Fundamentals of Econometrics

Lecture 3: Multiple Linear Regression Model

Dependent Variable (Response Variable)

Independent Variables (Predictors)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \varepsilon$$

Y intercept

Slope Coefficient

Error Term

Section 1

Multiple Regression Analysis: Estimation

Motivation

- Key assumption in the single variable model is that all unobserved factors are uncorrelated with the observed variable.
- **Is this a problem?**
  - Remember our wage model: $wage = \beta_0 + \beta_1 \cdot educ + u$.
- Recall that $\hat{\beta}_1$ measures the impact of education on wage. In this model, we assume that education is the only factor that affects wage and that all other factors are uncorrelated with education.
- Other factors that affect the wage rate could very well be correlated with education.
- Effect of experience would partly be included in the effect of education on wage.

Consider the following model:

$$wage = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot occup + \beta_3 \cdot gender + \beta_4 \cdot race + \beta_5 \cdot ability + u$$

- Assuming $E(u|educ, occup, gender, race, ability) = 0$, then β_1 measures the impact of education holding the other predictors (occup, gender, race, ability) constant.

What are we attempting to do by including additional variables?

We are trying to replicate *ceteris paribus* conditions of an experiment by including additional variables.

Multiple Regression Model

We can generalize the simple regression model to:

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + u_i$$

- **K = number of explanatory variables**.
- β_k measures the impact on y of a one-unit change in x_k.
- Will again work with a random sample of size N.
- u is still an error term capturing unobservables.

An Example

We return to the wage and schooling example from the previous lecture but now posit:

$$wage = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot exper + \beta_3 \cdot tenure + u$$

```r
mult1 <- lm(formula = wage ~ educ + exper + tenure, data = wage1); summary(mult1)
```

```
##
## Call:
## lm(formula = wage ~ educ + exper + tenure, data = wage1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6068 -1.7747 -0.6279  1.1969 14.6536
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.87273    0.72896  -3.941 9.22e-05 ***
## educ         0.59897    0.05128  11.679  < 2e-16 ***
## exper        0.02234    0.01206   1.853   0.0645 .
## tenure       0.16927    0.02164   7.820 2.93e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.084 on 522 degrees of freedom
## Multiple R-squared:  0.3064, Adjusted R-squared:  0.3024
## F-statistic: 76.87 on 3 and 522 DF,  p-value: < 2.2e-16
```

Presenting the Results professionally

```r
stargazer(mult1,
      title = "OLS Regression Results", font.size = "tiny",
      type = "latex", header = FALSE, digits = 3)
```

Table 1: OLS Regression Results

	Dependent variable:
	wage
educ	0.599^{***}
	(0.051)
exper	0.022^{*}
	(0.012)
tenure	0.169^{***}
	(0.022)
Constant	-2.873^{***}
	(0.729)
Observations	526
R^2	0.306
Adjusted R^2	0.302
Residual Std. Error	3.084 (df = 522)
F Statistic	76.873^{***} (df = 3; 522)
Note:	*p<0.1; **p<0.05; ***p<0.01

Functional Form and Units of Measurement

- Now we can enable all sorts of different functional forms.

Examples:

- Quadratic: $wage = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot educ^2 + u$.
- Cubic: $wage = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot educ^2 + \beta_3 \cdot educ^3 + u$.

How do we assess the impact of education on wage?

In the quadratic model, there are two explanatory variables: $educ$ and $educ^2$. Wage is explained as a quadratic function of education.
So, by how much does wage increase for a one-unit increase in education?

$$\frac{\partial wage}{\partial educ} = \beta_1 + 2\beta_2 educ$$

This means the effect is not constant but depends on the level of education already attained.

```r
w2 <- lm(formula = wage ~ educ + I(educ^2) + I(educ^3),
         data = wage1)

stargazer::stargazer(w2, title = "OLS Regression Results",
                     type = "latex", header = FALSE, digits = 3)
```

Table 2: OLS Regression Results

	Dependent variable:
	wage
educ	0.369
	(0.619)
I(educˆ2)	−0.058
	(0.063)
I(educˆ3)	0.003^{*}
	(0.002)
Constant	3.111
	(1.979)
Observations	526
R^2	0.205
Adjusted R^2	0.201
Residual Std. Error	3.301 (df = 522)
F Statistic	45.004^{***} (df = 3; 522)
Note:	*p<0.1; **p<0.05; ***p<0.01

CEO salary, sales and CEO tenure

$$log(salary) = \beta_0 + \beta_1 \cdot log(sales) + \beta_2 \cdot ceoten + \beta_3 \cdot ceoten^2 + u$$

- Model assumes a constant elasticity of salary with respect to sales.
- The effect of CEO tenure on salary is not constant but depends on the level of tenure (quadratic as we saw earlier).

