Econometria 2
Matrix Algebra for OLS Estimator
Aula 1

Monitor: Carlos Nathaniel Rocha Cavalcante
Professor: Raphael Bottura Corbi

Universidade de São Paulo

Abril/2021

Dependent Variables

- Suppose the sample consists of n observations.
- The dependent variable is denoted as an $n \times 1$ (column) vector:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

- The subscript indexes the observation.

Independent Variables

- Suppose there are k independent variables and a constant term. In the spreadsheet there are $k+1$ columns and n rows.

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

- Mathematically that spreadsheet corresponds to an $n \times (k+1)$ matrix, denoted by $\mathbf{X}$:

Linear Regression Model

- In matrix terms, we have:

$$Y_{n\times 1} = X_{n\times(k+1)}\beta_{(k+1)\times 1} + \varepsilon_{n\times 1}$$

- In matrix terms, we have the estimate equation:

$$\hat{Y} = \hat{X}\beta$$

- The residual is:

$$\varepsilon = Y - \hat{Y} = Y - X\hat{\beta}$$

Least Squares Regression

- The least squares coefficient vector minimizes the sum of squared residuals:

$$Minimize_{\hat{\beta}} \;\; S(\beta) = \varepsilon'\varepsilon = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$\varepsilon'\varepsilon = Y'Y - \hat{\beta}X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\beta$$

$$S(\hat{\beta}) = Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2X'Y + 2XX\hat{\beta} = 0$$

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

- We could have obtained the same result as the orthogonality condition.

OLS FINITE SAMPLE PROPERTIES

- HYPOTHESIS. 1: LINEAR IN PARAMETERS.
- HYPOTHESIS. 2: NO PERFECT COLINEARITY. Matrix X has maximum rank.
- HYPOTHESIS. 3: HOMOSCEDASTICITY AND ABSENCE OF SERIAL CORRELATION.
- HYPOTHESIS. 4: ZERO CONDITIONAL AVERAGE $E(\varepsilon|X) = 0$;.

Proof: NO OLS BIAS

- Note that

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

$$E[\hat{\beta}] = E[(X^T X)^{-1} X^T Y]$$

$$E[\hat{\beta}] = E[(X^T X)^{-1} X^T (X\beta + \varepsilon)]$$

$$E[\hat{\beta}] = E[(X^T X)^{-1} X^T X\beta] + E[(X^T X)^{-1} X'\varepsilon]$$

- Because $E[X'\varepsilon] = 0$ and $E[(X^T X)^{-1} X^T X] = I$, we have:

$$E[\hat{\beta}] = \beta$$

The Variance-Covariance Matrix of the OLS Estimates

- Note that

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

$$Var(\hat{\beta}|X) = (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}$$

$$Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1} = \hat{S}^2$$

- Where we estimate σ^2 with $\hat{\sigma}^2$:

$$\hat{\sigma^2} = \frac{\varepsilon'\varepsilon}{n-k}$$

Other possibility to see the Variance-Covariance Matrix of the OLS Estimates

- We can derive the variance-covariance matrix of the OLS estimator, $\hat{\beta}$:

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = E[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)']$$
$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = E[((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}]$$

- **TRANSPOSITION PROPERTY**: (AB)' = B'A'

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = ((X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1}$$
$$Var(\hat{\beta}|X) = (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}$$
$$Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$
$$Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$$

  - Where We estimate σ^2 with $\hat{\sigma}^2$:
$$\hat{\sigma}^2 = \frac{\varepsilon'\varepsilon}{n-k}$$

Estimation of σ^2

- We know that :

$$\hat{\sigma}^2 = \frac{\varepsilon'\varepsilon}{n-k}$$

- Then:

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})(Y - X\hat{\beta})'}{n-k} =$$

What does the variance-covariance matrix of the OLS estimator look like?

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = \begin{bmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{var}(\hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \vdots & \vdots \\ \text{cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{var}(\hat{\beta}_k) \end{bmatrix}$$

R^2 QUALITY OF ADJUSTMENT

- Measures the degree of adjustment of the model to the data;

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

- In matrix:

$$y^T y = \hat{\varepsilon}^t \hat{\varepsilon} + \hat{y}^T \hat{y}$$

- Where:

$$y_{n\times 1} = Y - \bar{Y} \;\; \hat{y}_{n\times 1} = \hat{Y} - \bar{Y} \;\; \varepsilon_{n\times 1} = Y - \hat{Y}$$

R^2 QUALITY OF ADJUSTMENT

- Degree of adjustment:

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

- We know that in summatorial language:

$$SSE = \sum(\hat{Y}_i - \bar{Y})^2$$

$$SSR = \sum(Y_I - \hat{Y})^2$$

$$SST = \sum(Y_I - \bar{Y})^2$$

R^2

- Degrees of Freedom:

$$df_{SSE} = k - 1$$

$$df_{SSR} = n - k$$

$$df_{SST} = n - 1$$

R^2 Matrix

- So, we will show R^2. We know that:

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

- We know that in matrix language:

$$SSE = \hat{\beta}^T X^T Y - n\bar{Y}^2$$

$$SSR = Y^t Y - \hat{\beta}^T X^T Y$$

$$SST = Y^T Y - n\bar{Y}^2$$

Properties of R^2

1. $R^2 \in [0,1]$;
2. Good adjustment $\rightarrow R^2 \approx 1$;
3. R^2 always tends to increase with new explanatory variables;
4. R^2 never experimental with new explanatory variables.

R^2 *Adjustment*

- Corrects limitation of the degree of adjustment:

$$\bar{R}^2 = 1 - \frac{\hat{\varepsilon}^T \varepsilon (n-1)}{y^T y (n-k)}$$

- $\bar{R}^2 = R^2$ se $k = 1$;
- $\bar{R}^2 < R^2$ se $K > 1$;
- $\bar{R}^2$ it can decrease if I include variables that are not very explanatory;
- $\bar{R}^2$ can be negative.

T-test and confidence interval

- In matrix terms:

$$Y_{n\times 1} = X_{n\times(k+1)}\beta_{(k+1)\times 1} + \varepsilon_{n\times 1} \quad where \quad \varepsilon_{n\times 1} \sim N\left(0, \sum\right)$$

- From the Variance-Covariance Matrix, the main diagonal corresponding to:

$$SE(\hat{\beta}_{k\times 1}) = diag(\sqrt{\hat{S}^2_{\hat{\beta}}})$$

T-test and confidence interval

- How to calculate the confidence interval?

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - \beta_i}{SE(\hat{\beta}_i)} \quad for \;\; i = \{1, 2, 3...\}$$

- in a compact way, the confidence interval for any β is:

$$\hat{\beta}_i \pm t_{(\frac{\alpha}{2};n-2)} \times SE(\hat{\beta}_i)$$