Multiple Regression Analysis: Estimation. Wooldridge (2013), Chapter 3

- Ordinary Least Squares (OLS) Estimator
- Interpreting Multiple Regression
- A “Partialling Out” Interpretation of the OLS estimator - Frisch-Waugh (1933) Theorem
- Simple vs Multiple Regression Estimate
- The R-squared
- Unbiasedness of the OLS estimator
- Too Many or Too Few Variables
- Variance of the OLS Estimators

The Multiple Regression model takes the form

$$E\left(y|x_1, ..., x_k\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

or equivalently

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u,$$

where $E\left(u|x_1, ..., x_k\right) = 0$.
Parallels with Simple Regression:

- y is the dependent variable (regressand).
- $x_1, ..., x_k$ are the k regressors.
- u is still the error term (or disturbance).
- β_0 is still the intercept.
- β_1 to β_k all called slope parameters.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u,$$

where $E\left(u|x_1, ..., x_k\right) = 0$.

Examples:

- $y-$sales, the regressors are advertising expenditure, income, price relative to competitors.
- $y-$ personal consumption, the regressors are disposable income, wealth, interest rates.
- $y-$ Investment, the regressors are interest rates and profits (past and future).
- $y-$ Wages, the regressors are schooling, experience, ability and gender.

To estimate $\beta_0, \beta_1, \beta_2, ..., \beta_k$ we choose $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ that minimize

$$\mathcal{S}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik}\right)^2$$

The first order conditions are

$$\begin{aligned} -\frac{2}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik}) &= 0 \\ -\frac{2}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik})x_{ij} &= 0 \\ j &= 1, ..., k \end{aligned}$$

- This is a system of equations with $k+1$ equations and $k+1$ variables: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$. The Ordinary Least Squares estimator is obtained by solving the system of equations for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$.

The first order conditions can be written as

$$-\frac{2}{n}\sum_{i=1}^{n}\hat{u}_i = 0, \quad (1)$$

$$-\frac{2}{n}\sum_{i=1}^{n}\hat{u}_i x_{ij} = 0, \quad (2)$$

$$j = 1, ..., k,$$

where $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik}$.(*residuals*)

Remarks:

- Beyond the two-variable case it is not possible to write out an explicit formula for the OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ (without the use of matrix algebra), although a solution exists.
- Equation (1) implies that the sum and the mean of the residuals are zero.
- Equations (1) and (2) imply that the covariances between the residuals and each regressor are zero.

The OLS regression line (*fitted values*) is now defined as

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k.$$

Writing it in terms of changes we obtain

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + ... + \hat{\beta}_k \Delta x_k.$$

Holding x_i ,$i = 1, ...k$ and $i \neq j$ fixed implies that

$$\Delta\hat{y} = \hat{\beta}_j \Delta x_j,$$

$j = 1, ..., k$. Thus each β has a ceteris paribus interpretation.

- In most cases, we will indicate the estimation of a relationship through OLS by writing as

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k. \tag{3}$$

- Sometimes, for the sake of brevity, it is useful to indicate that an OLS regression has been run without actually writing out the equation.
- We will often indicate that equation (3) has been obtained by OLS in saying that we run the regression of y on $x_1, x_2, ..., x_k$

- Regression of Wages on years of Education and years of Work Experience:

Dependent variable: Wages

Estimation Method: Ordinary Least Squares

Regressors	Estimates
Intercept	-5.56732
Education	0.97685
Experience	0.10367

- Another year of Education is predicted to increase the mean of wages by \$0.97685, holding Experience fixed.
- Another year of Experience is predicted to increase the mean of wages by \$0.10367, holding Education fixed.

