Econ 103: Introduction to Econometrics
Week 8

Francisco Díaz-Valdés*

Fall 2025

1 Measuring Goodness-of-Fit

This analysis can also be done using the Multiple Regression Framework. For simplicity, consider the simple linear regression model and that assumptions SR1-SR6 hold.

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

How much does the variation in x_i explain the variation of y_i? Let's separate y_i into two components: explainable and unexplainable. That is,

$$y_i = \underbrace{\mathbb{E}[y_i \mid \boldsymbol{x}]}_{\text{explainable}} + \underbrace{\varepsilon_i}_{\text{unexplainable}}$$

where $\mathbb{E}[y_i \mid \boldsymbol{x}]$ is the explainable or systematic component of y_i and ε_i is the random, unsystematic, and unexplainable component of y_i. While both components are unobservable to us, we can estimate the unknown parameters β_1 and β_2, and the residual of the estimation $\hat{\varepsilon}_i$. Once the model is estimated, we can write:

$$y_i = \hat{y}_i + \hat{\varepsilon}_i$$

where $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i = \hat{\mathbb{E}}[y_i \mid \boldsymbol{x}]$ and $\hat{\varepsilon}_i = y_i - \hat{y}_i$. Following the decomposition:

$$\begin{aligned}
& y_i = \hat{y}_i + \hat{\varepsilon}_i \quad / -\overline{y} \\
\Leftrightarrow \quad & y_i - \overline{y} = \hat{y}_i - \overline{y} + \hat{\varepsilon}_i \quad /()^2 \\
\Leftrightarrow \quad & (y_i - \overline{y})^2 = (\hat{y}_i - \overline{y})^2 + 2(\hat{y}_i - \overline{y})\hat{\varepsilon}_i + \hat{\varepsilon}_i^2 \quad / \sum_{i=1}^{N} \\
\Leftrightarrow \quad & \sum_{i=1}^{N}(y_i - \overline{y})^2 = \sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{N}\hat{\varepsilon}_i^2 + 2\sum_{i=1}^{N}(\hat{y}_i - \overline{y})\hat{\varepsilon}_i
\end{aligned}$$

Before continuing with the decomposition, it is worth noticing that $2\sum_{i=1}^{N}(\hat{y}_i - \overline{y})\hat{\varepsilon}_i = 0$ (why?[1])

Continuing the work with the decomposition:

$$\underbrace{\sum_{i=1}^{N}(y_i - \overline{y})^2}_{=\text{SST}} = \underbrace{\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2}_{=\text{SSR}} + \underbrace{\sum_{i=1}^{N}\hat{\varepsilon}_i^2}_{=\text{SSE}}$$

*Many thanks to all previous TAs for providing the notes. All mistakes are my own. Please contact me at fdiaz-valdes@g.ucla.edu if you spot any typos or mistakes. This TA is based on your textbook Principles of Econometrics.

[1] **Hint**: Use the two First-Order conditions when the model has an intercept.

Where,

(a) SST: Total sum of squares. A measure of the total variation of y.

(b) SSR: Sum of squares due to the regression. Part of the total variation of y that is "explained" by the regression.

(c) SSE: Sum of squares due to error. Part of the total variation of y that the regression cannot explain

For convenience, we can write the decomposition as follows: SST = SSR + SSE.

Define R^2 as the fraction of the total variation of y that is "explained" by the variation of x within the regression model. Some people call R^2 as the **coefficient of determination**. Formally,

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

Note that $0 \leq R^2 \leq 1$. The closer the R^2 is to 1, the closer the sample values of y_i will lie on the fitted regression.

Caution: The R^2 remains the same or increases when we add more regressors to the model. That is, generally speaking, the more regressors the model has, the higher the R^2. This renders the R^2 "useless" as a measure of how good a model is, as the R^2 can be arbitrarily increased. For this reason, Mordecai Ezekiel proposed the adjusted R^2. The intuition of the adjusted R^2 is that it attempts to compensate for the increase in the R^2 due to adding regressors by penalizing it. Formally, denoting by $\overline{R}^2$ (or by R^2_{adj}) is given by:

$$\overline{R}^2 = 1 - (1 - R^2)\frac{N-1}{N-K}$$

Where N is the number of observations, and K is the number of regressors (the number of βs to be estimated). It can be shown that $\overline{R}^2$ may be negative if too many insignificant regressors are added.

Exercises:

1. In the simple regression model, show that the squared of the sample correlation between x and y, $\hat{\rho}(x, y)^2$, is equal to R^2.

2. Suppose you run a regression with only a constant, without any regressors. Does it make sense to try to compute the R^2? What would be its value?

3. Another definition for the adjusted R^2 is:

$$\overline{R}^2 = 1 - \frac{(SSE)/(N-K)}{(SST)/(N-1)}$$

   Show that this definition is equal to the one given before.