Notion of Linearity

Meaning of "linear" regression is that the coefficients are linear in the parameters, not in the variables.

Determinants of college GPA

$$\widehat{colGPA} = 1.29 + 0.45 \cdot hsGPA + 0.009 \cdot ACT$$

```r
lm(colGPA ~ hsGPA + ACT, data = gpa1) |> coef()
```

```
## (Intercept)       hsGPA         ACT
## 1.286327767 0.453455885 0.009426012
```

Interpretations:

- Another point on the student's high school GPA will lead to a 0.453 increase in college GPA, holding ACT score constant.
- Holding high school GPA constant, a **10-point** increase in ACT score will lead to a **0.09** increase in college GPA.

Deriving the OLS Estimators

- Sample regression function is now:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_K x_{iK}$$

- Minimize sum of squared residuals:

$$\sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_K x_{iK})^2$$

- FOC:

$$\sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \cdots - \hat{\beta}_K x_{iK}) = 0$$

$$\sum_{i=1}^{N} x_{ij} (y_i - \beta_0 - \cdots - \hat{\beta}_K x_{iK}) = 0 \quad j = 1, \ldots, K$$

Matrix Algebra

- **Matrix**: A rectangular array of numbers with N rows and K columns.
- **Vector**: A matrix with a single column or row.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

- **Transpose**: A matrix obtained by interchanging rows and columns of a matrix.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \; A' = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}, \; B = \begin{pmatrix} 2 & -4 \\ -1 & 5 \\ 3 & 0 \end{pmatrix} \; B' = \begin{pmatrix} 2 & -1 & 3 \\ -4 & 5 & 0 \end{pmatrix}$$

Matrix multiplication can be written as $C = AB$ but note that the column dimension of A must be equal to the row dimension of B.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}, \ B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix}$$

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \end{pmatrix}$$

The sum of squares can be obtained by multiplying a column vector by its transpose:

$$U = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \ U' = \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} \longrightarrow U'U = \begin{pmatrix} u_1^2 + u_2^2 + u_3^2 \end{pmatrix}$$

Matrix Addition requires the matrices to have the same dimensions.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \; B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$C = A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

The regression model can then be given in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1K} \\ 1 & x_{21} & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{NK} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}$$

Deriving the OLS Estimators

- Minimize (in matrix form):

$$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

- FOC:

$$-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

OLS Estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Covariances of $\mathbf{X}$

Let us consider a model with two explanatory variables (covariates):

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i.$$

The three first order conditions are given by:

$$\hat{\beta}_0 : \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} \right) = 0 \tag{1}$$

$$\hat{\beta}_1 : \frac{1}{N} \sum_{i=1}^{N} x_{1i} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} \right) = 0 \tag{2}$$

$$\hat{\beta}_2 : \frac{1}{N} \sum_{i=1}^{N} x_{2i} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} \right) = 0 \tag{3}$$

Solving Equation (1) for $\hat{\beta}_0$:

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2} \tag{4}$$

Substitute Equation (4) into (2):

$$\frac{1}{N} \sum_{i=1}^{N} x_{1i} - \left(y_i - \left[\bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \right] + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} \right) = 0$$

$$\frac{1}{N} \sum_{i=1}^{N} x_{1i}(y_i - \bar{y}) - \frac{1}{N} \hat{\beta}_1 \sum_{i=1}^{N} x_{1i} \left(x_{1i} - \hat{\beta}_1 \bar{x}_1 \right) - \frac{1}{N} \hat{\beta}_2 \sum_{i=1}^{N} x_{1i} \left(x_{2i} - \hat{\beta}_2 \bar{x}_2 \right) = 0$$

$$cov(x_1, y) - \hat{\beta}_1 var(x_1) - \hat{\beta}_2 cov(x_1, x_2) = 0 \tag{5}$$

$$\boxed{\hat{\beta}_1 = \frac{cov(x_1, y)}{var(x_1)} - \hat{\beta}_2 \frac{cov(x_1, x_2)}{var(x_1)}} \tag{6}$$

Likewise,

$$\boxed{\hat{\beta}_2 = \frac{cov(x_2, y)}{var(x_2)} - \hat{\beta}_1 \frac{cov(x_1, x_2)}{var(x_2)}}$$

Digression: Partialling Out

The Frisch-Waugh-Lovell Theorem: We can use the concept of partialling out to estimate the marginal effect of an independent variable, say x_1, on the dependent variable y, controlling for another variable x_2.

