Econometrics Notes

Carlos Gustavo Salas Flores

March 2026 - May 2026

1 Simple Linear Regression Model

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

Where y is the dependent variable, x is the independent variable, and u is the error term. And β_0 our intercept (or starting point) and β_1 is our slope parameter.

So long as, the u does not depend on x i.e.

$$\frac{\Delta u}{\Delta x} = 0 \text{ or } \mathbb{E}[u|x] = 0$$

Then, it will allow us to study how a change in x affects y.

$$\frac{\Delta y}{\Delta x} = \beta_1 \tag{2}$$

1.1 Causal interpretation

- $\mathbb{E}[u|x] = 0$. This is the **zero conditional mean assumption**.
- $\mathbb{E}[u] = 0$. So long as we have β_0, the **average value of u in the population is zero**.

1.2 Population Regression Function

The conditional mean independence assumption implies that

$$\begin{aligned}\mathbb{E}[y|x] &= \mathbb{E}[\beta_0 + \beta_1 x + u|x] \\ &= \beta_0 + \beta_1 x + \mathbb{E}[u|x]\end{aligned}$$

Since $\mathbb{E}[u|x] = 0$

$$\mathbb{E}[y|x] = \beta_0 + \beta_1 x$$

Now, if we try to estimate the Linear Model using data, we have that for a sample of n datapoints. We need to make the following assumptions.

1. $\mathbb{E}[u|x] = \mathbb{E}[u] = 0$
2. $Cov(x, u) = \mathbb{E}[xu] = 0$

1 means that there is an intercept. **2** means that x and u are uncorrelated. We can re-write them as

1. $\mathbb{E}[y - \beta_0 - \beta_1 x] = 0$
2. $\mathbb{E}[x(y - \beta_0 - \beta_1 x)] = 0$

Now, let's estimate β_0 and β_1, those estimates will be $\hat{\beta_0}$ and $\hat{\beta_1}$ to differenciate them from the real values.

We can use our definitions above to find out these values.

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{\beta_0} - \hat{\beta_1}x] = 0$$

This implies that the mean of y i.e. $\bar{y}$ is equal to..

$$\bar{y} = \hat{\beta_0} + \hat{\beta_1}\bar{x}$$

Note that $\bar{x}$ is the mean of x. Hence, we can simplify this by taking $\hat{\beta_0}$ as

$$\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x} \tag{3}$$

As for $\frac{1}{n}\sum_{i=1}^{n} x[y_i - \hat{\beta_0} - \hat{\beta_1}x]$, we can take our results we just got and substitute in $\hat{\beta_0}$.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} x_i[y_i - \hat{\beta_0} - \hat{\beta_1}x_i] &= 0 \\
\frac{1}{n}\sum_{i=1}^{n} x_i[y_i - \left[\bar{y} - \hat{\beta_1}\bar{x}\right] - \hat{\beta_1}x_i] &= 0 \\
\frac{1}{n}\sum_{i=1}^{n} x_i\left[[y_i - \bar{y}] - \left[\hat{\beta_1}\bar{x} + \hat{\beta_1}x\right]\right] &= 0 \\
\frac{1}{n}\sum_{i=1}^{n} x_i\left[y_i - \bar{y}\right] = \frac{1}{n}\sum_{i=1}^{n} x_i\left[\hat{\beta_1}\bar{x} + \hat{\beta_1}x\right] & \\
= \sum_{i=1}^{n} x_i\left[\hat{\beta_1}\bar{x} + \hat{\beta_1}x\right] & \\
= \hat{\beta_1}\sum_{i=1}^{n} x_i\left[\bar{x} + x\right] &
\end{aligned}$$

We know that $\sum_{i=1}^{n} x_i \left[y_i - \bar{y}\right] = \sum_{i=1}^{n} [x_i - \bar{x}] \left[y_i - \bar{y}\right]$ and $\sum_{i=1}^{n} x_i \left[x_i - \bar{x}\right] = \sum_{i=1}^{n} [x_i - \bar{x}]^2$, hence...

