# Econometria
# Tema 4(1): Errors d'Especificació i Canvi Estructural

Ramon Alemany

Grau Estadística UB-UPC

**Curs 2017-18**

# Presentació

# Bibliografia


- GREENE, W. (1999)
  **Análisis econométrico. 3a Ed.**
  *Capítols 7 i 8*

- WOOLDRIDGE, J. (2009)
  **Introducción a la Econometría. Un enfoque moderno. 4a Ed.**
  *Capítol 9*

- STOCK, J. & WATSON, M. (2012)
  **Introducción a la Econometría. 3a Ed.**
  *Capítols 8 i 9*

# Errors d'especificació de la forma funcional

## 1. Errors d'especificació de la forma funcional

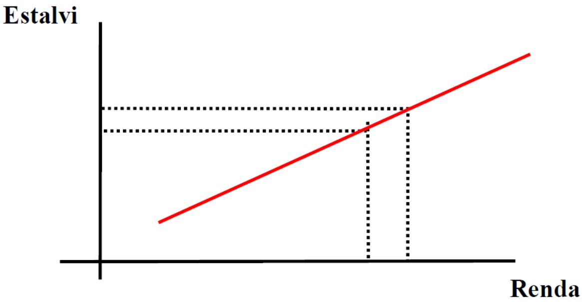
Direm que existeix un error en la forma funcional quan especifiquem una relació entre les variables endògena i exògenes que resulta ser diferent de la forma funcional verdadera.

Fins el moment, hem suposat que treballaríem sempre amb models lineals o linealitzables.

Per exemple, si volguéssim modelitzar la relació entre l'estalvi i la renda pels habitants d'una regió, definiríem el següent model:

# Errors d'especificació de la forma funcional

$$\text{Estalvi}_i = \beta_1 + \beta_2 \text{Renda}_i + u_i \quad i = 1, 2, 3, \ldots, N$$

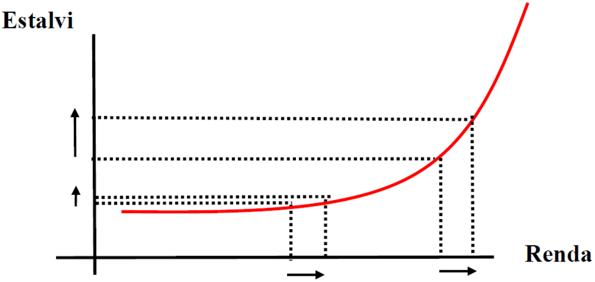



De manera que, amb independència del nivell de renda dels individus (i sota el supòsit que $\beta_2 > 0$), l'augment d'1 unitat de la renda implicaria un augment de $\beta_2$ unitats d'estalvi.

# Errors d'especificació de la forma funcional

Però podria succeir que, per nivells baixos de renda, un increment unitari de renda es destinés en la seva major part a consum mentre que, per nivells elevats, un increment unitari de renda es destinés en la seva pràctica totalitat a estalvi. En aquest cas, la relació no seria lineal sinó quadràtica:

$$\text{Estalvi}_i = \beta_1 + \beta_2 \text{Renda}_i + \beta_3 \text{Renda}_i^2 + u_i$$



# Errors d'especificació de la forma funcional

## Conseqüències

**Conseqüències de cometre un error en l'especificació de la forma funcional:**

Els estimadors MQO serien **esbiaixats**, **inconsistents**, i es podrien veure afectades també les hipòtesis d'esfericitat i de normalitat del terme de pertorbació.

És necessari doncs contrastar l'existència d'un error en l'especificació de la forma funcional...

# Errors d'especificació de la forma funcional

## Test RESET

**Regression Equation Specification Error Test (RESET) de Ramsey**

$H_0$: Forma funcional lineal correcta

$H_A$: Forma funcional lineal no correcta

Concretament, per contrastar si la forma funcional lineal és la correcta o, per contra, la forma funcional correcta és la quadràtica, construirem el contrast RESET seguint els passos següents:

# Errors d'especificació de la forma funcional
## Test RESET

1r) Estimem per MQO el MRLM:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

2n) Calculem la variable endògena ajustada $\hat{y}_i$

3r) Especifiquem una regressió auxiliar introduint la variable endògena ajustada elevada al quadrat com un regressor més del MRLM inicial:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \gamma \hat{y}_i^2 + u_i$$

4t) Contrastem la significació individual de $\gamma$:

$$\begin{cases} H_0 : \gamma = 0 \\ H_A : \gamma \neq 0 \end{cases}$$

No Rebuig $H_0 \Longrightarrow$ Forma funcional lineal correcta

Rebuig $H_0 \Longrightarrow$ Forma funcional quadràtica

# Errors d'especificació de la forma funcional

## Test RESET

Si volem contrastar no linealitats d'ordre superior aleshores:

3r) Especifiquem una regressió auxiliar:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + \gamma_3 \hat{y}_i^4 + u_i$$

4t) Contrastem la significació conjunta dels paràmetres $\gamma$:

$$\begin{cases} H_0: & \gamma_1 = \gamma_2 = \gamma_3 = 0 \\ H_A: & \exists \gamma_j \neq 0 \end{cases}$$

No Rebuig $H_0$ $\Rightarrow$ Forma funcional lineal correcta
Rebuig $H_0$ $\Rightarrow$ Forma funcional no lineal (quadràtica, cúbica,...)

# Formes funcionals no lineals

## Model lineal estàndard

Model Lineal: $\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i}$

Efecte Marginal: $\boldsymbol{\frac{\partial Y}{\partial X} = \beta_1}$

Un canvi unitari en $X$ suposa un canvi de $\boldsymbol{\beta_1}$ unitats en $\boldsymbol{Y}$

$$\widehat{TestScore} = 625{,}38 + 1{,}8785 \cdot AvgInc_i$$

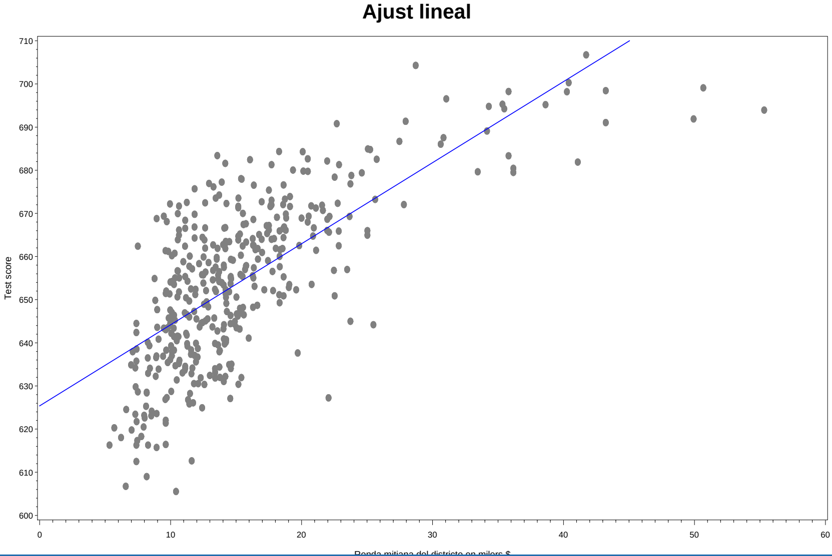
$TestScore$: Aprox entre 600 i 700 punts. Mitjana $=$ 654.16

$AvgInc$: Renda per càpita en el districte en milers de $

Un increment de 1000$ en la renda per càpita del districte suposa un increment de 1.8785 punts en el TestScore.

