Universidad de San Andrés – Departamento de Economía – Econometría - Semestre de otoño

MCO en el modelo lineal con 2 variables

Mariana Marchionni

mariana@depeco.econo.unlp.edu.ar

El modelo lineal con 2 variables

$$Y_i = \alpha + \beta X_i + u_i \qquad\qquad i = 1,\ldots,N$$

- α y β son constantes desconocidas que describen cómo es la relación entre *Y* y *X*
- u_i es un término aleatorio no observable, que representa el hecho de que la relación entre *Y* y X no es exacta (determinística)

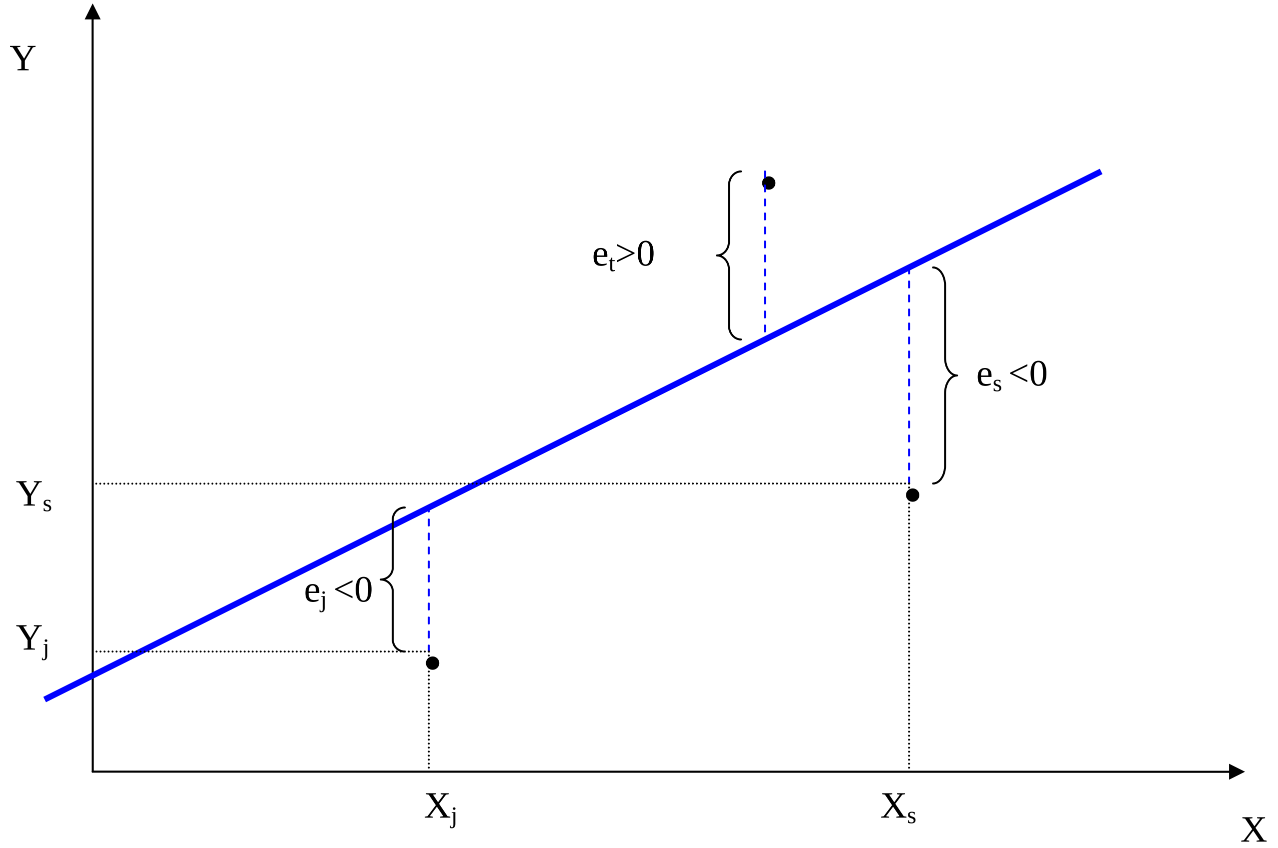
Estimación de los parámetros

- Necesitamos estimar α y β a partir de la muestra disponible
- La muestra consiste en un conjunto de observaciones (Y_i , X_i) con i=1,…N
- Denotemos como $\hat{\alpha}$ y $\hat{\beta}$ a los estimadores de α y β respectivamente
- Definamos algunas magnitudes:

$$\hat{Y}_i \equiv \hat{\alpha} + \hat{\beta} X_i \qquad \rightarrow \text{Estimación de } Y_i$$

$$e_i \equiv Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i) \qquad \rightarrow \text{Error de estimación o residuo}$$

- Cada posible valor que asignemos a $\hat{\alpha}$ y $\hat{\beta}$ define una recta en el plano (*Y,X*)