$$\frac{1}{n} \sum_{i=1}^{n} x_i \left[y_i - \bar{y}\right] = \hat{\beta}_1 \sum_{i=1}^{n} x_i \left[\bar{x} + x\right]$$
$$\sum_{i=1}^{n} [x_i - \bar{x}] \left[y_i - \bar{y}\right] = \hat{\beta}_1 \sum_{i=1}^{n} [x_i - \bar{x}]^2$$
$$Cov(x, y) = \hat{\beta}_1 Var(x)$$

So, finally, we can find $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{Cov(x, y)}{Var(x)} \tag{4}$$

see, the similarities with $\beta_1 = \frac{\Delta y}{\Delta x}$.

Intuitively we can see that when $Cov(x, y) > 0$ (positive correlation), then the $\hat{\beta}_1 > 0$. Same for the negative case.

See that **this also implies that** $Var(x) > 0$.

1.3 Properties

Fitted values and residuals

1. $\hat{y_i} = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is the predicted value.
2. $\hat{u_i} = y_i - \hat{y_i}$ Error. Deviations from regression line. Residuals.
3. $\sum_1^n \hat{u_i} = 0$. Deviations from regression add up to zero.
4. $\sum_1^n x_i \hat{u_i} = Cov(x, u) = 0$.
5. $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$. Sample averages of y and x lie on regression line.

Additionally

1. $\bar{\hat{y}} = \bar{y}$
2. $Cov(y, u) = 0$.
3. $y_i = \hat{y_i} + \hat{u_i}$

1.4 Evaluation of OLS

There are several metrics to evaluate the "goodness" of our predictions. The most important one for this course being the R^2.

First, let's define SST, SSE, and SSR.

- $SST = \sum_1^n [y_i - \bar{y}]^2$. Distance from observations to mean.
- $SSE = \sum_1^n [\hat{y}_i - \bar{y}]^2$ Distance from predictions to mean.
- $SSR = \sum_1^n [\hat{y}_i - \bar{y}]^2 = \sum_1^n \hat{u}^2$ Distance between Observations and Predictions.

We can then find the following relation

$$\begin{aligned} SST &= \sum_1^n [y_i - \bar{y}]^2 \\ &= \sum_1^n [y_i - \bar{y} + \hat{y}_i - \hat{y}_i]^2 \\ &= \sum_1^n [y_i - \hat{y}_i + \hat{y}_i - \bar{y}]^2 \\ &= \sum_1^n [\hat{u}_i + (\hat{y}_i - \bar{y})]^2 \end{aligned}$$

We expand the binomial term

$$= \sum_1^n [\hat{u}_i]^2 + 2\sum_1^n [\hat{u}_i(\hat{y}_i - \bar{y})] + \sum_1^n [\hat{y}_i - \bar{y}]^2$$

Then since $2\sum_1^n [\hat{u}_i(\hat{y}_i - \bar{y})] = Cov(u, \hat{y} - \bar{y}) = 0$, then

$$\begin{aligned} &= \sum_1^n [\hat{u}_i]^2 + \sum_1^n [\hat{y}_i - \bar{y}]^2 \\ &= SSR + SST \end{aligned}$$

In short,

$$SST = SSR + SST \tag{5}$$

The R^2 score is defined as

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \tag{6}$$

Notice that $0 \leq R^2 \leq 1$. An R^2 score closer to 1 means a good fit, while a score closer to 0 means a poor fit.

1.5 Functional Form-Incorporating nonlinearities

1.5.1 Semi-logarithmic form

In the case we apply *log* to the dependent variable.

$$\log(y) = \beta_0 + \beta_1 x + u \tag{7}$$

This changes the interpretation of the regression coefficient.

$$\beta_1 = \frac{\Delta \log(y)}{\Delta x} = \frac{1}{y} \cdot \frac{\Delta y}{\Delta x} = \frac{\frac{\Delta y}{y}}{\Delta x} \tag{8}$$

For instance, here we would interpret the result as: for every change in x, we change y by $\log(y)$ as a percentage. E.g. say $\beta_1 = 0.15$, that means a change in x will increase y by a percentage of 15%.