# Formes funcionals no lineals

## Model lineal estàndard



# Formes funcionals no lineals

## Model Quadràtic

Model Quadràtic: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$

Efecte Marginal: $\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X_i$

L'efecte sobre el valor esperat de $Y$ davant d'un canvi unitari en $X$ dependrà del valor de $X$ en el què es mesuri el canvi:

$$\widehat{TestScore} = 607{,}30 + 3{,}85 \cdot AvgInc_i - 0{,}04231 \cdot AvgInc_i^2$$

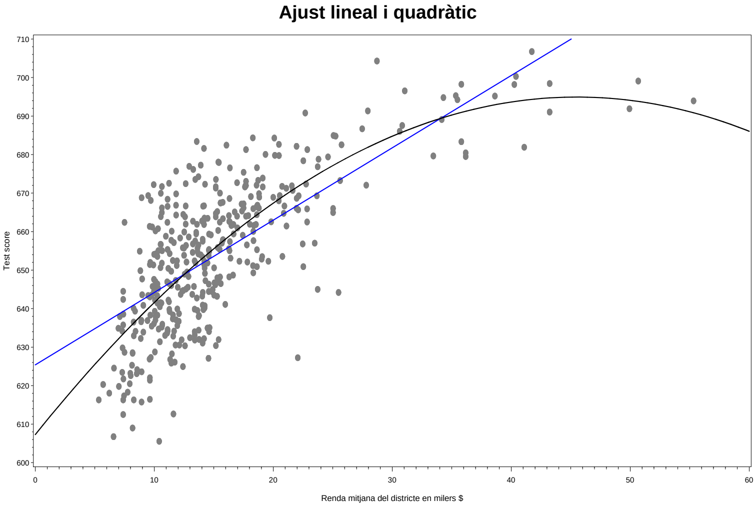
# Formes funcionals no lineals

## Model Quadràtic

| **$\Delta$ 1000 $ renda per càpita** | **$\Delta$ TestScore** |
|---|---|
| de 5000 a 6000 $ | $(3.85 - 2 \cdot 0.04231 \cdot 5) = 3.42$ |
| de 25000 a 26000 $ | $(3.85 - 2 \cdot 0.04231 \cdot 25) = 1.73$ |
| de 45000 a 46000 $ | $(3.85 - 2 \cdot 0.04231 \cdot 45) = 0.04$ |
| de 50000 a 51000 $ | $(3.85 - 2 \cdot 0.04231 \cdot 50) = -0.38$ |
| de 60000 a 61000 $ | $(3.85 - 2 \cdot 0.04231 \cdot 60) = -1.23$ |

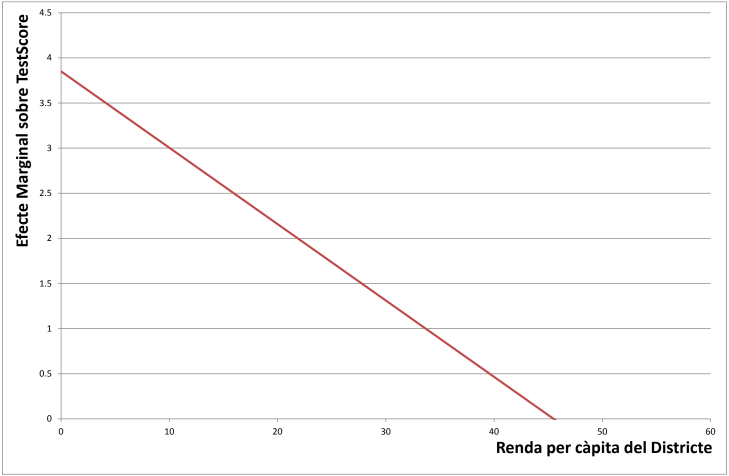
# Formes funcionals no lineals

## Model Quadràtic




Regression Equation:
testscr = 625.3836 + 1.87855*avginc
testscr = 607.3017 + 3.850995*avginc - 0.042308*avginc^2

# Formes funcionals no lineals
## Model Quadràtic



# Formes funcionals no lineals

## Model Cúbic

Model Cúbic: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$

Efecte Marginal: $\dfrac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X_i + 3\beta_3 X_i^2$

L'efecte sobre el valor esperat de $Y$ davant d'un canvi unitari en $X$ dependrà del valor de $X$ en el què es mesuri el canvi:

$$\widehat{TestScore} = 600{,}08 + 5{,}02 \cdot AvgInc_i - 0{,}09581 \cdot {AvgInc_i}^2 + 0{,}00068548 \cdot {AvgInc_i}^3$$

