

# **Tema 3.1: Modelo lineal general: hipótesis y estimación**

Universidad Complutense de  
Madrid

2013

# Introducción

El **objetivo** es especificar y estimar un Modelo Lineal General (MLG) en donde una variable de interés (endógena) es explicada por un conjunto de variables explicativas (exógenas).

La **relación de causalidad** entre estas variables es unidireccional. Es decir, las variables exógenas pueden influir en la endógena, pero no a la inversa.

Antes de poder estimar los parámetros que caracterizan este modelo (es decir, medir el efecto que tiene cada explicativa sobre la variable endógena) es necesario definir un conjunto de supuestos llamados **hipótesis del MLG**.

# Hipótesis del modelo de regresión lineal general (I)

**H1: Linealidad en los parámetros:** Establece la linealidad en los parámetros en la relación entre la variable endógena (a explicar) y las exógenas (explicativas). Por ejemplo, en la relación entre Consumo y Renta:

$$C_t = \beta_0 + \beta_1 R_t + \varepsilon_t$$

Los parámetros que definen esta relación son  $\beta_0$  y  $\beta_1$   
No hay que confundir esta hipótesis de linealidad en los parámetros con una relación lineal entre las variables.

# Hipótesis del modelo de regresión lineal general (II)

Por ejemplo, en las relaciones siguientes, sólo la primera es formalmente lineal. Sin embargo, las tres **cumplen la hipótesis de linealidad en los parámetros**:

$$y = \beta_1 + \beta_2 x$$

$$y = \beta_1 + \beta_2 e^x$$

$$y = \beta_1 + \beta_2 \ln x$$

donde  $\ln x$  denota el logaritmo neperiano de  $x$ . A veces, la transformación logarítmica, linealiza el modelo (función de producción Cobb-Douglas).

# Hipótesis del modelo de regresión lineal general (III)

La **función original de Cobb-Douglas** no es lineal ni en las variables ni en los parámetros:

$$Y = AK^{\alpha}L^{\beta}$$

donde  $Y$  es el nivel de producción,  $K$  el stock de capital y  $L$  el trabajo. Si **transformamos logarítmicamente esta relación**, se cumple la hipótesis de linealidad en los parámetros. Es decir:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L$$

Los parámetros son  $A$  (coeficiente de eficiencia),  $\alpha$  es la elasticidad del output-capital y  $\beta$  la elasticidad del output-trabajo.

# Hipótesis del modelo de regresión lineal general (IV)

Un modelo en donde **no se cumple la hipótesis de linealidad en los parámetros** sería, por ejemplo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_1^2 X_2$$

en donde el coeficiente asociado a la variable  $X_2$  no es  $\beta_2$ , sino  $\beta_1^2$ , es decir, el cuadrado del coeficiente asociado a la variable  $X_1$

En este caso, la relación es lineal en las variables, pero no en los parámetros. La transformación logarítmica del modelo no resuelve el problema.

# Hipótesis del modelo de regresión lineal general (V)

**H2: Especificación correcta.** Esta hipótesis supone que las variables explicativas “x” del modelo son aquellas variables relevantes que explican a la endógena, “y”. Y que están todas. No existe ninguna variable “x” que no explique nada de la “y”. El modelo está bien planteado o especificado.

Esto supone admitir que siempre existe una teoría económica que nos dicta los todos y cada uno de los factores que explican a otra variable.

También supone que siempre disponemos de datos coherentes de todas las variables explicativas a introducir.

# Hipótesis del modelo de regresión lineal general (VI)

El incumplimiento de esta hipótesis se da en muchos casos.

**Ejemplo:** Si uno quiere estimar con datos de sección cruzada una función de consumo keynesiana, además de la renta familiar, existen otras muchas variables que explican el comportamiento del consumo de una familia.

Por ejemplo, el número de hijos de la familia, si la mujer trabaja o no, si se vive en el campo o la ciudad, si se tiene hipoteca o no, etc. Normalmente, no es posible tener información muestral sobre todas estas variables.