In the case we apply *log* to the independent variable.

$$y = \beta_0 + \beta_1 \log(x) + u \tag{9}$$

Then, a percent change of x will lead to a change in y. E.g. a $\beta_1 = 0.33$ means that a 1% change in x (0.01 change in $\log(x)$) will lead to a 0.0033 change in y.

$$\beta_1 = \frac{y}{\Delta \log(x)} = \frac{y}{\frac{\Delta x}{x}} \tag{10}$$

1.5.2 Log-logarithmic form

$$\log(y) = \beta_0 + \beta_1 \log(x) + u \tag{11}$$

Here, a percentage change in x, results in a percentage change in y.

1.6 Unbiasdness of OLS

We need to make 4 assumptions about OLS

- SLR.1 Linear in parameters. In the population model, the dependent variable, y, is related to the independent variable x and u as

$$y = \beta_0 + \beta_1 x + u \tag{12}$$

- SLR.2 Random Sampling. We have a random sample of size n composed of

$$\{(x_i, y_i) : i = 1, ...n\} \tag{13}$$

- SLR.3 Sample Variation in the Explanatory Variable. The sample outcomes in x are not all the same, i.e.

$$Var(x) > 0 \tag{14}$$

- SLR.4 Zero conditional mean. Meaning, the error u has an unexpected value of zero given any value of x.

$$\mathbb{E}[u|x] = 0 \tag{15}$$

Now, we are ready to show that the estimators of OLS are unbiased. To this end, we use the fact that $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$

We have that

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{SST_x} \tag{16}$$

1.6.1 Variance and Std

$$Var(y|x) = Var(\beta_0 + \beta_1 x + u|x)$$

Then, since $Var(\beta_0) = 0$ and $Var(\beta_1 x) = 0$

$$Var(y|x) = Var(u|x)$$

The variances of the OLS estimators are then

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{17}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2}{\sum_{i}^{n}(x_i - \bar{x})^2} \tag{18}$$

And the standard deviations are

$$\sigma(\hat{\beta}_1) = sd(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{19}$$

$$\sigma(\hat{\beta}_0) = sd(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \frac{\sigma\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}{\sqrt{\sum_{i}^{n}(x_i - \bar{x})^2}} \tag{20}$$

1.6.2 Proof

$$\begin{aligned} Var(\hat{\beta}_1) &= \frac{1}{SST_x^2} Var\left(\sum_{i=1}^{n} d_i u_i\right) \\ &= \frac{1}{SST_x^2} \sum_{i=1}^{n} d_i^2 Var(u_i) \\ &= \frac{1}{SST_x^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \sigma^2 \\ &= \frac{\sigma^2}{SST_x^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{\sigma^2}{SST_x^2} SST_x \\ &= \frac{\sigma^2}{SST_x} \end{aligned}$$

In reality, however, we only have $\hat{\sigma}$. Which is given by

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i}^{n} \hat{u_i}^2 \tag{21}$$

The $n - l$ is discounting by the degree of freedom. Intuitively, if we only have two samples, then we will have no variance because our estimator will cut across both of them.

1.6.3 Standard Error

Using the results from above and our definition of standard deviations, we can build the standard errors *se*, which are the estimators of *sd*.

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SST_x}} = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \tag{22}$$

$$se(\hat{\beta}_0) = \frac{\hat{\sigma} \sum_{i=1}^{n} x_i^2}{\sqrt{SST_x}} = \frac{\hat{\sigma} \sum_{i=1}^{n} x_i^2}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \tag{23}$$

1.6.4 Regression on binary explanatory variable

Let's say that $x \in \{0, 1\}$.