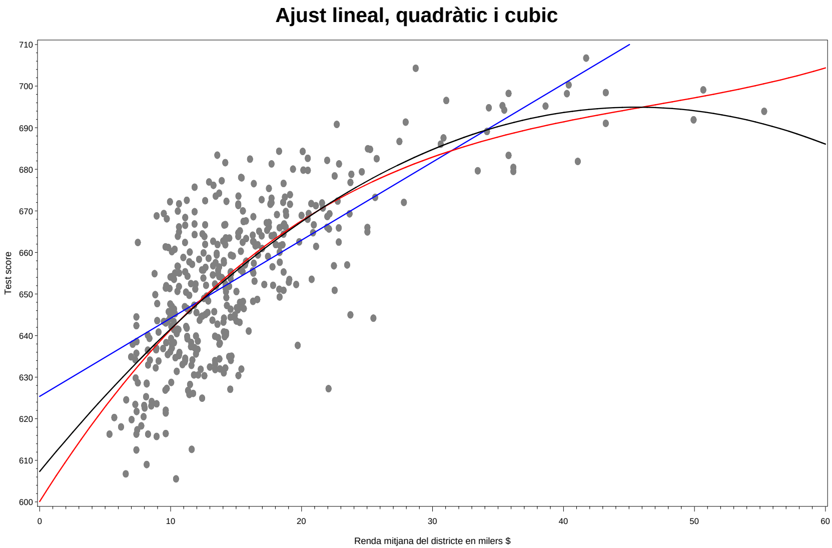
# Formes funcionals no lineals

## Model Cúbic

| $\Delta$ 1000 $ renda per càpita | $\Delta$ TestScore |
|---|---|
| de 5000 a 6000 $ | (5.02 - 2 $\cdot$ 0.09581 $\cdot$ 5 + 3 $\cdot$ 0.00068548 $\cdot 5^2$)= 4.11 |
| de 25000 a 26000 $ | (5.02 - 2 $\cdot$ 0.09581 $\cdot$ 25 + 3 $\cdot$ 0.00068548 $\cdot 25^2$)= 1.51 |
| de 45000 a 46000 $ | (5.02 - 2 $\cdot$ 0.09581 $\cdot$ 45 + 3 $\cdot$ 0.00068548 $\cdot 45^2$)= 0.56 |
| de 50000 a 51000 $ | (5.02 - 2 $\cdot$ 0.09581 $\cdot$ 50 + 3 $\cdot$ 0.00068548 $\cdot 50^2$)= 0.58 |
| de 60000 a 61000 $ | (5.02 - 2 $\cdot$ 0.09581 $\cdot$ 60 + 3 $\cdot$ 0.00068548 $\cdot 60^2$)= 0.93 |

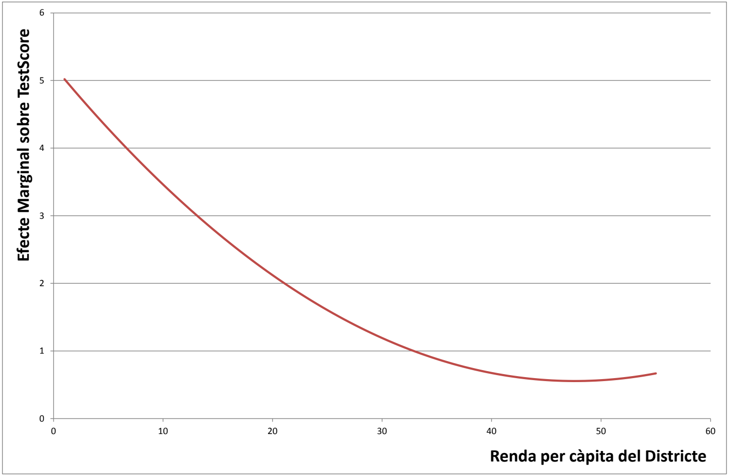
# Formes funcionals no lineals

## Model Cúbic



# Formes funcionals no lineals

## Model Cúbic



# Formes funcionals no lineals

## Model Lineal-Log

Model Lineal-Log: $\boldsymbol{Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i}$

Efecte Marginal: $\boldsymbol{\Delta Y = \beta_1 \Delta \ln(X)}$ $\boldsymbol{\frac{\partial Y}{\partial X} = \beta_1 \frac{1}{X}}$

$$\boldsymbol{\Delta Y \approx \frac{\beta_1}{100}\,\%\Delta X}$$

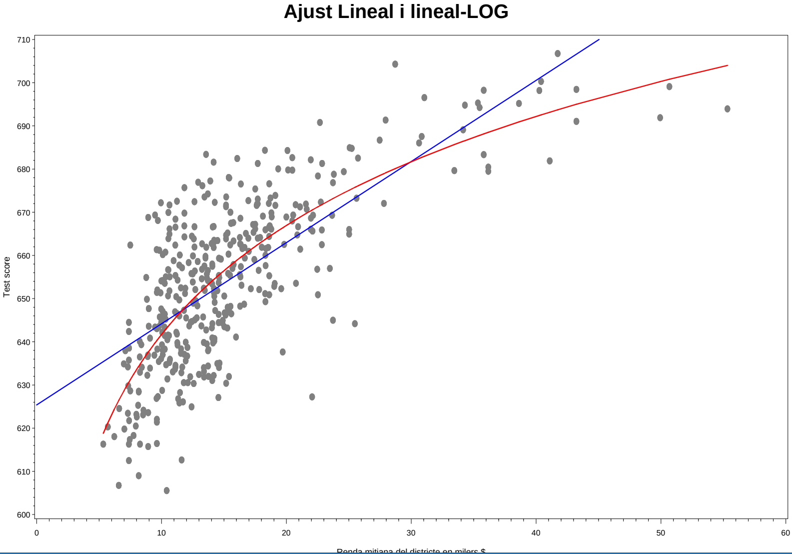
Una variació d'un 1 % en $X$ suposa un canvi de $\frac{\beta_1}{100}$ unitats en $Y$:

$$\widehat{TestScore} = 557{,}83 + 36{,}42 \cdot \ln(AvgInc_i)$$

Un increment d'un 1 % en la renda per càpita del districte suposa un increment de 0.3642 punts en el TestScore.

# Formes funcionals no lineals

## Model Lineal-Log



# Formes funcionals no lineals

## Model Log-Lineal

Model Log-Lineal: $\mathbf{ln}(\boldsymbol{Y_i}) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 X_i} + \boldsymbol{u_i}$

Efecte Marginal: $\boldsymbol{\Delta}\,\mathbf{ln}(\boldsymbol{Y}) = \boldsymbol{\beta_1 \Delta X} \quad \frac{\boldsymbol{\partial Y}}{\boldsymbol{\partial X}} = \boldsymbol{\beta_1 Y}$

$$\boldsymbol{\%\Delta Y \approx 100\beta_1 \Delta X}$$

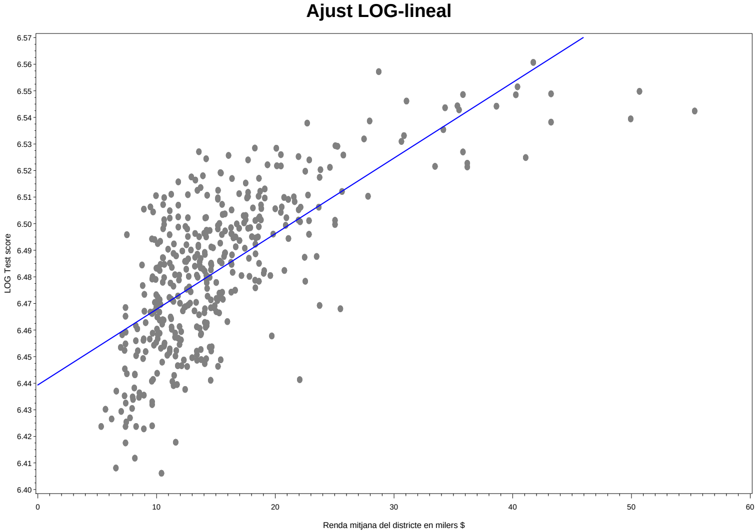
Un increment d'una unitat en $X$ suposa un canvi de $100\beta_1\,\%$ en $Y$:

$$\ln(\widehat{TestScore}) = 6{,}44 + 0{,}00284 \cdot AvgInc_i$$

Un increment de 1000 \$ en la renda per càpita del districte suposa un increment del 0.284 % en el TestScore.