# Hipótesis del modelo de regresión lineal general (VII)

**H3: Grados de libertad positivos:** Los grados de libertad ( $gl$ ) de un modelo se definen como la diferencia entre el número de datos ( $n$ ) y el número de variables explicativas ( $k$ ). Es decir,  $gl = n - k \geq 0$

Esta hipótesis supone que, como mínimo, es necesario disponer de tantos datos como parámetros a estimar. No obstante, es preferible disponer de más datos que parámetros a estimar. En el ejemplo de la función de consumo keynesiana hay que estimar dos parámetros. Con un único dato, no sería posible estimar de forma única ambos parámetros. ¿Por qué? Con dos datos, se tiene una estimación única de ambos parámetros.

# Hipótesis del modelo de regresión lineal general (VIII)

**H4: Parámetros constantes.** Esta hipótesis supone que los parámetros no cambian con el tiempo (con datos temporales) o con el individuo (en secciones cruzadas). Si trabajamos con  $n$  datos en la función de consumo keynesiana, suponer que la propensión marginal a consumir es constante en el tiempo, implica que se obtiene una estimación que ha de interpretarse como la propensión marginal a consumir media en ese período de tiempo considerado. Si el período muestral con el que se trabaja es muy amplio y heterogéneo (por ejemplo, incluye períodos de crisis y de auge), es más difícil mantener esta hipótesis que si la muestra es homogénea.

# Hipótesis del modelo de regresión lineal general (IX)

## H5: Independencia lineal entre las variables explicativas

Esta hipótesis implica que cada variable explicativa contiene información adicional sobre la endógena que no está contenida en otras. Si hubiera información repetida, habría variables explicativas dependientes linealmente de otras. Se puede resumir la información muestral sobre las variables explicativas en una matriz:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdot & x_{1k} \\ x_{21} & x_{22} & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdot & x_{nk} \end{bmatrix}$$

La primera columna contiene información de la 1ª variable explicativa, la primera fila contiene la 1ª observación de todas las  $x$ 's del modelo.

# Hipótesis del modelo de regresión lineal general (X)

La hipótesis de independencia lineal entre los regresores matemáticamente implica que cada columna de la matriz  $X$  es linealmente independiente del resto.

Encontrar una dependencia lineal en las columnas de la matriz  $X$  supone que **alguna variable explicativa es redundante**. Por ejemplo, si se cumple que  $x_{t1} = 3x_{tk}, \forall t$ , la primera y la última variable contienen la misma información, salvo por un cambio de escala. Si ocurre esto, la matriz  $X^T X$  no tiene inversa.

# Hipótesis del modelo de regresión lineal general (XI)

**H6: Regresores no estocásticos.** Esta hipótesis implica que los datos de las variables explicativas son **fijos en muestras repetidas**. Es decir, el valor de las variables explicativas es constante en la función de distribución de la endógena.

El MLG establece que  $Y = f(X) + \varepsilon$ , donde  $f$  es una función lineal en los parámetros. Como el error es aleatorio, la función de distribución de la endógena dependerá de la función de distribución del error y del valor conocido de la matriz de datos  $X$ . Esta es una hipótesis crucial en la Econometría de este curso.

# Hipótesis del modelo de regresión lineal general (XII)

Existen **tres situaciones** en Econometría donde no es posible mantener esta hipótesis:

- (A) Modelos de ecuaciones simultáneas
- (B) Modelos dinámicos (con retardos de la endógena como exógenas)
- (C) Modelos con errores de medida en los regresores

# Hipótesis del modelo de regresión lineal general (XIII)

(A) Modelos de ecuaciones simultáneas. Por ejemplo, un modelo de demanda (primera ecuación) y de oferta de un bien (segunda ecuación) que se intercambia en un mercado competitivo en equilibrio (tercera ecuación), se puede escribir:

$$q_t^d = a + bp_t + \varepsilon_{1t}$$

$$p_t = c + dq_t^o + \varepsilon_{2t}$$

$$q_t^d = q_t^o, \forall t = 1, 2, \dots, n$$

donde se observa una relación bidireccional entre el precio y la cantidad intercambiada. Por ejemplo, el precio en la ecuación de demanda es un regresor estocástico por ser la variable endógena de la ecuación de oferta.