Then $\mathbb{E}[y|x = 0] = \beta_1$ and $\mathbb{E}[y|x = 1] = \beta_0 + \beta_1$.

In this context, β_1 is the difference between these two groups.

$$\mathbb{E}[y|x = 1] - \mathbb{E}[y|x = 0] = \beta_1 \tag{24}$$

In difference of differences i.e., control and treatment effects.

The Average Treatment Effect (ATE) is defined as:

$$\tau_{ate} = \mathbb{E}[y(1)] - \mathbb{E}[y(0)] \tag{25}$$

In a binary variable example. Let's say $y = \beta_0 + \beta_1 x$, we say that those who did $x = 1$ have β_1 higher than those who did $x = 0$.

2 Multiple Regression Analysis: Estimation

2.0.1 Sample regression function and OLS regression

The population regression function is defined as:

$$\begin{aligned}\mathbb{E}[y|x] &= \mathbb{E}[\beta_0 + \beta_1 x_1 + \beta_2 x_2 ... + u|x] \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \mathbb{E}[u|x]\end{aligned}$$

Then, the Sample Regression Function is

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

Ressiduals

$$\hat{u} = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - ... - \hat{\beta}_p x_p \tag{26}$$

Minimize the sum of squared residuals

$$min \sum_{i}^{n} \hat{u_i}^2 \tag{27}$$

By how much does the j^{th} variable change if the j^{th} variable changes by one unit.

$$\beta_j = \frac{\Delta y}{\Delta x_j} \tag{28}$$

The multiple linear regression model manages to hold the values of other explanatory variables fixed even if, in reality, they are correlated with the explanatory variable under consideration.

For the **partial effect** interpretation, we still have to assume that unobserved factors do not change if the explanatory variables are changed.

2.1 2-dimensional case

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \tag{29}$$

To understand the **partial effect** interpretation. We can draw the following equation

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

Notice how when $\Delta x_2 = 0$, then

$$\Delta\hat{y} = \hat{\beta}_1\Delta x_1$$

and similarly, $\Delta x_1 = 0$ gives

$$\Delta\hat{y} = \hat{\beta}_2\Delta x_2$$

2.2 N-dimensional case

Let's expand this property to the N-dimensional case.

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

This leads to

$$\Delta\hat{y} = \hat{\beta}_1\Delta x_1 + \hat{\beta}_2\Delta x_2 + ... + \hat{\beta}_p\Delta x_p$$

And, similarly, we can notice that for every $j \in [1, 2, ..., p]$ and letting every other coefficient $k \in [0, 1, ..., p]$ and $\neq j$ with a $\Delta x_k = 0$. Then,

$$\Delta\hat{y} = \hat{\beta}_j\Delta x_j \tag{30}$$

Therefore, the covariance between each independent variable and OLS residual is 0; consequently,

$$\sum_i^n x_{ij}\hat{u}_i = 0$$

2.3 Partialling out interpretation of multidimensional linear regression

Based off Eq. 30, we can show that the estimated coefficient of an explanatory variable in a multiple regression can be obtained in two steps:

1. **Regress the explanatory variable** on all other explanatory variables.
2. Regress y on the residuals from the regression.

Let's consider the case where $y \in \mathbb{R}^2$.
Consider again the case with $k = 2$ independent variables, let's focus on $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_i^n \hat{r}_{i1} y_i}{\sum_i^n \hat{r}_{i1}^2}$$

Where $\hat{r}_{i1}$ are the OLS residuals from a simple regression of x_1 on x_2. We can say that $\hat{r}_{i1}$ is the part of x_{i1} that is correlated with x_{i2}, or that it i x_{i1} after the effects of x_{i2} have been *partialled out*.