Consider the case $k = 2$, i.e.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

There is an interesting interpretation for $\hat{\beta}_1$:

- Let $\hat{r}_{i1}$ be the residuals from the regression of x_1 on x_2. The fitted values are $\hat{x}_1 = \hat{\gamma}_0 + \hat{\gamma}_2 x_2$.
- Notice that for $i = 1, ..., n$

$$x_{i1} = \underbrace{\hat{x}_{i1}}_{\substack{\text{part of } x_1 \text{ that can} \\ \text{be explained by } x_2}} + \underbrace{\hat{r}_{i1}}_{\substack{\text{part of } x_1 \text{ that cannot} \\ \text{be explained by } x_2}}$$

- It can be shown that the OLS estimator for β_1, $\hat{\beta}_1$, is equal to the estimator of the slope when we run a regression of y_i on $\hat{r}_{i1}$. That is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1} y_i}{\sum_{i=1}^n \hat{r}_{i1}^2}.$$

- It can be shown that the OLS estimator for β_1, $\hat{\beta}_1$, is equal to the estimator of the slope when we run a regression of y_i on $\hat{r}_{i1}$. That is
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2},$$
- What is the interpretation of this?
- We're estimating the effect of x_1 on y after removing from x_1 the effect of x_2.

Compare the simple regression

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

with the multiple regression

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

Generally $\tilde{\beta}_1 \neq \hat{\beta}_1$ unless $\hat{\beta}_2 = 0$ (i.e. no partial effect of x_2) or x_1 and x_2 are uncorrelated in the sample.

Example:

- Regression of Wages on Education

Dependent valiable: Wages

Estimation Method: Ordinary Least Squares, sample size: 528

Regressors	Estimates
Intercept	-1.60468
Education	0.81395

- Regression of Wages on Education and Experience

Dependent valiable: Wages

Estimation Method: Ordinary Least Squares, sample size: 528

Regressors	Estimates
Intercept	-5.56732
Education	0.97685
Experience	0.10367

As in the simple regression model we can think of each observation as being made up of an explained part, and an unexplained part, $y_i = \hat{y}_i + \hat{u}_i$.
We then define the following:

- $\sum_{i=1}^{n} (y_i - \bar{y})^2$ is the *total sum of squares* (SST).
- $\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ is the *explained sum of squares* (SSE).
- $\sum_{i=1}^{n} \hat{u}_i^2$ is the *residual sum of squares* (SSR).

(Same definitions as in the linear regression model)
Then

$$SST = SSE + SSR.$$

Prove this result in the simple regression model!

Proof:
Recall that in the simple regression model we had

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0, \\ \frac{1}{n}\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0. \end{aligned}$$

But since $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i &= 0, \\ \frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i &= 0. \end{aligned}$$

By definition, we have

$$\begin{aligned}\hat{u}_i &= y_i - \hat{y}_i, \\ y_i &= \hat{y}_i + \hat{u}_i.\end{aligned}$$

Therefore

$$\begin{aligned}\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i + \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i \\ &= \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \bar{\hat{y}}\end{aligned}$$

because $\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0$ and $\bar{\hat{y}}$ is the average of the fitted values.

We prove now that

$$\sum_{i=1}^{n} \hat{u}_i \hat{y}_i = 0.$$

Notice that $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, therefore

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) &= \sum_{i=1}^{n} \left(\hat{\beta}_0 \hat{u}_i + \hat{\beta}_1 x_i \hat{u}_i\right) \\
&= \sum_{i=1}^{n} \hat{\beta}_0 \hat{u}_i + \sum_{i=1}^{n} \hat{\beta}_1 x_i \hat{u}_i \\
&= \hat{\beta}_0 \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i \hat{u}_i \\
&= 0
\end{aligned}$$

because $\sum_{i=1}^{n} \hat{u}_i = 0$ and $\sum_{i=1}^{n} x_i \hat{u}_i = 0$.

Now we are going to prove that

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

Given that $y_i = \hat{y}_i + \hat{u}_i$, we have

$$\begin{aligned} \sum_{i=1}^{n} (y_i - \bar{y})^2 &= \sum_{i=1}^{n} (\hat{y}_i + \hat{u}_i - \bar{y})^2 \\ &= \sum_{i=1}^{n} [(\hat{y}_i - \bar{y}) + \hat{u}_i]^2 \\ &= \sum_{i=1}^{n} \left[(\hat{y}_i - \bar{y})^2 + \hat{u}_i^2 + 2(\hat{y}_i - \bar{y})\hat{u}_i \right] \\ &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} \hat{u}_i^2 + \sum_{i=1}^{n} 2(\hat{y}_i - \bar{y})\hat{u}_i \end{aligned}$$