# Formes funcionals no lineals

## Model Log-Lineal



# Formes funcionals no lineals

## Model Log-Log

Model Log-Log: $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$

Efecte Marginal: $\Delta \ln(Y) = \beta_1 \Delta \ln(X) \quad \frac{\partial Y}{\partial X} = \beta_1 \frac{Y}{X}$

$$\%\Delta Y \approx \beta_1 \, \%\Delta X$$

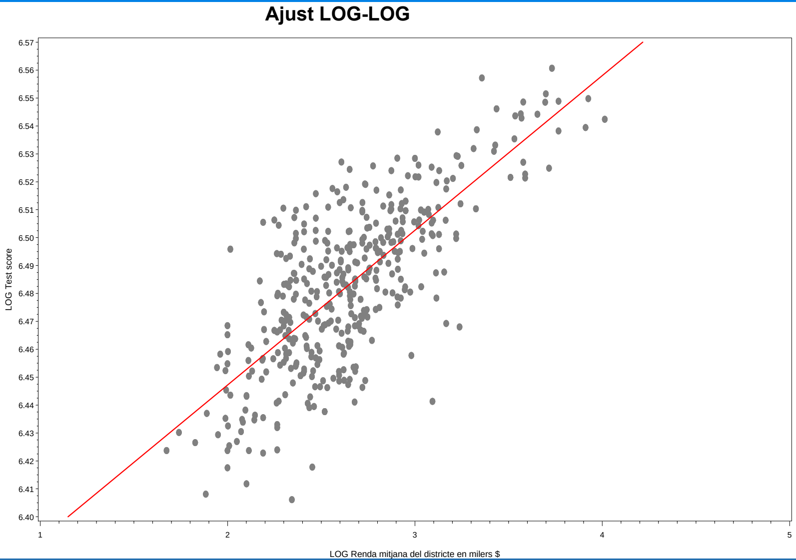
Un increment d'un 1 % en $X$ suposa un canvi de $\beta_1$ % en $Y$:

$$\ln(\widehat{TestScore}) = 6{,}34 + 0{,}0554 \cdot \ln(AvgInc_i)$$

Un increment d'un 1 % en la renda per càpita del districte suposa un increment del 0.0554 % en el TestScore.

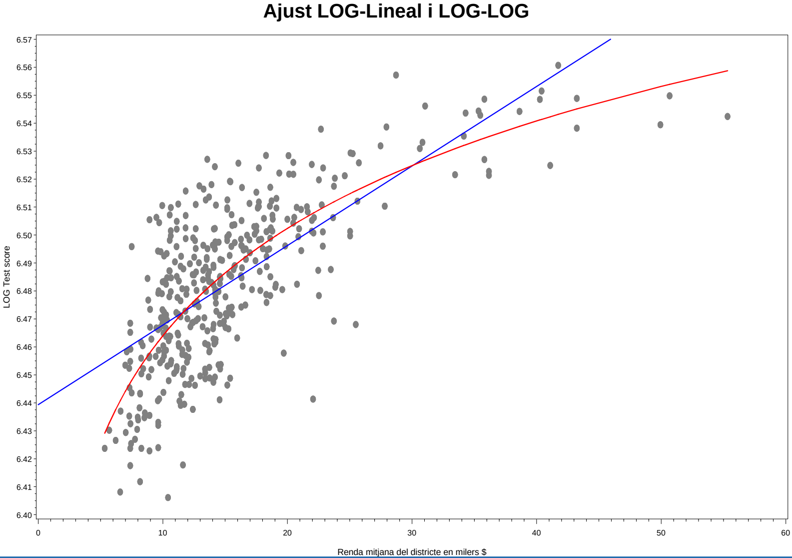
# Formes funcionals no lineals

## Model Log-Log



# Formes funcionals no lineals

## Model Log-Log



# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

**PREDICCIÓ:**

Donat el model Log-Lineal: $\mathbf{\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i}$

De l'estimació del model tindrem que:

$$\widehat{\ln(Y_i)} = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Podríem pensar que per fer prediccions del valor de $Y$ seria suficient amb fer:

$$\hat{Y}_i = exp(\widehat{\ln(Y_i)}) = exp(\hat{\beta}_0 + \hat{\beta}_1 X_i)$$

Però aquesta expressió **subestimaria** la predicció de $Y$.

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

a) $u_i$ **és Normal** $\Longrightarrow u_i \sim N(0, \sigma^2)$

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

aleshores:

$$Y_i = exp(\beta_0 + \beta_1 X_i + u_i) = exp(\beta_0 + \beta_1 X_i) \cdot exp(u_i)$$

El valor esperat de la variable dependent original serà:

$$E(Y_i | X_i) = exp(\beta_0 + \beta_1 X_i) \cdot E[exp(u_i)]$$

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

Però si $u_i \sim N(\mu, \sigma^2)$ aleshores $exp(u_i) = w_i$ serà **lognormal**

amb mitjana $E(w_i) = exp(\mu + \frac{\sigma^2}{2})$

i variança $V(w_i) = \left[ [exp(\sigma^2) - 1] \cdot exp(2\mu + \sigma^2) \right]$

Per tant, si $u_i \sim N(0, \sigma^2)$ aleshores $E[exp(u_i)] = exp(\frac{\sigma^2}{2})$ i:

$$E(Y_i|X_i) = exp(\beta_0 + \beta_1 X_i) \cdot exp(\frac{\sigma^2}{2})$$

i la predicció de $Y$ haurà de ser:

$$\hat{Y}_i = exp(\widehat{\ln(Y_i)}) \cdot exp(\frac{\hat{\sigma}^2}{2})$$

la qual és esbiaixada però **consistent**.

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

b) $u_i$ **no és Normal**

Si assumim que el terme de pertorbació és independent del conjunt d'explicatives, aleshores:

$$E(Y_i|X_i) = \alpha_0 \cdot exp(\beta_0 + \beta_1 X_i)$$

on $\alpha_0 = E[exp(u_i)]$. I amb una estimació consistent de $\hat{\alpha}_0$ podem predir consistentment $Y$:

$$\hat{Y}_i = \hat{\alpha}_0 \cdot exp(\widehat{\ln(Y_i)})$$

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

**Predicció de $Y$ quan la variable dependent és $ln(Y)$**

(1) Obtindrem els valors ajustats $\widehat{\ln(Y_i)}$ a partir de l'estimació del model $\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$.

(2) Per cada observació $i$ calcularem $\hat{m}_i = exp(\widehat{\ln(Y_i)})$.

(3) Farem una regressió auxiliar de $Y_i$ sobre $\hat{m}_i$ sense terme independent.

$$Y_i = \gamma \hat{m}_i + \epsilon_i$$

i per tant el coeficient $\gamma$ serà una estimació de $\alpha_0$.