# Hipótesis del modelo de regresión lineal general (XIV)

(B) Modelos dinámicos en los que aparecen como regresores sucesivos retardos de la variable endógena. Por ejemplo, si en la relación entre consumo y renta se supone un modelo dinámico como:

$$C_t = \beta_1 + \beta_2 C_{t-1} + \beta_3 Y_t + \varepsilon_t$$

donde el mismo modelo implica que:

$$C_{t-1} = \beta_1 + \beta_2 C_{t-2} + \beta_3 Y_{t-1} + \varepsilon_{t-1}$$

En la primera ecuación no se puede asumir que el consumo retardado sea un regresor determinista, porque el propio modelo indica que depende de un error  $\varepsilon_{t-1}$



# Hipótesis del modelo de regresión lineal general (XV)

(C) Modelos con errores de medida en las variables explicativas. Bajo la hipótesis de renta permanente de Friedman, el consumo sólo depende del componente permanente de la renta ( $Y_t^P$ )

$$C_t = b Y_t^P + \varepsilon_t$$

$$Y_t = Y_t^P + Y_t^T$$

donde el componente transitorio  $Y_t^T$  o las desviaciones aleatorias alrededor de la renta media de un agente no es observable. Como la renta permanente es una variable no observable, para estimar la primera ecuación hay que sustituirla por  $Y_t^P = Y_t - Y_t^T$  y por tanto, es un regresor estocástico.

# Hipótesis del modelo de regresión lineal general (XVI)

## H7: Hipótesis referentes al error

(H7.1) Esperanza nula en todo instante de tiempo

$$E(\varepsilon_t) = 0, \forall t = 1, 2, \dots, n$$

Ya que el error es tratado como la suma de muchos efectos individuales sobre la endógena, donde el signo de cada uno es desconocido, no existe ninguna razón para esperar cualquier valor distinto de cero. Una situación en la que se incumple esta hipótesis, es cuando a su vez, se incumple otra, como es omitir en el modelo una variable explicativa relevante.

# Hipótesis del modelo de regresión lineal general (XVII)

Como ejemplo, supongamos que la verdadera función de consumo es:

$$C_t = \beta_0 + \beta_1 R_t + \beta_2 i_t + \varepsilon_t$$

donde  $i_t$  es un tipo de interés y el error tiene esperanza nula. Si en lugar de trabajar con este modelo, se omite la variable explicativa del tipo de interés, se tiene que:

$$C_t = \beta_0 + \beta_1 R_t + v_t \quad v_t = \varepsilon_t + \beta_2 i_t$$

Por tanto, la esperanza del nuevo error no puede ser nula, ya que  $E(v_t) = E(\varepsilon_t) + E(\beta_2 i_t) = \beta_2 i_t$  dadas las hipótesis de parámetros constantes y regresores fijos.

# Hipótesis del modelo de regresión lineal general (XVIII)

(H7.2) Varianza constante u Homoscedasticidad.

$$\text{var}(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2, \forall t = 1, 2, \dots, n$$

Si la varianza de los errores cambia con el tiempo (o con el individuo) hablamos de **heteroscedasticidad**. Esta es muy frecuente con datos de sección cruzada. Si tenemos la función de consumo familiar, es habitual que aquellas familias con mayor nivel de renta tengan mayor variabilidad en su consumo  $C_t = \beta_0 + \beta_1 R_t + \varepsilon_t$

Puesto que el error del modelo está relacionado con el consumo, lo que ocurrirá es que a mayor renta, mayor varianza en el consumo y por tanto, mayor varianza en el error.

# Hipótesis del modelo de regresión lineal general (XIX)

## (H7.3) Ausencia de autocorrelación

$$\text{COV}(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0, \forall t, s = 1, 2, \dots, n, t \neq s$$

Si hay autocorrelación, el error en un momento del tiempo ayudaría a predecir el error en un momento posterior y los errores tendrían inercia. Si no hay autocorrelación, la historia pasada no ayuda a predecir el comportamiento futuro y los errores son completamente aleatorios e imprevisibles. Por ejemplo, un patrón de errores autocorrelacionados sería aquél en el que siempre crecen con el tiempo o siempre decrecen.