Now notice that

$$\begin{aligned}
\sum_{i=1}^{n} 2\left(\hat{y}_i - \bar{y}\right)\hat{u}_i &= 2\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)\hat{u}_i \\
&= 2\sum_{i=1}^{n}\left(\hat{y}_i\hat{u}_i - \bar{y}\hat{u}_i\right) \\
&= 2\left(\sum_{i=1}^{n}\hat{y}_i\hat{u}_i - \sum_{i=1}^{n}\bar{y}\hat{u}_i\right) \\
&= 2\left(\sum_{i=1}^{n}\hat{y}_i\hat{u}_i - \bar{y}\sum_{i=1}^{n}\hat{u}_i\right) \\
&= 0
\end{aligned}$$

because $\sum_{i=1}^{n}\hat{y}_i\hat{u}_i = 0$ and $\sum_{i=1}^{n}\hat{u}_i = 0$.

Multiple Regression Analysis

Goodness-of-Fit

- Can compute the fraction of the total sum of squares (SST) that is explained by the model, call this the *R-squared* of regression:

$$R^2 = SSE/SST = 1 - SSR/SST,$$

where

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

- R^2 is a measure of *Goodness of fit*: proportion of the variance of the dependent variable that is explained by the model.
- The R^2 is called the *coefficient of determination*.
- $0 \leq R^2 \leq 1$.

It can be shown that R^2 is equal to the squares of the correlation between $\hat{y}$ and y

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

(also valid for the simple regression model). It can also be shown that $\bar{\hat{y}} = \bar{y}$.

- 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 way to compare models.
- An alternative measure usually reported by any statistical software is the *adjusted R-squared*:

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} \\ &= 1 - \frac{(n-1)}{(n-k-1)}(1-R^2).\end{aligned}$$

- $\bar{R}^2$ penalizes the number of regressors included.
- However, $\bar{R}^2$, is not not between 0 and 1. In fact, it can be negative.

Example: Regression of Wages on Education and Experience

Dependent valiable: Wages

Estimation Method: Ordinary Least Squares, sample size: 528

Regressors	Estimates
Intercept	-5.56732
Education	0.97685
Experience	0.10367

$$R^2 = 0.209, \ \bar{R}^2 = 0.206$$

- Population model is linear in parameters: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$.
- We can use a random sample of size $n, \{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : i = 1, 2, \ldots, n\}$, from the population model, so that the sample model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i.$$

- $E(u|x_1, x_2, \ldots\ x_k) = 0$, implying that all of the explanatory variables are exogenous.
- None of the x's is constant, and there are no exact linear relationships among them (no perfect *multicolinearity*).

Proposition

Under the above assumptions the OLS estimators for β_0, β_1, ...β_k are unbiased, that is

$$E\left(\hat{\beta}_j\right) = \beta_j,\ j = 1, ..., k.$$

(prove this result in the simple regression model).

Proof:
Recall that in the simple regression model we had

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},\end{aligned}$$

We proved before that $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i - \bar{x}) x_i$.
We prove now that $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x}) y_i$.
Notice that

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n \left[(x_i - \bar{x}) y_i - (x_i - \bar{x}) \bar{y}\right] \\ &= \sum_{i=1}^n (x_i - \bar{x}) y_i - \sum_{i=1}^n (x_i - \bar{x}) \bar{y} \\ &= \sum_{i=1}^n (x_i - \bar{x}) y_i - \bar{y} \sum_{i=1}^n (x_i - \bar{x}) \\ &= \sum_{i=1}^n (x_i - \bar{x}) y_i\end{aligned}$$

because, as we proved before, $\sum_{i=1}^n (x_i - \bar{x}) = 0$.