We can proceed as follows:

1. Regress $x1$ on x_2 and obtain the residuals $\hat{r}_1$. This will give us the part of x_1 that is not in x_2. Similar logic for x_2.
2. Regress y on $\hat{r}_1$ to obtain the coefficient $\hat{\beta}_1$. This will tell us the effect of x_1 on y after controlling for x_2.

Why does this work?

1. The residuals from the first stage regression are orthogonal to (uncorrelated with) x_2 (i.e. $cov(x_2, \hat{r}_1) = 0$). Again, $\hat{r}_1$ contains everything in x_1 that is not in x_2.
2. The slope coefficient $\hat{\beta}_1$ in the second stage regression tells us the effect of $\hat{r}_1$ (everything in x_1 that is not in x_2) on y.

Example: Partialling Out

Consider the following model:

$$wage = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot exper + u$$

Step 1 (a): Regress education on experience to obtain the residuals $\hat{r}_{educ}$.

```
hatr_educ <- lm(educ ~ exper, data = wage1) |> residuals()
```

Step 1 (b): Regress experience on education to obtain the residuals $\hat{r}_{exper}$.

```
hatr_exper <- lm(exper ~ educ, data = wage1) |> residuals()
```

Step 2: Separately regress wage on $\hat{r}_{educ}$ and $\hat{r}_{exper}$ to obtain $\hat{\beta}_1$ and $\hat{\beta}_2$.

```
c((lm(wage ~ hatr_educ, data = wage1) |> coef())[2],
(lm(wage ~ hatr_exper, data = wage1) |> coef())[2])
```

```
##  hatr_educ hatr_exper
##  0.6442721  0.0700954
```

Example: Partialling Out

How does this compare to the OLS estimates of the multiple regression of wage on both education and experience?

```r
lm(wage ~ educ + exper, data = wage1) |> coef()
```

```
## (Intercept)        educ       exper
##  -3.3905395   0.6442721   0.0700954
```

Goodness-of-fit

The idea is the same as in simple regression.

- **Total Sum of Squares (SST)**: The total variation in the dependent variable.
- **Explained Sum of Squares (SSE)**: The variation in the dependent variable explained by the model.
- **Residual Sum of Squares (SSR)**: The variation in the dependent variable not explained by the model.

$$SST = SSE + SSR$$

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Problem:

- R^2 **almost always** increase as we add more variables to the model.
- Therefore R^2 can be a poor indicator of whether to include more RHS variables.

- Consider the following multivariate model:

$$wage = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot exper + \beta_3 \cdot tenure + u$$

```r
lm(wage ~ educ + exper + tenure, data = wage1) |> summary()
```

```
##
## Call:
## lm(formula = wage ~ educ + exper + tenure, data = wage1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6068 -1.7747 -0.6279  1.1969 14.6536
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.87273    0.72896  -3.941 9.22e-05 ***
## educ         0.59897    0.05128  11.679  < 2e-16 ***
## exper        0.02234    0.01206   1.853   0.0645 .
## tenure       0.16927    0.02164   7.820 2.93e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.084 on 522 degrees of freedom
## Multiple R-squared:  0.3064, Adjusted R-squared:  0.3024
## F-statistic: 76.87 on 3 and 522 DF,  p-value: < 2.2e-16
```

$$wage = -2.87 + 0.60 \cdot educ + 0.02 \cdot exper + 0.17 \cdot tenure$$
$$n = 522, \quad R^2 = 0.3064$$

- A one-year increase in education is associated with a \$0.60 increase in average hourly wage, *holding experience and tenure constant.*
- A one-year increase in experience is associated with a \$0.02 increase in average hourly wage, *holding education and tenure constant.*
- A one-year increase in tenure is associated with a 0.17 increase in average hourly wage, *holding education and experience constant.*
- The R^2 indicates that 30.64% of the variation in wage is *jointly* explained by the individual's education, experience, and tenure.

Section 2

MLR Assumptions

MLR Assumptions

1. **Linearity**: The model is linear in the parameters.
2. **Random Sampling**: The data are a random sample from the population.
3. **No Perfect Collinearity**: None of the explanatory variables is constant and there are no exact linear relationships among the explanatory variables.
   - If an explanatory variable is a perfect linear combination of other variables it is redundant and can be dropped.
   - Constant variables are also ruled out as they are collinear with the intercept.
4. **Zero Conditional Mean**: $E(u|x_1, x_2, \ldots, x_K) = 0$.
   - This is the same as the assumption in the simple regression model. The value of explanatory variables must contain no information about the mean of the unobserved factors.