Es frecuente el incumplimiento de esta hipótesis con datos de series temporales y menos en secciones cruzadas

# Especificación del MLG (I)

- Para estimar por MCO un MLG, es conveniente usar una **notación matricial**. Los **datos de la variable a explicar** (endógena o dependiente) se almacenan en un vector **Y** y los **datos de las variables explicativas** (exógenas o independientes) en una matriz llamada **X**:

$$\begin{pmatrix} y_1 \\ \cdot \\ y_n \end{pmatrix} \quad \begin{pmatrix} x_{11} & \cdot & x_{1k} \\ \cdot & \cdot & \cdot \\ x_{n1} & \cdot & x_{nk} \end{pmatrix}$$

- Cuando la 1ª columna de la matriz **X** es una columna de unos, el **modelo tiene término constante**.

## Especificación del MLG (II)

- Las perturbaciones en un vector  $\varepsilon$  de tamaño  $(n \times 1)$  y los parámetros en un vector  $\beta$  de tamaño  $(k \times 1)$ :

$$\begin{pmatrix} \varepsilon_1 \\ \cdot \\ \varepsilon_n \end{pmatrix} \quad \begin{pmatrix} \beta_1 \\ \cdot \\ \beta_k \end{pmatrix}$$

- El modelo lineal general (MLG) escrito en forma matricial es  $Y = X\beta + \varepsilon$

## Especificación del MLG (III)

- Escribiendo el sistema de  $n$  ecuaciones:

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdot & x_{1k} \\ x_{21} & x_{22} & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdot & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_n \end{pmatrix}$$

Es decir, tenemos un sistema de ecuaciones donde cada una establece la relación entre la endógena y las exógenas en un momento del tiempo concreto.



# Especificación del MLG (IV)

- Las hipótesis sobre las perturbaciones en notación matricial son:

$$E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{var}[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 \mathbf{I}$$

donde

$$\text{var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \text{var}(\varepsilon_1) & \text{cov}(\varepsilon_1\varepsilon_2) & \cdot & \text{cov}(\varepsilon_1\varepsilon_n) \\ \text{cov}(\varepsilon_2\varepsilon_1) & \text{var}(\varepsilon_2) & \cdot & \text{cov}(\varepsilon_2\varepsilon_n) \\ \cdot & \cdot & \cdot & \cdot \\ \text{cov}(\varepsilon_n\varepsilon_1) & \text{cov}(\varepsilon_n\varepsilon_2) & \cdot & \text{var}(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \cdot & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & \cdot & E(\varepsilon_2\varepsilon_n) \\ \cdot & \cdot & \cdot & \cdot \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \cdot & E(\varepsilon_n^2) \end{pmatrix}$$

# Especificación del MLG (V)

Bajo la hipótesis de ausencia de autocorrelación de los errores, fuera de la diagonal principal de la matriz de varianzas-covarianzas hay ceros. Bajo la hipótesis de homoscedasticidad, los elementos de la diagonal principal de la matriz de varianzas son iguales entre sí y constantes en el tiempo:

$$\text{var}(\varepsilon) = \begin{pmatrix} \sigma^2 & 0 & . & 0 \\ 0 & \sigma^2 & . & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & . & \sigma^2 \end{pmatrix}$$

# Estimación por MCO (I)

Dada la formulación matricial del MLG, el **objetivo** es, de nuevo, **obtener la expresión analítica del estimador MCO** de  $\beta$ . Para ello, se define el vector de residuos de tamaño  $(n \times 1)$  como la diferencia entre el valor real de la endógena y su valor ajustado por el modelo:

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

La **función objetivo** sigue siendo minimizar la suma de cuadrados de los residuos con respecto a los  $k$  parámetros

$$\min \sum_{t=1}^n \hat{\varepsilon}_t^2 = \min \hat{\varepsilon}^T \hat{\varepsilon} = \min (Y - X \hat{\beta})^T (Y - X \hat{\beta})$$

## Estimación por MCO (II)

Operando en la función objetivo, tenemos:

$$\min Y^T Y - \hat{\beta}^T X^T Y - Y^T X \hat{\beta} + \hat{\beta}^T X^T X \hat{\beta}$$

donde el segundo y el tercer término son iguales (uno es el traspuesto del otro). Reagrupando, obtenemos que:

$$\min Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X \hat{\beta}$$

# Estimación por MCO (III)

- Condiciones de primer orden

$$\frac{\partial \hat{\varepsilon}^T \hat{\varepsilon}}{\partial \hat{\beta}} = -2X^T Y + 2X^T X \hat{\beta} = 0$$