Therefore

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\, y_i}{\sum_{i=1}^{n}(x_i - \bar{x})\, x_i}$$

Recall that in the simple regression model we have

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + u_i, i = 1, ..., n \\ E(u_i | x_i) &= 0. \end{aligned}$$

We have to prove that $E(\hat{\beta}_1) = \beta_1$ and $E(\hat{\beta}_0) = \beta_0$
Write $\tilde{x} = (x_1, x_2, ..., x_n)$, therefore by the law of iterated expectations we have

$$E(\hat{\beta}_1) = E(E(\hat{\beta}_1 | \tilde{x}))$$

Now

$$
\begin{aligned}
E\left(\hat{\beta}_1|\tilde{x}\right) &= E\left(\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)y_i}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)x_i}|\tilde{x}\right) \\
&= \frac{1}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)x_i}E\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)y_i|\tilde{x}\right) \\
&= \frac{1}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)x_i}\sum_{i=1}^{n}E\left(\left(x_i-\bar{x}\right)y_i|\tilde{x}\right) \\
&= \frac{1}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)x_i}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)E\left(y_i|\tilde{x}\right)
\end{aligned}
$$

and $E\left(y_i|\tilde{x}\right)=E\left(y_i|x_1,x_2,...,x_i,...,x_n\right)=E\left(y_i|x_i\right)$ because y_i is independent from x_j for $j\neq i$ as we assumed that we use a random sample $\left\{\left(x_i,y_i\right)\right\}_{i=1}^{n}$

Now notice that $E\left(y_i|x_i\right)=\beta_0+\beta_1 x_i$, therefore

$$
\begin{aligned}
E\left(\hat{\beta}_1|\tilde{x}\right) &= \frac{1}{\sum_{i=1}^n\left(x_i-\bar{x}\right)x_i}\sum_{i=1}^n\left(x_i-\bar{x}\right)\left(\beta_0+\beta_1 x_i\right)\\
&= \frac{\sum_{i=1}^n\left(x_i-\bar{x}\right)\beta_0+\sum_{i=1}^n\left(x_i-\bar{x}\right)\beta_1 x_i}{\sum_{i=1}^n\left(x_i-\bar{x}\right)x_i}\\
&= \frac{\beta_0\sum_{i=1}^n\left(x_i-\bar{x}\right)+\beta_1\sum_{i=1}^n\left(x_i-\bar{x}\right)x_i}{\sum_{i=1}^n\left(x_i-\bar{x}\right)x_i}\\
&= \frac{\beta_1\sum_{i=1}^n\left(x_i-\bar{x}\right)x_i}{\sum_{i=1}^n\left(x_i-\bar{x}\right)x_i}\\
&= \beta_1
\end{aligned}
$$

therefore

$$
\begin{aligned}
E\left(\hat{\beta}_1\right) &= E\left(E\left(\hat{\beta}_1|\tilde{x}\right)\right)\\
&= E\left(\beta_1\right)\\
&= \beta_1
\end{aligned}
$$

Concerning the estimator of the intercept parameter notice that, by the law of iterated expectations, we have

$$E\left(\hat{\beta}_0\right) = E\left(E\left(\hat{\beta}_0|\tilde{x}\right)\right)$$

Also

$$\begin{aligned} E\left(\hat{\beta}_0|\tilde{x}\right) &= E\left(\bar{y} - \hat{\beta}_1\bar{x}|\tilde{x}\right) \\ &= E\left(\bar{y}|\tilde{x}\right) - E\left(\hat{\beta}_1\bar{x}|\tilde{x}\right) \\ &= E\left(\bar{y}|\tilde{x}\right) - E\left(\hat{\beta}_1|\tilde{x}\right)\bar{x} \end{aligned}$$

$$
\begin{aligned}
E\left(\bar{y}|\tilde{x}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} y_i|\tilde{x}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(y_i|\tilde{x}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \left(\beta_0 + \beta_1 x_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \beta_0 + \frac{1}{n}\sum_{i=1}^{n} \beta_1 x_i \\
&= \frac{1}{n}\left(\underbrace{\beta_0 + \beta_0 + \ldots + \beta_0}_{\times n}\right) + \beta_1 \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{n}{n}\beta_0 + \beta_1 \bar{x} \\
&= \beta_0 + \beta_1 \bar{x}
\end{aligned}
$$

Therefore

$$\begin{aligned} E\left(\hat{\beta}_0 | \tilde{x}\right) &= \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} \\ &= \beta_0 \end{aligned}$$

therefore

$$\begin{aligned} E\left(\hat{\beta}_0\right) &= E\left(E\left(\hat{\beta}_0 | \tilde{x}\right)\right) \\ &= E\left(\beta_0\right) \\ &= \beta_0 \end{aligned}$$

- What happens if we include variables in our specification that don't belong?
- There is no effect on our parameter estimate, and OLS remains unbiased.
- What if we exclude a variable from our specification that does belong?
- OLS will usually be biased.