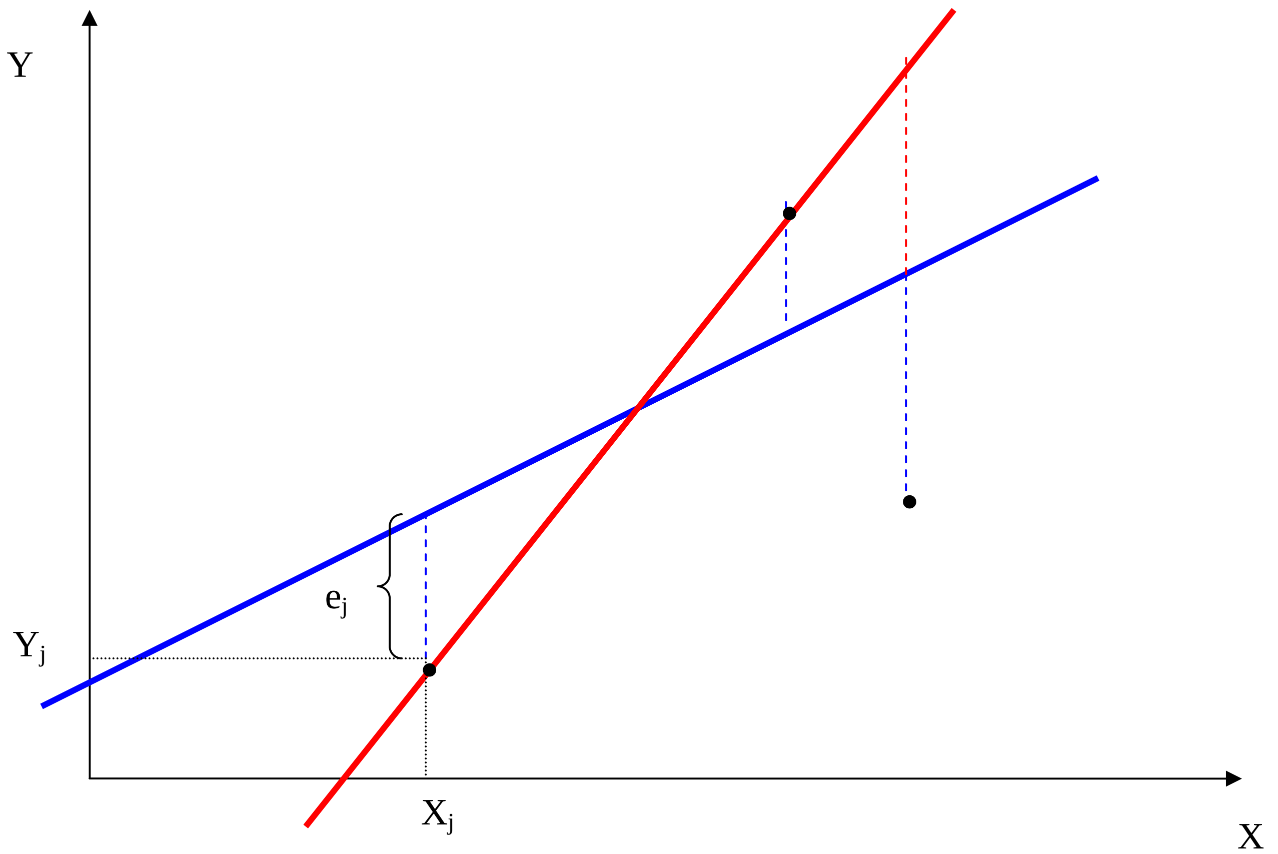
Parece natural buscar estimadores que cometan “pocos” errores.

Criterio propuesto: elegir $\hat{\alpha}$ y $\hat{\beta}$ que minimicen:

$$SCR(\hat{\alpha},\hat{\beta}) = \sum_{i=1}^{n} e_i^{\ 2} = \sum_{i=1}^{n} \left[Y_i - (\hat{\alpha} + \hat{\beta} X_i) \right]^2$$

SCR es la **S**uma de los **C**uadrados de los **R**esiduos

Gráficamente: la recta azul produce una menor SCR que la roja.



Formalmente, el problema planteado es:

$$\underset{\hat{\alpha},\hat{\beta}}{Minimizar} \quad SCR(\hat{\alpha},\hat{\beta}) = \sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n}\left[Y_i - (\hat{\alpha} + \hat{\beta} X_i)\right]^2$$

Condiciones de Primer Orden:

1. $$\frac{\partial SCR(\hat{\alpha},\hat{\beta})}{\partial \hat{\alpha}} = -2\sum_{i=1}^{n}\left[Y_i - (\hat{\alpha} + \hat{\beta} X_i)\right] = -2\sum_{i=1}^{n} e_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} e_i = 0$$