2.4 Comparison of simple and multiple regression estimates

- Simple Regression: $\tilde{y} = \tilde{\beta_0} + \tilde{\beta_1} x_1$
- Multiple Regression: $\hat{y} = \hat{\beta_0} + \hat{\beta_1} x_1 + \hat{\beta_2} x_2$
- Relationship: $\tilde{\beta_1} = \hat{\beta_1} + \hat{\beta_2}\tilde{\delta_1}$

The relationship between $\tilde{\beta_1}$ and $\hat{\beta_1}$, also shows two distinct cases where they are equal.

1. The partial effect of x_2 on $\hat{y}$ is zero on the sample. That is, $\beta_2 = 0$.
2. x_1 and x_2 are uncorrelated in the sample. That is, $\hat{\delta_1} = 0$

Goodness of Fit works the same for Multilinear Regression than it does for Linear Regression.

An alternative interpretation of the R^2 is the squared correlation between the actual and the predicted value of the dependent variable.

$$R^2 = \frac{(\sum^n (y_i - \bar{y})(\hat{y} - \bar{\hat{y}}))^2}{\sum^n (y_i - \bar{y})^2 \sum^n (\hat{y} - \bar{\hat{y}})^2}$$

Note about R^2, adding additional variables to the model will always increase the value of R^2 or at least leave it unchanged.

3 Multiple Regression Estimation: Analysis

3.1 Expected value of the OLS estimators

First, we assume:

- MLR.1. **Linear relation**. The relationship between y and the explanatory variables is linear
- MLR.2. **Random sampling**. The data is a random sample drawn from the population.
- MLR.3. **No perfect collinearity**.
  - None of the independent variables are constant. i.e. for every i in the sample $Var(x_i) > 0$ i.e. collinear with intercept.
  - No exact linear relationships among the independent variables, either one-to-one or many-to-one. *Partial correlation is allowed*
- MLR.4. **Zero conditional mean**. The value of the explanatory variables must contain no information about the mean of the unobserved factors.

- If MLR.4 holds, then we are said to have *Exogenous explanatory variables*.
- **Endogeneity** is a violation of this rule. We are said to have **explanatory variables** if they are correlated with the error u.

3.2 Unbiassedness

Under assumptions MLR.1 - 4.

$$\mathbb{E}[\hat{\beta}_j] = \beta_j, j = 0, 1, ..., k$$

for any value of the population parameter β_j. In other words, the OLS estimators are unbiased estimators of the population parameters.

3.2.1 Violations

Remarks: Violations of MLR.3 (Perfect Collinearity)

- **If the sample is small**. If one variable is accidentally a multiple of another one. E.g.

  $$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u$$

  here, the expend might be an exact multiple of avg income.

- If the sum of two or more variables is always the same e.g. $share_A + share_B = 1$ or $study + work + leisure + sleep = 168$.

- Using the wrong logarithm e.g.

  $$\beta_1 \log(inc) + \beta_2 \log(inc)^2 = \beta_1 \log(inc) + \beta_2 2 \log(inc)$$

  Instead of

  $$\log(inc) + \log(inc^2)$$

Remarks: Violations of MLR.4 (Zero conditional mean)

- **Functional form mispecification**. If the functional relationship between the explained and explanatory variables is mispecified. e.g. forgetting the inc^2 term in

  $$const = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$$

  Or using $wage$ instead of $\log(wage)$ when log wage is the true relationship in the model. Our estimators will be biased.

- **Omitted variable bias**. Omitting an important factor that is correlated with x_1, x_2, ..., x_k.

Other Violations

- Reverse causality. This occurs when $x \to y$ but $y \to x$ at the same time. Since y is a function of the error term u, x also becomes a function of u. This creates a mechanical correlation between the regressor and the error term.
- Measurement error. Such errors will become part of the equation's total error.
- Sample selection. Sample is not a random sample.

3.2.2 Including Irrelevant Variables i.e. overspecifying the model

Let's say we overspecify a model

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

however, x_3 has no effect on y after x_1 and x_2 have been controlled for. Thus $\beta_3 = 0$. In terms of the conditional expectations, $\mathbb{E}[y|x_1, x_2, x_3] = \mathbb{E}[y|x_1, x_2]$.