Suppose that we know that the model is

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

where $E(u|x_1, x_2) = 0$ but we estimate $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$.

- As it was shown before

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1) y_i}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2}.$$

- Then **conditional on the regressors**

$$\begin{aligned} E\left(\tilde{\beta}_1\right) &= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1) x_{i2}}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2} \\ &= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2} \end{aligned}$$

as we can show that
$\sum_{i=1}^{n} (x_{i1} - \bar{x}_1) x_{i2} = \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)$.

- Thus

$$
\begin{aligned}
E\left(\tilde{\beta}_1\right) &= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)\left(x_{i2}-\bar{x}_2\right)}{\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)^2} \\
&= \beta_1 + \beta_2 \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)\left(x_{i2}-\bar{x}_2\right)}{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)^2} \\
&= \beta_1 + \beta_2 \frac{S_{x_1,x_2}}{S_{x_1}^2}.
\end{aligned}
$$

where S_{x_1,x_2} is the sample covariance between x_1 and x_2 and $S_{x_1}^2$ is the sample variance of x_1.

$$
\begin{aligned}
E\left(\tilde{\beta}_1\right) &= \beta_1 + \beta_2 \frac{S_{x_1,x_2}}{S_{x_1}^2} \\
&= \beta_1 + \beta_2 \frac{S_{x_1,x_2}}{S_{x_2} S_{x_1}} \frac{S_{x_2}}{S_{x_1}} \\
&= \beta_1 + \beta_2 Corr(x_1, x_2) \frac{S_{x_2}}{S_{x_1}}.
\end{aligned}
$$

Summary of Direction of Bias

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) < 0$
$\beta_2 > 0$	Positive Bias	Negative Bias
$\beta_2 < 0$	Negative Bias	Positive Bias

- Two cases where bias is equal to zero:
  - $\beta_2 = 0$, that is x_2 doesn't really belong in model.
  - x_1 and x_2 are uncorrelated in the sample.
- If $corr(x_2, x_1)$ and β_2 have the same sign, bias will be positive.
- If $corr(x_2, x_1)$ and β_2 have the opposite sign, bias will be negative.
- The More General Case: Technically, can only obtain the sign of the bias for the more general case if all of the included x's are uncorrelated.

The Variance-covariance matrix of the OLS estimator $(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)$ has the form:

$$\begin{bmatrix} Var\left(\hat{\beta}_0\right) & Cov(\hat{\beta}_0, \hat{\beta}_1) & ... & Cov(\hat{\beta}_0, \hat{\beta}_k) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var\left(\hat{\beta}_1\right) & ... & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(\hat{\beta}_0, \hat{\beta}_k) & Cov(\hat{\beta}_1, \hat{\beta}_k) & ... & Var\left(\hat{\beta}_k\right) \end{bmatrix}$$

- Let $\mathbf{x}$ stand for $(x_1, x_2, \ldots x_k)$.
- Assume $Var(u|\mathbf{x}) = \sigma^2$ (Homoskedasticity).
- Assuming that $Var(u|\mathbf{x}) = \sigma^2$ also implies that $Var(y|\mathbf{x}) = \sigma^2$.
- The 4 assumptions for unbiasedness, plus this homoskedasticity assumption are known as the Gauss-Markov assumptions.

Given the Gauss-Markov Assumptions

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

where the $SST_j = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$ and R_j^2 is the R^2 from the regressing x_j on all other $x's$.

Components of OLS Variances:

- The error variance: a larger σ^2 implies a larger variance for the OLS estimators.
- The total sample variation: a larger SST_j implies a smaller variance for the estimators.
- Linear relationships among the independent variables: a larger R_j^2 implies a larger variance for the estimators.