(4) Un cop disposem de l'estimació de $\hat{\alpha}_0$ podem obtenir la predicció per $Y_i$:

$$\hat{Y}_i = \hat{\alpha}_0 \cdot exp(\widehat{\ln(Y_i)})$$

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

**BONDAT DE L'AJUST:**
**El $R^2$ quan la variable dependent és $ln(Y)$**

L'objectiu és obtenir una mesura de la bondat de l'ajust en el model on la variable dependent és $ln(Y_i)$ i que pugui ser comparable amb el $R^2$ d'un model on la variable dependent és $Y_i$.

Una possible forma de fer-ho és:

(a) Amb la regressió auxiliar de l'etapa (3) del procediment anterior podem obtenir els valors ajustats $\hat{Y}_i = \hat{\alpha}_0 \cdot \hat{m}_i$.
(b) Calculem el coeficient de correlació entre $\hat{Y}_i$ i el valor observat de $Y_i$
(c) El quadrat d'aquest coeficient de correlació **pot ser comparat** amb el $R^2$ del model en el què la variable dependent és la $Y_i$.

(Recordeu que el $R^2$ en el model ajustat $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ es pot calcular també com el coeficient de correlació al quadrat entre $Y_i$ i $\hat{Y}_i$.)

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

EXEMPLE: Models de TestScore en funció de AvgInc

$TestScore_i = \beta_0 + \beta_1 AvgInc_i + u_i$

$$\widehat{TestScore} = 625{,}38 + 1{,}8785 \cdot AvgInc_i$$

$$R^2 = 0{,}5076$$

Un increment de 1000$ en la renda per càpita del districte suposa un increment de 1.8785 punts en el TestScore.

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

$$\ln(TestScore_i) = \beta_0 + \beta_1 AvgInc_i + u_i$$

$$\ln(\widehat{TestScore}) = 6{,}44 + 0{,}00284 \cdot AvgInc_i$$

$$R^2 = 0{,}4982$$

Un increment de 1000 $ en la renda per càpita del districte suposa un increment del 0.284 % en el TestScore.

# Formes funcionals no lineals

## Model Log-Lineal: Predicció i $R^2$

L'estimació de la regressió auxiliar:

$$TestScore = 1{,}00020 \cdot exp(\ln(\widehat{TestScore}))$$

$$\hat{\alpha}_0 = 1{,}00020$$

$$R^2 = 0{,}5012$$

# Especificació errònia de les explicatives

Omissió de variables rellevants

## 2. Especificació errònia de les explicatives

### 2.1. Omissió de variables rellevants

Direm que existeix un error per omissió de variables rellevants quan no hem inclòs al model al menys una variable que explica de forma significativa el comportament de la variable endògena.

Aquest error es pot cometre bàsicament per dos motius. Que s'ignori que una o més variables són rellevants per explicar l'endògena o bé que, tot i que se sap que són rellevants, manquin dades fiables sobre aquestes variables.

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

Quines conseqüències té l'omissió de variables rellevants?

1r) Sobre les propietats dels estimadors MQO de $\boldsymbol{\beta}$

2n) Sobre les propietats de l'estimador MQO de $\sigma_u^2$

Per veure quines conseqüències té l'omissió de variables rellevants partim del cas concret en què el model verdader té dues variables explicatives i el model que erròniament especifiquem en té només una.

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

Tindrem els següents models en desviacions:

Model correcte: $\tilde{y}_i = \beta_1 \tilde{x}_{1i} + \beta_2 \tilde{x}_{2i} + u_i$

Model especificat: $\tilde{y}_i = \beta_2 \tilde{x}_{2i} + u_i$

Aleshores, l'estimador del model especificat serà:

$$\hat{\beta}_2 = \frac{\sum \tilde{x}_{2i}\tilde{y}_i}{\sum \tilde{x}_{2i}^2} = \frac{\sum \tilde{x}_{2i}(\beta_1 \tilde{x}_{1i} + \beta_2 \tilde{x}_{2i} + u_i)}{\sum \tilde{x}_{2i}^2} =$$

$$= \frac{\beta_1 \sum \tilde{x}_{2i}\tilde{x}_{1i} + \beta_2 \sum \tilde{x}_{2i}^2 + \sum \tilde{x}_{2i} u_i}{\sum \tilde{x}_{2i}^2} = \beta_1 \frac{\sum \tilde{x}_{2i}\tilde{x}_{1i}}{\sum \tilde{x}_{2i}^2} + \beta_2 + \frac{\sum \tilde{x}_{2i} u_i}{\sum \tilde{x}_{2i}^2}$$

Prenent esperances:

$$\mathrm{E}(\hat{\beta}_2) = \beta_1 \frac{\sum \tilde{x}_{2i}\tilde{x}_{1i}}{\sum \tilde{x}_{2i}^2} + \beta_2$$

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

Ara doncs **l'estimador MQO de $\beta$ és esbiaixat**.

El biaix depèn de:

- La magnitud de $\beta_1$, és a dir, del paràmetre associat a la variable omesa. A major valor del paràmetre, més gran serà el biaix.
- Si $x_1$ i $x_2$ estan incorrelacionades aleshores:
  $$\frac{\sum \tilde{x}_{2i}\tilde{x}_{1i}}{\sum \tilde{x}_{2i}^2} = 0$$ i el biaix serà 0.

En conclusió, l'estimació serà esbiaixada excepte en el cas que les variables incloses i omeses siguin ortogonals entre sí.

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

La variància en el model especificat serà:

$$Var(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum \tilde{x}_{2i}^2}$$

mentre que en el model correcte seria:

$$Var(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum \tilde{x}_{2i}^2 (1 - r_{12})}$$

Per tant, com més correlacionades estiguin les variables $x_1$ i $x_2$ més gran serà la diferència entre ambdues variàncies.

Per altra part, si les variables $x_1$ i $x_2$ estan incorrelacionades (són ortogonals) les variàncies seran iguals.

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

D'aquesta manera, la variància dels estimadors del model que té una omissió de variables rellevants és més petita o igual que la del model ben especificat. Únicament seran iguals en el cas de que les variables incloses i omeses siguin ortogonals entre sí.

A aquesta conclusió es podia haver arribat també si es té en compte que l'estimació d'un model amb omissió de variables rellevants equival a l'estimació d'un model restringit, suposant que els paràmetres de les variables omeses són 0.

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

### Els estimadors MQO de $\beta$ són inconsistents

Donat que els estimadors dels paràmetres del model amb omissió de variables rellevants són esbiaixats i que aquest biaix no tendeix a 0 quan la grandària de la mostra augmenta, direm que són estimadors inconsistents en Error Quadràtic Mig.

$$\lim_{N\to\infty} \text{EQM}(\hat{\beta}_2) = \lim_{N\to\infty} \left[ (\text{biaix}(\hat{\beta}_2))^2 + \text{Var}(\hat{\beta}_2) \right] =$$

$$= \lim_{N\to\infty} \left[ \beta_1 \frac{\sum \tilde{x}_{2i}\tilde{x}_{1i}}{\sum \tilde{x}_{2i}^2} \right]^2 + \lim_{N\to\infty} \left[ \frac{\sigma_u^2}{\sum \tilde{x}_{2i}^2} \right] \neq 0$$

Únicament seran consistents quan les variables incloses i omeses siguin ortogonals.