- La solución analítica a las condiciones de primer orden es:

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

Este es un sistema llamado de  $k$  ecuaciones con  $k$  incógnitas (llamado sistema de ecuaciones normales)

$$\hat{\beta}_1, \dots, \hat{\beta}_k$$

## Estimación por MCO (IV)

- La forma más sencilla de resolver este sistema de ecuaciones es premultiplicar el mismo por la inversa de la matriz  $X^T X$  de tamaño  $(k \times k)$ , teniendo que:

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

- Esta es la fórmula más conocida del estimador MCO de los parámetros del MLG. Exige calcular los elementos de la matriz  $X^T X$ , invertirla y calcular los elementos del vector  $X^T Y$

# Estimación por MCO (V)

## Propiedades estadísticas del estimador MCO

**Linealidad:** El estimador MCO de  $\beta$  es lineal. La linealidad consiste en poder escribir el estimador como una **combinación lineal fija** de los valores de la variable endógena.

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

# Estimación por MCO (VI)

- **Insesgadez:** El estimador MCO de  $\beta$  es insesgado. Es decir, la media de la distribución muestral del estimador de  $\beta$  coincide con el verdadero  $\beta$ .
- Si la  $E(\hat{\beta}) \neq \beta$ , las estimaciones que conseguimos con el estimador no son iguales al verdadero vector de parámetros ni siquiera en media. A la diferencia  $E(\hat{\beta}) - \beta$  se le denomina sesgo.
- Si  $E(\hat{\beta}) - \beta > 0$  estamos en media sobrestimando el verdadero valor de los parámetros al estimar por MCO.
- Si  $E(\hat{\beta}) - \beta < 0$  estamos en media infraestimando el verdadero valor de los parámetros al estimar por MCO.



## Estimación por MCO (VII)

La insesgadez es una propiedad deseable, pero no a toda costa. Por ejemplo, podemos tener dos estimadores alternativos de un parámetro, uno insesgado y otro sesgado.

Si los valores que toma el estimador sesgado oscilan menos alrededor de la media que el insesgado, el primero tendría menos varianza que el segundo.

Un pequeño sesgo compensa menor varianza.

# Estimación por MCO (VIII)

Prueba de insesgadez:

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

O bien, 
$$\hat{\beta} = \beta + (X^T X)^{-1} X^T \varepsilon$$

Tomando esperanzas:

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

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

# Estimación por MCO (IX)

- **Eficiencia:** El estimador MCO de  $\beta$  es eficiente. Es decir, tiene varianza mínima dentro de la familia de estimadores lineales e insesgados de  $\beta$
- Es decir, podemos encontrar un estimador diferente al MCO con un pequeño sesgo, pero menor varianza que el MCO. Esto es lo que demuestra el **Teorema de Gauss y Markov**. Pero antes, hay que derivar la expresión de la matriz de varianzas-covarianzas del estimador MCO de .

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

$$\text{var}(\hat{\beta}) = (X^T X)^{-1} X^T E(\varepsilon \varepsilon^T) X (X^T X)^{-1}$$

# Estimación por MCO (X)

Y, finalmente, aplicando las hipótesis de que las perturbaciones tienen esperanza nula, varianza constante y ausencia de autocorrelación:

$$\text{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

Esta es la expresión de la mínima varianza de un estimador lineal e insesgado de  $\beta$ . Como  $\sigma^2$  es desconocido hay que encontrar un estimador.

**Estimador MCO de la varianza residual** Un estimador intuitivo de la varianza de las perturbaciones consiste en dividir la suma de cuadrados de los residuos MCO por  $n$ . No obstante, para que dicho estimador sea insesgado, hay que ponderar esa suma de cuadrados por los grados de libertad.