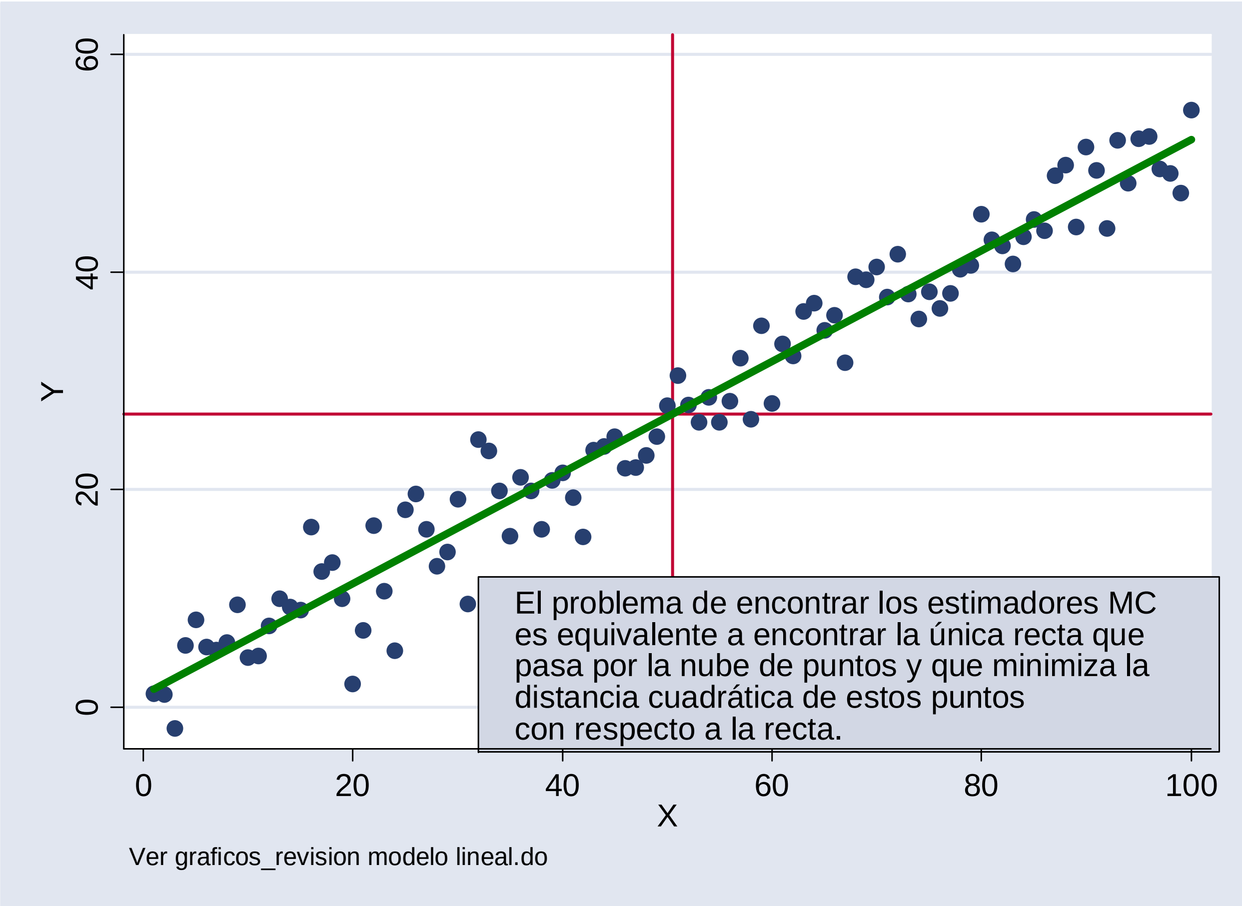
2. $$\frac{\partial SCR(\hat{\alpha},\hat{\beta})}{\partial \hat{\beta}} = -2\sum_{i=1}^{n} X_i\left[Y_i - (\hat{\alpha} + \hat{\beta} X_i)\right] = -2\sum_{i=1}^{n} X_i e_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i e_i = 0$$

Las anteriores CPO (1) y (2) constituyen un sistema de **dos ecuaciones lineales con dos incógnitas**. Puede mostrarse que la solución a este problema viene dada por:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{YX}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \qquad \text{y} \qquad \hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{X}$$

y se los denomina ***estimadores mínimo cuadráticos* (MC)** de α y β.

Interpretación geométrica de los estimadores MCO



Bondad del ajuste o cuán bien se ajusta a los datos el modelo estimado

Resultado importante: si el método de estimación que se usa es el de MC se cumple que

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

SCT = SCE + SCR

Intuitivamente: la variabilidad total de la variable explicada alrededor de su media muestral (la SCT) se puede descomponer en dos partes. Una atribuible al modelo (SCE) y otra a la variabilidad de los errores (SCR). Una medida de la bondad del ajuste viene dada por $R^2 \equiv \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$

R^2 = 1 sólo cuando SCR=0, indicando un ajuste perfecto del modelo a los datos.

Distintos ajustes