Collinearity

Recall the voting example from the previous lecture. Where

$$shareA : 100 * (expendA/(expendA + expendB))$$

If we decided to add the expenditure shares of both candidates A and B as explanatory variables, we would have a problem.

$$voteA = \beta_0 + \beta_1 \cdot shareA + \beta_2 \cdot shareB + u$$

Either `shareA` or `shareB` will have to be dropped from the model because there is an exact linear relationship between the two variables: $shareA + shareB = 1$.

Do we have a collinearity problem here?

$$y = \beta_0 + \beta_1 income + \beta_2 income^2 + u$$

Including Irrelevant Variables

Suppose we estimate

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 x_3 + u$$

When the true model is

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + u$$

Are β_0, β_1, and β_2 still unbiased?

Table 3: Wage Regression Results - Irrelevant Variable

	Dependent variable:	
	wage	
	(1)	(2)
educ	0.599***	0.598***
	(0.051)	(0.051)
exper	0.022*	0.022*
	(0.012)	(0.012)
tenure	0.169***	0.172***
	(0.022)	(0.022)
BMI		−0.030
		(0.028)
Constant	−2.873***	−1.451
	(0.729)	(1.527)
Observations	526	526
R^2	0.306	0.308
Adjusted R^2	0.302	0.303
Residual Std. Error	3.084 (df = 522)	3.084 (df = 521)
F Statistic	76.873*** (df = 3; 522)	57.950*** (df = 4; 521)
Note:		*p<0.1; **p<0.05; ***p<0.01

Omitting a Relevant Variable

Suppose the true model is

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + u \tag{7}$$

but we estimated

$$\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 \cdot x_1 + \widetilde{u} \tag{8}$$

If x_2 is correlated with x_1, then $\widetilde{\beta}_1$ will be biased. To show this:

1. Consider the auxiliary regression:

$$x_2 = \gamma_0 + \gamma_1 \cdot x_1 + r_2 \tag{9}$$

2. Substitute Equation (9) into Equation (7):

$$\begin{aligned} y &= \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot (\gamma_0 + \gamma_1 \cdot x_1 + r_2) + u && (10) \\ &= \underbrace{(\beta_0 + \beta_2 \cdot \gamma_0)}_{\widetilde{\beta}_0} + \underbrace{(\beta_1 + \beta_2 \cdot \gamma_1)}_{\widetilde{\beta}_1} \cdot x_1 + \underbrace{\beta_2 \cdot r_2 + u}_{\widetilde{u} = \text{error term}} && (11) \end{aligned}$$

Omitted Variable Bias

Again, consider the following model:

$$\text{(True Model) } wage = \beta_1 \cdot educ + \beta_2 \cdot abil + u \tag{12}$$

$$\text{(Estimated Model) } wage = \delta_0 + \delta_1 \cdot educ + v \tag{13}$$

In Equations (12) and (13), β_2 and δ_1 should be positive.

$$wage = (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1) \cdot educ + (\beta_2 \cdot v + u) \tag{14}$$

The return to education, β_1 will be overestimated since $\beta_2 \cdot \delta_1 > 0$. Our biased model, in Equation (13), will give us the impression that people with many years of education earn very high wages, but a part of this could be due to the fact that people with more education are also more able, on average.

Table 4: Wage Regression Results - Omitted Variable Bias

	Dependent variable:		
	wage		exper
	(1)	(2)	(3)
educ	0.644***	0.541***	−1.468***
	(0.054)	(0.053)	(0.204)
exper	0.070***		
	(0.011)		
Constant	−3.391***	−0.905	35.461***
	(0.767)	(0.685)	(2.628)
Observations	526	526	526
R^2	0.225	0.165	0.090
Adjusted R^2	0.222	0.163	0.088
Note:			*p<0.1; **p<0.05; ***p<0.01

Omitted Variable Bias

We can summarize the omitted variable bias as follows:

x	$corr(x_1, x2) > 0$	$corr(x_1, x2) < 0$
$\beta_2 > 0$	Positive Bias	Negative Bias
$\beta_2 < 0$	Negative Bias	Positive Bias

MLR Assumptions

5. **Homoskedasticity**: $Var(u|x_1, x_2, \ldots, x_K) = \sigma^2$.