# Estimación por MCO (XI)

- Dada cualquier muestra de  $Y$  y  $X$  en el MLG, los pasos en la estimación MCO son:

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

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

$$\text{var}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}$$

# Propiedades algebraicas de la estimación MCO (I)

En **modelos con término constante**, si estimamos por MCO se cumple que:

- (1) La media muestral de la variable endógena real coincide con la media muestral de la variable endógena ajustada por el modelo:  $\bar{y} = \hat{\bar{y}}$
- (2) La media muestral de los residuos es nula:  $\hat{\bar{\varepsilon}} = 0$

En **modelos con o sin término constante** si estimamos por MCO se cumple que:

- (3) Los residuos son ortogonales a las variables explicativas:  $X^T \hat{\varepsilon} = 0$
- (4) Los residuos son ortogonales a los valores ajustados de la endógena:  $\hat{Y}^T \hat{\varepsilon} = 0$

# Propiedades algebraicas de la estimación MCO (II)

(5) En un MRL con o sin término constante, si se estima por MCO se cumple la siguiente identidad:

$$\sum y_t^2 = \sum \hat{y}_t^2 + \sum \hat{\varepsilon}_t^2$$

(6) En un MRL con término constante, si se estima por MCO se cumple la siguiente identidad:

$$\sum (y_t - \bar{y})^2 = \sum (\hat{y}_t - \bar{y})^2 + \sum \hat{\varepsilon}_t^2$$

O bien

$$ST = SE + SR$$

# Bondad del ajuste (I)

- **Bondad del ajuste:** La  $SR$  (suma de cuadrados de residuos) puede ser una medida de bondad de ajuste. Pero no es una buena medida, ya que los residuos tienen escala y esta suma cambia ante un simple cambio de escala en los datos de la endógena.
- Además, la  $SR$  como mínimo es nula, pero su valor máximo no está acotado. Si queremos **una medida adimensional y acotada**, se puede definir un ratio de sumas de cuadrados. Si el modelo tiene término constante: la media de los residuos es cero y las medias de la variable  $Y$  real coincide con la de la  $Y$  ajustada.



## Bondad del ajuste (II)

- La **medida de ajuste más conocida** es el llamado R-cuadrado ó coeficiente de determinación del modelo y viene definido como:

$$R^2 = \frac{SE}{ST} = 1 - \frac{SR}{ST}$$

donde se ha usado la propiedad de que  $ST = SR + SE$

La  $ST$  es la suma de cuadrados de la endógena alrededor de su media, la  $SE$  es la suma de cuadrados de la variable ajustada alrededor de su media y la  $SR$  es la suma de cuadrados de residuos MCO (se cumple con término constante).

## Bondad del ajuste (III)

- El valor del R-cuadrado (multiplicado por 100) se interpreta como el porcentaje de la varianza de la endógena que queda explicada por el modelo.
- Además, está acotado entre cero y uno. Si el  $R^2 = 0$ , el ajuste es nulo, ya que la  $SE = 0$ .
- Si el  $R^2 = 1$ , el ajuste es perfecto, ya que la  $ST = SE$ , o bien, la  $SR = 0$ .
- Ajustes intermedios darían lugar a un  $R^2 = 0.50$ . En este caso, un 50% de la variabilidad de la endógena queda explicada por los regresores incluidos en el modelo.
- Si el modelo es lineal simple, el  $R^2$  es el cuadrado del coeficiente de correlación lineal entre las variables  $Y$  y  $X$ .

# Limitaciones del $R^2$ (I)

- **Limitaciones del  $R^2$ :** Es muy usado, pero tiene problemas. Es engañoso mirar el  $R^2$  sin mirar los datos.
- El **ejemplo más famoso** en la literatura econométrica es la relación entre el N° de nacimientos en un año en los EEUU y el N° de cigueñas en ese mismo año y estados. La estimación del modelo que explica el N° de nacimientos en función del N° de cigueñas proporciona un  $R^2$  muy elevado.
- La **razón** es que en ese año la **correlación lineal muestral entre ambas variables fue muy alta** y aunque no hay ninguna relación causal entre ambas, el coeficiente de determinación es bueno, pero engañoso.

## Limitaciones del $R^2$ (II)

- Incluso cuando tiene sentido relacionar determinadas variables ( por ejemplo, Consumo y Renta per cápita en EEUU), el coeficiente de determinación puede ser excesivamente alto si en el período muestral considerado ambas variables evolucionan de forma muy parecida o presentan una tendencia común.
- Recuérdese que en la regresión lineal simple estimada en el Tema 2 explicando el Consumo per cápita en función de la Renta per cápita en EEUU, el  $R^2$  de la regresión era igual a 0.997. Es decir, más del 99% de la variabilidad del Consumo viene explicada sólo por la Renta. Esto no significa que el modelo sea excepcional.