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

**L'estimador MQO de $\sigma_u^2$ és esbiaixat**

Pel model correctament especificat tenim que:

$$\hat{\sigma}_u^2 = \frac{e'e}{N-k}$$

suposant que hi ha hagut l'omissió d'una variable explicativa, en el model especificat de manera errònia tindrem:

$$\hat{\sigma}_u^2 = \frac{e^{*\prime}e^*}{N-(k-1)}$$

i es pot demostrar que aquest és un estimador esbiaixat.

# Especificació errònia de les explicatives

## Omissió de variables rellevants: Conseqüències

Per últim, cal esmentar que totes les variables omeses queden amagades sota el terme de pertorbació, podent afectar al compliment de les hipòtesis bàsiques relacionades amb el terme d'error (esperança nul·la, homoscedasticitat, no autocorrelació i normalitat).

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants

### 2.2. Inclusió de variables irrellevants

Direm que hi ha un error d'inclusió de variables irrellevants quan s'inclouen variables explicatives en el model que resulten ser irrellevants per explicar el comportament de la variable endògena.

Quines conseqüències té l'omissió de variables rellevants?

1r) Sobre les propietats dels estimadors MQO de $\boldsymbol{\beta}$

2n) Sobre les propietats de l'estimador MQO de $\sigma_u^2$

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants: conseqüències

Per veure quines conseqüències té la inclusió de variables irrellevants partim del cas concret en què el model verdader té una variable explicativa i el model que erròniament especifiquem en té dues.

Tindrem els següents models en desviacions:

Model correcte: $\tilde{y}_i = \beta_2 \tilde{x}_{2i} + u_i$

Model especificat: $\tilde{y}_i = \beta_1 \tilde{x}_{1i} + \beta_2 \tilde{x}_{2i} + u_i$

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants: conseqüències

**L'estimador MQO de $\beta_1$ i $\beta_2$ és no esbiaixat.**

El vector d'estimadors en el model erroni és:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum \tilde{x}_{1i}^2 & \sum \tilde{x}_{1i}\tilde{x}_{2i} \\ \sum \tilde{x}_{1i}\tilde{x}_{2i} & \sum \tilde{x}_{2i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum \tilde{x}_{1i}\tilde{y}_i \\ \sum \tilde{x}_{1i}\tilde{y}_i \end{bmatrix}$$

i per tant pels estimadors de $\beta_1$ i $\beta_2$ tindrem:

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants: conseqüències

$$\hat{\beta}_1 = \frac{\sum \tilde{x}_{2i}^2 \sum \tilde{x}_{1i}\tilde{y}_i - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{2i}\tilde{y}_i}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} =$$

$$= \frac{\sum \tilde{x}_{2i}^2 \sum \tilde{x}_{1i}(\beta_2 \tilde{x}_{2i} + u_i) - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{2i}(\beta_2 \tilde{x}_{2i} + u_i)}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} =$$

$$= \frac{(\beta_2 \sum \tilde{x}_{2i}^2 \sum \tilde{x}_{1i}\tilde{x}_{2i} + \sum \tilde{x}_{2i}^2 \sum \tilde{x}_{1i}u_i)}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} -$$

$$- \frac{(\beta_2 \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{2i}^2 + \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{2i}u_i)}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} =$$

$$= \frac{\sum \tilde{x}_{2i}^2 \sum \tilde{x}_{1i}u_i - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{2i}u_i}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2}$$

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants: conseqüències

$$\hat{\beta}_2 = \frac{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}\tilde{y}_i - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{1i}\tilde{y}_i}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} =$$

$$= \frac{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}(\beta_2 \tilde{x}_{2i} + u_i) - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{1i}(\beta_2 \tilde{x}_{2i} + u_i)}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} =$$

$$= \frac{(\beta_2 \sum \tilde{x}_1^2 \sum \tilde{x}_2^2 + \sum \tilde{x}_1^2 \sum \tilde{x}_2 u) - (\beta_2 (\sum \tilde{x}_1 \tilde{x}_2)^2 + \sum \tilde{x}_1 \tilde{x}_2 \sum \tilde{x}_1 u))}{\sum \tilde{x}_1^2 \sum \tilde{x}_2^2 - (\sum \tilde{x}_1 \tilde{x}_2)^2} =$$

$$= \frac{\beta_2 (\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2)}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} + \frac{(\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}u_i - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{1i}u_i}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} =$$

$$= \beta_2 + \frac{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}u_i - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{1i}u_i}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2}$$

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants: conseqüències

És a dir, que tenim:

$$\mathsf{E}(\hat{\beta}_1) = \frac{\sum \tilde{x}_{2i}^2 \sum \tilde{x}_{1i} \mathsf{E}(u_i) - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{2i} \mathsf{E}(u_i)}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} = 0$$

$$\mathsf{E}(\hat{\beta}_2) = \beta_2 + \frac{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i} \mathsf{E}(u_i) - \sum \tilde{x}_{1i}\tilde{x}_{2i} \sum \tilde{x}_{1i} \mathsf{E}(u_i)}{\sum \tilde{x}_{1i}^2 \sum \tilde{x}_{2i}^2 - (\sum \tilde{x}_{1i}\tilde{x}_{2i})^2} = \beta_2$$

Com la variable $x_1$ s'ha inclòs erròniament el seu paràmetre poblacional és 0 i, per tant, l'estimador és no esbiaixat.

Així doncs davant la inclusió de variables irrellevants en el model els estimadors MQO seran **no esbiaixats**.

A més, atès que els estimadors són no esbiaixats, direm que són estimadors **consistents en Error Quadràtic Mig**.

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants: conseqüències

D'altra banda, la variància en el model especificat serà:

$$Var(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum \tilde{x}_{2i}^2 (1 - r_{12})}$$

mentre que en el model correcte seria:

$$Var(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum \tilde{x}_{2i}^2}$$

Per tant, la variància de l'estimador amb inclusió de variables irrellevants és més gran que la del model correcte. Així doncs, hi haurà pèrdua d'eficiència excepte en el cas que les variables $x_1$ i $x_2$ estiguin incorrelacionades, supòsit en el qual les variàncies serien iguals.

# Especificació errònia de les explicatives

## Inclusió de variables irrellevants: conseqüències

**L'estimador MQO de $\sigma_u^2$ és no esbiaixat**

Es pot demostrar que en el cas d'inclusió de variables irrellevants l'estimador MQO de la variància del terme de pertorbació és un estimador no esbiaixat.