- The variance of the error term is constant across all values of the explanatory variables.
- The value of the explanatory variables must contain no information about the variance of the unobserved factors.
- This is the same as the assumption in the simple regression model.

Example

$$var(u|educ, exper, tenure) = \sigma^2$$

- This assumption may also be hard to justify.

Sample Variances

The variances of the OLS estimators are given by:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)},\ j = 1, \ldots, k$$

- σ^2 is the variance of the error term.
- SST_j is the total sample variance of x_j, $\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2$, and
- R_j^2 is the R^2 from regressing x_j on all other explanatory and a constant term.

Sample Variances: Matrix Algebra

Recall, OLS Estimator is given by: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ then

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right] \\
&= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{u})\right] \\
&= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}\right] \\
&= \boldsymbol{\beta} + E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}\right] \\
var(\hat{\boldsymbol{\beta}}) &= E\left[(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}))'\right] \\
&= E\left[(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u} - \beta)(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u} - \beta)'\right] \\
&= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}\boldsymbol{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{u}\boldsymbol{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_N\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

Sample Variances

Components of the OLS variances:

1. **Error Variance**: σ^2.
   - A high error variance increases the sampling variance because there is more "noise" in the equation.
   - A large error variance doesn't necessarily mean make the estimates imprecise.
   - The error variance does not decrease with sample size.
2. **Total variation in the explanatory variable**: SST_j.
   - More sample variation leads to more precise estimates.
   - The total variation in the explanatory variable automatically increases with sample size.
     - Increasing the sample size is therefore a way to get more precise estimates.

Components of the OLS variances (contd.):

3. **Linear relationship between the explanatory variables**: R_j^2.

- Regress x_j on all other explanatory variables and a constant term.
- The R^2 from this regression will be higher when x_j can be better explained by the other explanatory variables.
- Under perfect multicollinearity, the variance of the slope estimator will approach infinity.

Multicollinearity

Multicollinearity occurs when two or more explanatory variables are highly, but not perfectly correlated.

Consider $y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + u$, where $corr(x_2, x_3)$ is exceptionally large.

Implications for the estimates:

- Estimators are still unbiased.
- The variance of the estimators, β_2 and β_3 may be very large.
- Variance estimator for β_1 is okay as long as x_1 is not correlated with x_2 and x_3.

Multicollinearity

- Only the sampling variance of the variables involved in multicollinearity will be inflated; the estimates of the other effects may still be precise.
- Note that multicollinearity does not violate MLR.3 in the strict sense.
- Multicollinearity may be detected through "variance inflation factors" (VIFs).

$$VIF_j = \frac{1}{1 - R_j^2}$$

An arbitrary rule of thumb is that a VIF greater than 10 indicates multicollinearity.

```r
car::vif(model)
```

Estimating σ^2

- Recall from the univariate regression model:

$$\hat{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{N}\hat{u}_i^2 \tag{15}$$

- We subtracted 2 because we estimated two (2) parameters: the intercept and the slope. Thus, two first-order conditions.
- In estimating the multivariate model, we estimate $K+1$ parameters (including the intercept)– also the number of first-order conditions.
- Therefore, the formula for $\hat{\sigma}^2$ in the multivariate model is:

$$\hat{\sigma}^2 = \frac{1}{N-K-1}\sum_{i=1}^{N}\hat{u}_i^2 \tag{16}$$

Assumptions MLR 1 – 5 imply that:

$$E\left[\hat{\sigma}^2\right] = \sigma^2$$

Sample Variances

- The true sampling variation of the estimated β_j is given by:

$$sd(\hat{\beta}_j) = \sqrt{var(\hat{\beta}_j)} = \sqrt{\sigma^2 / SST_j(1 - R_j^2)}$$

- The estimated sampling variation of the estimated β_j is given by:

$$se(\hat{\beta}_j) = \sqrt{\widehat{var}(\hat{\beta}_j)} = \sqrt{\widehat{\sigma^2} / SST_j(1 - R_j^2)}$$

Gauss Markov Theorem

- Under MLR Assumptions 1 – 4, the OLS estimators are **unbiased**.
- However, under these assumptions there may be many other estimators that are also unbiased.
- We want to choose the estimator that has the **smallest variance**.
- called the **BLUE** estimator (Best Linear Unbiased Estimator).
- In order to answer this question one usually limits oneself to linear estimators, i.e. estimators linear in the dependent variable.
- Under MLR Assumptions 1-5, the OLS estimators are **BLUE**.
    - see Wooldridge for proof.