## Limitaciones del $R^2$ (III)

Otro problema distinto del R-cuadrado es que nunca empeora cuando en el modelo introducimos variables explicativas adicionales. Es decir, aunque una nueva variable no sea muy relevante, su incorporación hace que, en el peor de los casos, el R-cuadrado no cambie, o bien, con un poco de suerte, aumente. La inclusión de variables explicativas irrelevantes no empeora el R-cuadrado.

Introducir un nuevo regresor en el modelo tiene dos efectos: (1) disminuyen los grados de libertad y éste es negativo y (2) disminuye la suma de cuadrados de residuos y éste es positivo. Si el peso del efecto negativo es mayor que la mejora en el ajuste, no compensará introducir esta nueva variable y a la inversa.

# Coeficiente de determinación corregido (I)

- La solución a éste último problema es utilizar el llamado **R2 ajustado o corregido de grados de libertad** que se calcula como:

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

- **Los dos efectos son:** (a) Si aumenta el número de regresores en el modelo, disminuyen los grados de libertad y esto se penaliza  $\uparrow k \Rightarrow \downarrow n-k \Rightarrow \uparrow \frac{n-1}{n-k} \Rightarrow \downarrow \bar{R}^2$
- y (b) Esos nuevos regresores pueden mejorar el modelo en términos de ajuste,  $\uparrow k \Rightarrow \downarrow SR \Rightarrow \uparrow R^2 \Rightarrow \uparrow \bar{R}^2$

# Coeficiente de determinación corregido (II)

Al final, si el efecto negativo es mayor (en valor absoluto) que el positivo, el R-cuadrado corregido bajará y no compensaría introducir el nuevo regresor.

Si el efecto positivo es mayor que el negativo, compensa introducir el nuevo regresor pese a la pérdida de grados de libertad. Como ejemplo, en las 2 regresiones siguientes compensa quedarse con el modelo más amplio:

$$C_t = \beta_0 + \beta_1 R_t + v_t \quad \bar{R}^2 = 0.80$$

$$C_t = \beta_0 + \beta_1 R_t + \beta_2 i_t + \varepsilon_t \quad \bar{R}^2 = 0.88$$

# Criterios de información (I)

- Otra forma de resolver los problemas del R2 es usar algún **criterio de información**, que tiene en cuenta en la misma fórmula tanto el ajuste del modelo como el  $n^0$  de parámetros del mismo. Los más conocidos son los **criterios de Akaike (AIC)** y el de Schwarz (SIC) que se definen como:

$$AIC(k) = \log[\hat{\sigma}_k^2] + \frac{2k}{n}$$

$$SIC(k) = \log[\hat{\sigma}_k^2] + \frac{k \log[n]}{n}$$

- donde  $n$  es el tamaño de la muestra,  $k$  el número de regresores y  $\hat{\sigma}_k^2$  la estimación de la varianza residual por máxima verosimilitud (dividiendo la  $SR/n$ ).



## Criterios de información (II)

- La diferencia entre el estimador de la varianza residual por MCO y por MV consiste en que por MCO se pondera la suma de cuadrados de residuos por los grados de libertad y por MV se pondera por el tamaño de la muestra.
- El estimador MCO de esa varianza es insesgado, mientras que el MV tiene sesgo (aunque este sesgo tiende a desaparecer a medida que aumenta el tamaño de la muestra).

# Criterios de información (III)

- Ambos criterios tienen en cuenta un término de penalización por el  $n^0$  de parámetros, ya que el ajuste del modelo siempre mejora (es decir, la varianza residual baja) cuando aumenta el  $n^0$  de parámetros.
- En modelos anidados (es decir, donde la endógena es siempre la misma y se va añadiendo uno a uno los regresores), se escoge aquél que tenga un menor valor del *AIC* y del *SIC*.
- Para  $n$  mayor o igual a 8, el *SIC* supone una penalización más fuerte sobre la inclusión de variables extra que el criterio *AIC*. Por ello, el criterio *SIC* tiende a quedarse con modelos más pequeños que el *AIC*.