In short, **adding irrelavant explanatory variables does not affect the biasness of the model**, but it may affect the sampling variance.

3.3 Omitting relevant variables i.e. underspacifying the model

Let's say we have this true model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{31}$$

But we estimate omitting x_2

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 \tag{32}$$

Let's say then that x_1 and x_2 are correlated. Let's say that is

$$x_2 = \delta_0 + \delta_1 x_1 + v$$

So, if we plug in $x_2 = \delta_0 + \delta_1 x_1 + v$ into Eq. 31. We get

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2(\delta_0 + \delta_1 x_1 + v) \\ &= (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1)x_1 + (\beta_2 v + u) \end{aligned}$$

That means $\mathbb{E}[\hat{\beta}_0] = \beta_0 + \beta_2\delta_0$ and $\mathbb{E}[\hat{\beta}_1] = \beta_1 + \beta_2\delta_1$.

Note that the bias is positive or negative depending on $\delta_1\beta_2$. When the bias is positive, then there is an overestimation. When it's negative is an underestimation. If $\mathbb{E}[\hat{\beta}_0]$ is closer to zero then we say it's biased towards zero.

In most cases, however, the omitted variables have more complex relations, so it's hard to know the direction of the bias.

3.4 Variance of the OLS Estimators

- MLR.5 Homoskedasticity. Says the error u has the same variance $Var(u|x_1, ..., x_n) = \sigma^2$ given any value of the explanatory variables.

$$Var(u|x_1, ..., x_n) = Var(u|x_1, ..., x_{n-1})$$

3.4.1 Sampling variances of the OLS slope estimators

Under assumptions MLR.1-5 (**Gauss-Markov Assumptions**), and conditional on the sample value of the independent variables,

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)} \tag{33}$$

Where $j = 1, 2, ..., k$, where $SST_j = \sum_{i-1}^{n}(x_{ji} - \bar{x_j})$ is the total sample variation in x_j, and R_j^2 is the R-squared from regressing x_j on all other independent variables (including an intercept).

Components of the OLS Variances:

- **The Error Variance** σ^2. A high error variance increases the sampling variance because there is more noice in the equation. Makes estimates imprecise. It does not decrease with sample size. This is obvious since $Var(\hat{\beta}_j) \propto \sigma^2$.
- **Total Sample Variation in** x_j SST_j. Higher sample variation leads to more precise estimates. It automatically increases with sample size. Hence, bigger sample size, leads to better estimates i.e. lower $Var(\beta_j)$.
- **Linear Relationships among the Id. Var.** R_j^2. The R-squared will be higher the better explanatory variable can be linearly explained by other independent variables. Hence, when the explanatory variables are nearly perfectly correlated, we get $R_j^2 \to 1$, causing $Var(\hat{\beta}_j) \to \infty$. $R_j^2 \to 0$ is preferred.

3.4.2 Multicollinearity

When two or more independent variables are highly (but not perfectly) correlated i.e. $R_j^2 \to 1$ is called **Multicollinearity**.

IMPORTANT: It is NOT a violation of MLR.3

Let's thee the following **Example**.

We analize the avg test score of a school.

$$avgscore = \beta_0 + \beta_1 teachexp + \beta_2 matexp + \beta_3 otherexp + ...$$

Here, we expect all expenditures to be high or all to be low. Hence, the sampling variance is going to be very high.

In this case it might be a better approach to sum all the expenses altogether. In other situations, we might be better off dropping some variables, but we may get omitted variable bias.

So, back to the topic.

The problem of multicollinearity is not well-defined. However, we can use the **variance inflation factor** to help us detect collinearity.

3.4.3 Variance Inflation Factors

$$VIF_j = \frac{1}{1 - R^2} \tag{34}$$

Therefore, if we substitute $1 - R^2$ in Eq. 33.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} \cdot VIF_j \tag{35}$$

This shows that VIF_j is the factor that affects sample variance the most.

