table>

| n                                        | 11064                |
| $R^2$                                     | 0.351                |

• Estimated coefficient of Education is the same as the coefficient from the previous regression (that **controls** for experience)!
Comparison with Simple Linear regression

Estimate the following regressions:

\[ \tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 \]
\[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \]

If we run the regression \( \hat{x}_1 = \hat{y}_0 + \hat{\gamma}_2 \hat{x}_2 \) and it turns out that \( \hat{\gamma}_2 = 0 \), this means that \( x_1 \) and \( x_2 \) are uncorrelated. Then, the residuals of this equation are the deviations of \( x_1 \) from its mean, which implies that \( \tilde{\beta}_1 = \hat{\beta}_1 \) (check formula for \( \tilde{\beta}_1 \) in the two interpretations)

If \( \hat{\beta}_2 = 0 \), it will also be the case that \( \tilde{\beta}_1 = \hat{\beta}_1 \) (check first order conditions)

However, in general it will be the case that \( \tilde{\beta}_1 \neq \hat{\beta}_1 \)
More Variables or Less Variables?

Assuming the factors in $u$ are uncorrelated with all the variables we are considering (2 in the previous example), we have learned:

- If we include “irrelevant” variables in our model, the estimated effects will be zero or close to zero and the remaining parameter estimates will remain unchanged (check first order conditions!!)

- So, OLS estimators are still **unbiased** if we include these variables
More Variables or Less Variables?

Also:

- If we do not include a variable \textbf{and} this variable is uncorrelated with the included regressors, then the OLS estimators will be \textit{unbiased}

- Remember, if the “other” factors (in $u$) tell us nothing about the regressors (they are uncorrelated with the regressors) we can still interpret the estimated effects as “ceteribus paribus” effects. No bias in this case

- If we do not include a variable \textbf{and} this variable is correlated with the included regressors, then the OLS estimators will be biased

- More on this in the practical sessions
So, always more variables? Even if they are irrelevant (or almost irrelevant) and therefore do not induce bias in the other estimators? No!

Why? Variances of the estimators can become large!

Can show, under MLR.1 to MLR.5, that:

\[
Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j (1 - R_{j}^2)} \quad j=1,2,\ldots,k
\]

\[
SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2
\]

\[
R_{j}^2
\]

Is the coefficient of determination from regressing \(x_j\) on all the other regressors. Tells us how much the other regressors “explain” \(x_j\)
Understanding OLS Variances

\[ Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R^2_j)} \quad j=1,2,\ldots,k \]

So, why can variances increase? Strong linear relations among the independent variables are harmful: a larger \( R^2_j \) implies a larger variance for the estimators (almost multicollinearity). If we keep adding variables, \( R^2_j \) will always increase. If it goes close to 1 we are in trouble…

- If the new “irrelevant” variables are uncorrelated to the already included regressors, then the variance remains unchanged
- A larger \( \sigma^2 \) implies a larger variance of the OLS estimators (as before)
- A larger \( SST_j \) implies a smaller variance of the estimators (increases with sample size, so in large samples we should not be too worried)
The Gauss-Markov Theorem

- Under MLR.1 to MLR.5 (the so-called Gauss-Markov Assumptions) it can be shown that OLS is “BLUE”

- Best Linear Unbiased Estimator

- Thus, if the 5 assumptions are presumed to hold, use OLS

No other linear and unbiased estimator \( \hat{\beta} = A'y \) has a variance smaller than OLS

Variances here are matrices, we are saying that

\[ \text{Var}(\hat{\beta}|X) = \text{Var}(\hat{\beta}|X) - \text{Var}(\hat{\beta}|X) \]

is a positive semi-definite matrix (implies that all individual OLS parameter estimators have smaller variance than any other linear unbiased estimator for those parameters)