2 Multiple Regression Framework

The Multiple Regression Framework extends the Simple Regression Model by allowing the incorporation of more regressors. Generally speaking, a Multiple Regression Model with k regressors is given by (note that the constant is also a regressor: $x_{i1} = 1$ for all i):

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

We will make the following assumptions:

- **MR1: Econometric model**:
  We have observations $(y_i, x_i1, x_i2, \ldots, x_ik)$ for every individual $i = 1, \ldots, N$, where $x_{i1} = 1 \ \forall i$, and these observations satisfy

  $$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

- **MR2: Strict Exogeneity**:
  We will assume that $\mathbb{E}[\varepsilon_i \mid x_1, \ldots, x_k] = 0$, this implies that $\mathbb{E}[\varepsilon_i] = 0$ and $\text{cov}(\varepsilon_i, x_{jh}) = 0$, for $h = 1, \ldots, k$, and $(i, j) = 1, \ldots, N$.

  This assumption allows us to write:

  $$\mathbb{E}[y_i \mid (x_1, \ldots, x_k)] = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik}$$

  Thus, we can obtain the marginal effects, just as in the case with the simple regression model:

  $$\frac{\partial \mathbb{E}[y_i \mid (x_1, \ldots, x_k)]}{\partial x_j} = \beta_j$$

  When taking the partial derivative, we are holding $x_{-j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k)$ constant.

- **MR3: Conditional Homoskedasticity**:
  We assume that $\mathbb{V}(\varepsilon_i \mid (x_{i1}, \ldots, x_{ik})) = \sigma^2$ for every $i = 1, \ldots, N$.

- **MR4: Conditional Uncorrelated Errors**:
  We assume that $\text{cov}(\varepsilon_i, \varepsilon_j \mid x_1, \ldots, x_k) = 0$ for every $i \neq j$. If $i = j$, we get that

  $$\text{cov}(\varepsilon_i, \varepsilon_i \mid x_1, \ldots, x_k) = \mathbb{V}(\varepsilon_i \mid (x_{i1}, \ldots, x_{ik})) = \sigma^2$$

- **MR5: No Perfect Multicollinearity**:
  This means that there are no such constants $(c_1, \ldots, c_k)$, all different from zero, such that,

  $$c_1 x_{i1} + c_2 x_{i2} + \ldots + c_k x_{ik} = 0, \text{for all } i$$

  Intuitively, there is no redundant information. If one regressor can be constructed using the others, then it provides no new information and has to be removed from the model. A classic example of perfect multicollinearity problem arising is when you run a regression with a constant, a dummy variable and the complement of the dummy variable. Note that we can have multiple dummy variables for different categories, the problem arises when you add the complement of a dummy variable.

  For example, the following model will present perfect multicollinearity:

  $$y_i = \beta_1 + \beta_2 D_i + \beta_3 D_i^c + \varepsilon_i$$

where,

$$D_i = \begin{cases} 1 & \text{if} \quad i \in A \\ 0 & \text{if} \quad i \notin A \end{cases}$$

and,

$$D_i^c = \begin{cases} 0 & \text{if} \quad i \in A \\ 1 & \text{if} \quad i \notin A \end{cases}$$

- **MR6 (Optional): Error Normality**:
We assume that conditional on the regressors:

$$\varepsilon \mid (x_1, \ldots, x_k) \sim \mathcal{N}(0, \sigma^2)$$

Thus,

$$y \mid (x_1, \ldots, x_k) \sim \mathcal{N}\Big(\mathbb{E}[\, y \mid (x_1, \ldots, x_k)], \sigma^2\Big)$$

2.1 How do we estimate the Multiple Regression Model?

Same as in the simple regression model, we minimize this loss function:

$$\min_{\beta_1,\ldots,\beta_k} \mathcal{L}(\beta_1, \ldots, \beta_k) = \min_{\beta_1,\ldots,\beta_k} \sum_{i=1}^{N} \Big(y_i - (\beta_1 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik})\Big)^2$$

Thus, we have k equations (the partial derivatives of $\mathcal{L}$ for each β) and k unknowns.

Doing the minimization by hand and using the summation notation is time consuming. Instead, what the computer does is the following. Define,

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}, \quad \mathbf{X} = \begin{bmatrix} x_{11} & \ldots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{N1} & \ldots & x_{Nk} \end{bmatrix}_{N\times k}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}_{N\times 1}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}_{N\times 1}$$

Then, we write the model as

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

And then, we minimize

$$\min_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

After taking the first-order condition and some algebra, we get that the solution is given by

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$$

If Linear Algebra is not a requisite for taking this class, almost surely you will not be asked to do this. However, keep in mind that this is what the computer is doing when you ran a regression.

Also, this illustrate why perfect multicollinearity is a problem. When perfect multicollinearity is present, the matrix $\mathbf{X}^T\mathbf{X}$ does not have an inverse, so we cannot compute $\boldsymbol{\beta}$.