Multicollinearity does not necessarily affect other variables; it only affects the ones involved in multicollinearity.

3.5 Variances in misspecified models

Sometimes we will have to make a tradeoff between bias and variance when adding a new variable.

Let's see the case of a simple linear regression vs a 2-D linear regression.

Let's say we have

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ is the true model
- $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ is the estimated model 1
- $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$ is the estimated model 2

Maybe in the case above, the likely omitted bias in model 2 is overcompensated by a smaller variance.

See that for the 2-dimensional (model 1), the variances are

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST(1 - R^2)} \tag{36}$$

While model 2 is

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{SST} \tag{37}$$

Hence, in model 2, the variance is always smaller.

Let's say now that in this example we have two cases for β_2.

- $\beta_2 = 0$ i.e. Model 1 is an overestimation.

  Then, we known from the unbiasedness property coming from assumptions MLR.1-4 $\mathbb{E}[\hat{\beta}_1] = \mathbb{E}[\tilde{\beta}_1] = \beta_1$.

  However, $Var(\hat{\beta}_1) > Var(\tilde{\beta}_1)$. The model is biased.

- $\beta_2 \neq 0$ i.e. Model 2 is an underestimation.

  Then, $\mathbb{E}[\hat{\beta}_1] = \beta_1$ but $\mathbb{E}[\tilde{\beta}_1] \neq \beta_1$. The bias is still there, though.

  But still, $Var(\hat{\beta}_1) > Var(\tilde{\beta}_1)$.

However, in the practice we don't know σ^2, so we get an estimator of it $\hat{\sigma}^2$. Since $\sigma^2 = \mathbb{E}[u^2]$, the biased estimator is given by $\frac{1}{n}\sum \hat{u}^2$, but in order to obtain an unbiased estimator we must divide by the degree of freedom $(n-k-1)$ instead of n.

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

Where $k+1$ is the number of parameters including the intercept.
Under the **Gauss-Markov Assumptions** assumptions MLR.1-5.

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2 \tag{38}$$

Now that we have the sample variance $Var(\hat{\beta}_j)$, now we can get the standard deviation of $\hat{\beta}_j$, which is nothing but

$$\sqrt{Var(\hat{\beta}_j)} = \frac{\sigma}{\sqrt{SST_j(1-R_j^2)}} \tag{39}$$

But again, we don't know σ, so we replace it with the estimator so we get the **standard error** of β_j

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}} \tag{40}$$

Remember that $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$.

3.6 Gauss-Markov Theorem

Under assumptions MLR.1-4, OLS is unbiased. But, there are other unbiased estimators besides OLS. Are there any better estimators then? No. If we look at the linear estimators $\tilde{\beta}_j = \sum_{i=1}^{n} w_{ji} y_i$.

The **Gauss-Markov Theorem** tells us that under assumptions **MLR.1-5**, the best OLS estimator of β_j is $\hat{\beta}_j$. This is called the **Best Linear Unbiased Estimator (BLUE)**.

However, if any of them do not hold e.g., if there is Heteroskedasticity, then there are better estimators.

4 Multiple Regression Analysis: Inference

4.1 Sampling Distributions of OLS Estimators

To perform statistical inference, we need to know the full sampling distribution of $\hat{\beta}_j$. Of course, even under the **Gauss-Markov Assumptions**, the distribution of $\hat{\beta}_j$ can have virtually any shape but we know it depends on the distribution of the errors. We assume that the unobserved error is **normally distributed**.

Under assumptions MLR.1 - MLR.6, the **Classical Linear Model (CLM) assumptions**. The normality assumption is then that the error u is independent of all explanatory variables and is distributed with zero mean and variance σ^2

$$u \sim N(0, \sigma^2) \tag{41}$$

MLR.6 is especially important because it says that $\mathbb{E}[u|x_1, ..., x_k] = \mathbb{E}[u] = 0$, therefore it follows that $Var(u|x_1, ..., x_k) = Var(u) = \sigma^2$.