# Especificació errònia de les explicatives

| | **Omissió de variables rellevants** | **Inclusió de variables irrellevants** |
|---|---|---|
| $\hat{\beta}_{\text{MQO}}$ | Estimadors esbiaixats i inconsistents (excepte si omeses i incloses són ortogonals) | Estimadors no esbiaixats i consistents |
| $\text{Var}(\hat{\beta})$ | Estimadors amb **menor** variància que la del model correcte (excepte si omeses i incloses són ortogonals) | Estimadors amb **major** variància que la del model correcte (excepte si omeses i incloses són ortogonals) |
| $\hat{\sigma}^2_{\text{MQO}}$ | Estimador esbiaixat | Estimador no esbiaixat |

# Canvi Estructural

## 3. Canvi estructural

Una de les hipòtesis bàsiques del MRLM és l'existència d'estabilitat o permanència estructural, és a dir, que els paràmetres del model es mantenen constants per tota la mostra.

Malgrat això, en determinats casos aquesta hipòtesi no es compleix donant lloc a l'aparició d'inestabilitat estructural (**canvi estructural**) en el model.

# Canvi Estructural

Així, en el cas de dades de sèrie temporal, es poden produir determinats esdeveniments que afectin a l'economia, de manera que es produeixin canvis en l'estructura interna del procés generador de dades que volem analitzar (per exemple, l'adhesió d'un país com a membre d'una unió econòmica i monetària, el sorgiment d'una crisi borsària, l'esclat d'una guerra, un canvi de política fiscal o monetària, etc).

De manera similar, en cas de treballar amb dades de tall transversal, pot no complir-se la hipòtesi d'estabilitat estructural si la mostra està composada per dos o més subgrups d'individus amb comportaments diferents entre sí (per exemple, la incorporació dins d'una mateixa mostra de països industrials i països perifèrics, etc).

# Canvi Estructural

## Conseqüències

En aquest tipus de situacions, quines **conseqüències tindria sobre l'estimació del model** la no consideració de l'existència de canvi estructural?

L'existència de canvi estructural provoca:

- Estimadors MQO esbiaixats, inconsistents i ineficients.
- Inferència errònia (contrastos de significació afectats).
- Prediccions poc fiables i ajust molt baix.

# Canvi Estructural

## Test de CHOW

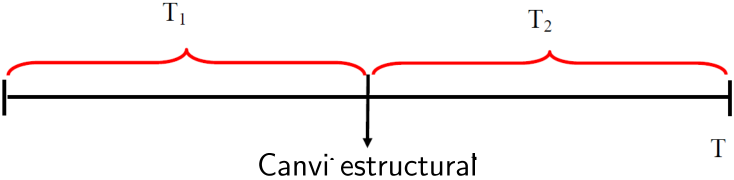
### Contrast de CHOW

$$\begin{cases} H_0: & \text{Permanència o estabilitat estructural} \\ H_A: & \text{Canvi estructural} \end{cases}$$

Partim del següent model:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t \quad t = 1, 2, \ldots, T$$

i sospitem d'un possible canvi estructural a partir de l'observació $T_1 + 1$ (essent $T = T_1 + T_2$).



# Canvi Estructural

## Test de CHOW

Fent ús del contrast de restriccions lineals, construirem el test de Chow seguint els passos següents:

1r) Estimar el model de regressió per tot el període mostral i obtenir la suma dels errors al quadrat:

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + u_t \quad {\scriptstyle t=1,2,\ldots,T} \quad \Rightarrow \mathsf{SQE}_T$$

2n) Estimar el model de regressió pel primer subperíode (amb les primeres $T_1$ observacions) i obtenir la suma dels errors al quadrat:

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + u_t \quad {\scriptstyle t=1,2,\ldots,T_1} \quad \Rightarrow \mathsf{SQE}_{T_1}$$

3r) Estimar el model de regressió pel segon subperíode (amb les $T_2$ observacions restants) i obtenir la suma dels errors al quadrat:

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + u_t \quad {\scriptstyle t=T_1+1,\ldots,T} \quad \Rightarrow \mathsf{SQE}_{T_2}$$

# Canvi Estructural

## Test de CHOW

4t) Construir el estadístic de prova pel contrast de Chow:

$$\mathsf{F}_0 = \frac{\mathsf{SQE}_T - (\mathsf{SQE}_{T_1} + \mathsf{SQE}_{T_2})/k}{(\mathsf{SQE}_{T_1} + \mathsf{SQE}_{T_2})/T - 2k} \sim F_{k,T-2k;\alpha}$$

Si $\mathsf{F}_0 \geq F_{k,T-2k;\alpha}$ Rebuig $H_0$ $\Longrightarrow$ Canvi estructural

Si $\mathsf{F}_0 < F_{k,T-2k;\alpha}$ No Rebuig $H_0$ $\Longrightarrow$ Estabilitat estructural

En cas de rebutjar la hipòtesis nul·la i concloure a favor de l'existència de canvi estructural, hauríem de quedar-nos amb els resultats de les estimacions per cada subperíode per separat.

# Canvi Estructural

## Test de CHOW

El test de Chow pot ser emprat per contrastar l'existència de dues o més ruptures estructurals. Així, en cas d'existir dues possibles ruptures (per tant, amb tres subperíodes potencialment diferents composats per $T_1, T_2$ i $T_3$ observacions respectivament), l'estadístic de prova seria:

$$\mathsf{F}_0 = \frac{\mathsf{SQE}_T - (\mathsf{SQE}_{T_1} + \mathsf{SQE}_{T_2} + \mathsf{SQE}_{T_3})/k}{(\mathsf{SQE}_{T_1} + \mathsf{SQE}_{T_2} + \mathsf{SQE}_{T_3})/T - 3k} \sim F_{k,T-3k;\alpha}$$

# Canvi Estructural

## Test de CHOW: limitacions

### Limitacions del Test de Chow

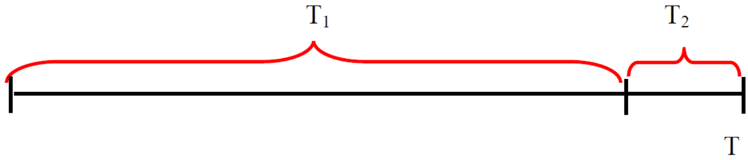
1) Pot captar altres errors d'especificació en la part determinista (errors en la forma funcional o omissió de variables rellevants,...) i acabar detectant un canvi estructural espuri. Per tant, en presència d'altres errors d'especificació, disminueix la seva potència [potència d'un contrast: probabilitat de rebutjar la hipòtesi nul·la essent falsa (1-Prob(error tipus $II$))].
2) És sensible a l'existència de heteroscedasticitat en el terme de pertorbació. En cas d'existir, s'hauria de corregir abans d'analitzar l'estabilitat estructural del model.
3) En ocasions no es disposa d'informació exacta per saber en quin moment precís es produeix el canvi estructural.