Under CLM assumptions, we can show that the OLS estimators are the **minimum variance unbiased estimators**, meaning OLS has the smallest variance among unbiased estimators.

Succinctly,

$$y|\boldsymbol{x} = N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k, \sigma^2) \tag{42}$$

We invoke the CTL to make this affirmation; however, it has its problems. One clear violation of the MLR.6 is when values are coming from an Exponential distribution for instance. However, if our sample is large enough, this should not be problem.

4.2 Normal Sampling Distributions

Under the CLM assumptions, MLR.1-6, conditional on the sample values of the independent variables.

$$\hat{\beta}_j \sim N(\beta_j, Var(\hat{\beta}_j)) \tag{43}$$

Therefore, it follows that

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0, 1) \tag{44}$$

4.3 T-test: Hypothesis testing about a single parameter

Let's say we have a linear model that satisfies all CLM assumptions. We know that OLS produces unbiased estimators of the β_j. While we cannot know for certainty, we can *hypothesize* about the value of β_j and test the hypothesis.

$$\frac{\hat{\beta}_j - \beta_J}{se(\beta_j)} \sim t_{n-k-1} = t_{df} \tag{45}$$

In most applications, our primary interest lies in the **null hypothesis**.

$$\boldsymbol{H}_0 : \beta_j = 0 \tag{46}$$

To test the null hypothesis against any other hypothesis, we run the **t statistic** or **t ratio** as

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{47}$$

4.4 One Sided Alternative Hypothesis

Let's consider the alternative to the **null hypothesis** to be

$$\boldsymbol{H}_1 : \beta_j > 0 \tag{48}$$

Then, what we are really saying is that $\boldsymbol{H_0} : \beta_j \leq 0$

We reject a null hypothesis at a **significance level** e.g., 5% level. We say we reject the null hypothesis at a 5% significance level if

$$t_{\hat{\beta}_j} > c_{0.05} \tag{49}$$

$c_{0.05}$ is the critical value.

The critical value depends on the direction of the test and the number of samples. Smaller samples require higher scores to pass the critical value.

4.5 Two-Sided Hypothesis

This is our case when $\boldsymbol{H_0} : \beta_j = 0$, so our two-sided alternative is nothing but

$$\boldsymbol{H_1} : \beta_j \neq 0 \tag{50}$$

Under this H_1, there is a ceteris paribus effect on y.

4.6 Testing other hypothesis about β_j

Although generally, we are interested in testing $\beta_j = 0$, sometimes we may be interested in testing $\beta_j = a$ e.g. 1. So we change our hypothesis testing to

$$\boldsymbol{H_0} : \beta_j = a_j \tag{51}$$

Where a_j is our hypothesized value of β_j, then the appropriate t-statistic is

$$t = \frac{\hat{\beta}_j - a}{se(\hat{\beta}_j)} \tag{52}$$

4.6.1 Summary of Hypothesis testing procedure

- Stating the alternative hypothesis
- Choose a significance level

Example

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_2 ros + x \tag{53}$$

Null is: After controlling for Sales and roe, ros has no effect on salary.
$H_0 : ros = 0$; $H_1 : ros > 0$

$$\log(salary) = 4.32(0.32) + 0.280 \log(sales)(0.035) + 0.174 roe(0.0041) + 0.00024 ros(0.00054) \tag{54}$$

$$t - score = \frac{0.00024}{0.00054} = 0.44 < c_{0.1} \tag{55}$$

4.7 P-value

The smallest significance level at which the null hypothesis would be rejected is called the **p-value**. A **p-value** $\in [0, 1]$ is a probability. For a one-sided test, just divide the two-sided p-value by 2.

5 Questions

- Is SST on x or y?.
- Homoskedasticity in Simple Linear Regression vs Multiple Linear Regression