# Canvi Estructural

## Test de CHOW: limitacions

4) Perd potència quan el moment de l'hipotètic canvi estructural es troba molt proper a algun dels extrems:

De vegades pot succeir que en una de les submostres hi hagi un número insuficient d'observacions, és a dir, que hi hagi menys observacions que paràmetres a estimar $(T_2 < k)$.




En aquest cas, podríem comparar únicament els errors obtinguts estimant el model amb tota la mostra amb els que s'obtindrien d'estimar-lo emprant el subperíode amb més observacions.

# Canvi Estructural

## Test de CHOW: limitacions

És a dir, en aquest cas l'estadístic de prova pel contrast de Chow serà:

$$\mathsf{F}_0 = \frac{\mathsf{SQE}_T - \mathsf{SQE}_{T_1} / T_2}{\mathsf{SQE}_{T_1} / T - k} \sim F_{T_2, T_1 - 2k; \alpha}$$

Si $\mathsf{F}_0 \geq F_{T_2, T_1 - 2k; \alpha}$ Rebuig $H_0$ $\Longrightarrow$ Canvi estructural

Si $\mathsf{F}_0 < F_{T_2, T_1 - 2k; \alpha}$ No Rebuig $H_0$ $\Longrightarrow$ Estabilitat estructural

# Residus recursius i contrastos d'estabilitat

## 5. Residus recursius i contrastos d'estabilitat

A més del Test de Chow, també podem contrastar l'estabilitat del model, i per tant, l'absència de canvi estructural, mitjançant els anomenats **residus recursius** i els estadístics CUSUM i CUSUMQ.

El **residu recursiu** corresponent a l'observació $t$ es defineix com l'error de predicció de $y_t$ utilitzant l'estimador MQO obtingut amb les observacions $1, 2, \ldots, t-1$.

$$\hat{e}_t = y_t - x_t' \hat{\beta}_{t-1}$$

# Residus recursius i contrastos d'estabilitat

La variància de l'error de predicció és:

$$\text{Var}(\hat{e}_t) = \sigma_u^2 \left[ 1 + x_t'(X_{t-1}'X_{t-1})^{-1} x_t \right]$$

on $X_{t-1}$ és una matriu $(t-1) \times k$ amb les $(t-1)$ primeres observacions de les variables explicatives.

Definim el **residu recursiu normalitzat** com:

$$\hat{w}_t = \frac{\hat{e}_t}{\sqrt{1 + x_t'(X_{t-1}'X_{t-1})^{-1} x_t}}$$

que sota la hipòtesi d'estabilitat i de normalitat del terme de pertorbació es distribueix segons $\hat{w}_t \sim N(0, \sigma_u^2)$ i no està autocorrelacionat.

# Residus recursius i contrastos d'estabilitat

L'estadístic CUSUM és:

$$\text{CUSUM}_t = \sum_{\text{r=k+1}}^{\text{t}} \frac{\hat{w}_r}{\hat{\sigma}}$$

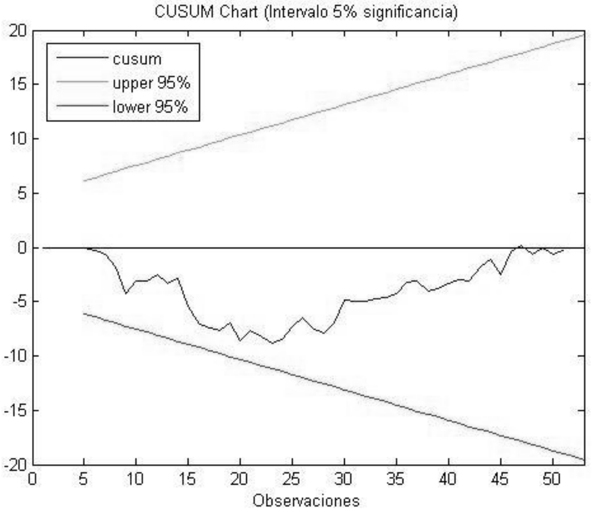
$$\hat{\sigma}^2 = \frac{\sum_{\text{r=k+1}}^{\text{t}} (\hat{w}_r - \bar{w})^2}{T-k} \qquad \bar{w} = \frac{\sum_{\text{r=k+1}}^{\text{t}} \hat{w}_r}{T-k}$$

El contrast es realitza amb el gràfic dels $\text{CUSUM}_t$ a través del temps, mitjançant unes bandes de confiança que, si són superades, comporten el rebuig de l'estabilitat.

Les bandes són: $\pm \left[ a\sqrt{T-k} + 2\frac{(t-k)}{\sqrt{T-k}} \right]$

$$\alpha = 1\,\%\; a = 1{,}143 \qquad \alpha = 5\,\%\; a = 0{,}948$$

# Residus recursius i contrastos d'estabilitat
## CUSUM



# Residus recursius i contrastos d'estabilitat

L'estadístic CUSUMQ és:

$$\text{CUSUMQ}_t = \frac{\sum_{\text{r=k}}^{\text{t}} \tilde{\hat{w}}_r^2}{\sum_{\text{r=k}}^{\text{T}} \tilde{\hat{w}}_r^2}$$

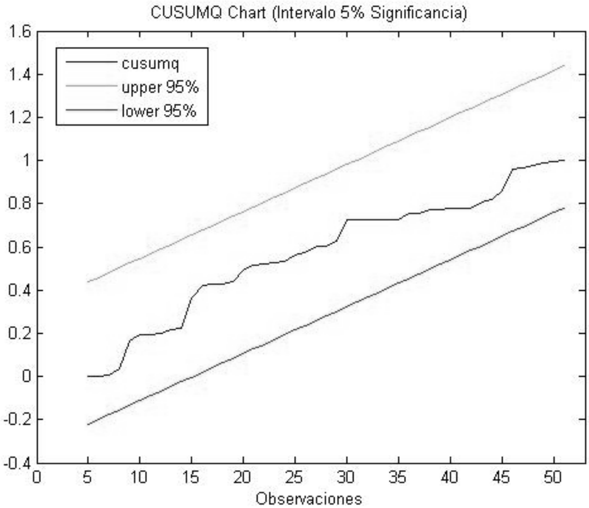
Novament el contrast es realitza amb el gràfic dels $\text{CUSUMQ}_t$ a través del temps, mitjançant unes bandes de confiança que, si són superades, comporten el rebuig de l'estabilitat.

Ara les bandes són: $\pm\left[a + 2\frac{(t-k)}{(T-k)}\right]$

i els valors d'$a$ es troben tabulats per diversos $\alpha$.

# Residus recursius i contrastos d'estabilitat

## CUSUMQ



# Econometria
# Tema 4(1): Errors d'Especificació i Canvi Estructural

Ramon Alemany

Grau Estadística UB-UPC

**Curs 2017-18**