Econometria II
Part 4: Autocorrelació

Riste Gjorgjiev
pareto.uab.es/~rgjorgjiev/econ_cat

Lorenzo Burlon
idea.uab.es/lburlon/teaching_econometria2.html

Universitat Autònoma de Barcelona

Autocorrelació

L'objectiu d'aquest tema és poder contestar les següents preguntes

- Què és l'Autocorrelació
- Quins són els seus efectes al MQO
- Què són sèries temporals
- Com podem detectar l'autocorrelació
- Què podem fer per evitar els efectes negatius

Un dels supòsits del MLG és $cov(u_i, u_j) = 0,\ i \neq j$

- les pertorbacions no són correlacionades

Què vol dir aquest supòsit en les regressions següents:

- producció sobre capital i treball (dades temporals)
- gastes en consum sobre ingrès (dades transversals)

En el cas contrari, diem que hi ha autocorrelació

Definició

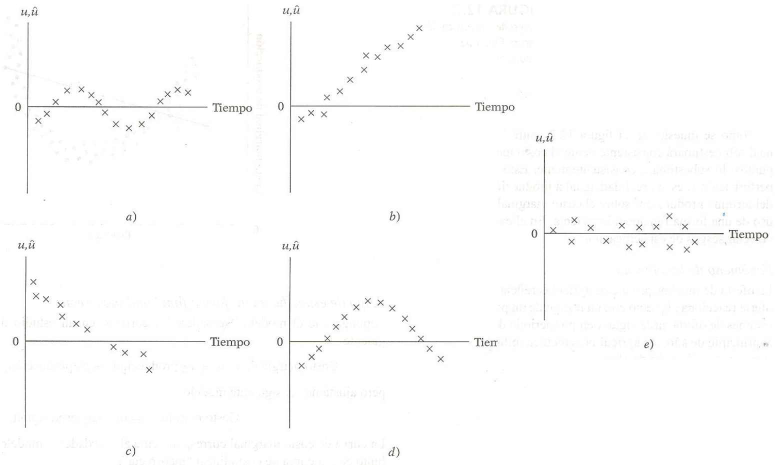
L'existència de correlació entre les perorbecions d'un model,

$$cov(u_i, u_j) \neq 0$$

es diu autocorrelació

Què vol dir l'autocorrelació en els exemples anteriors?

Exemples de pertorbacions correlacionades



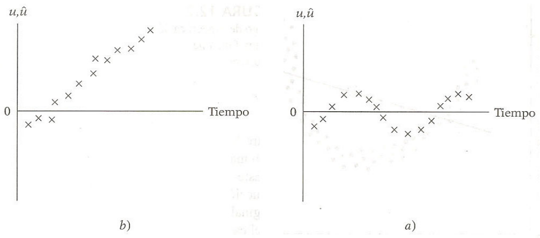
Raons per l'autocorrelació

- inèrcia de les variables econòmiques
- errors d'especificació: exclució de variables

$$\text{fer la regressió } Y = \beta_1 + \beta_2 X_2 + \nu \text{, quan}$$

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

  - aquí $\nu = \beta_3 X_3 + u$
  - segons dels valors de la variable X_3, ν podrà tenir una forma com els gràfics anteriors. Per exemple, quan serien com els següents:

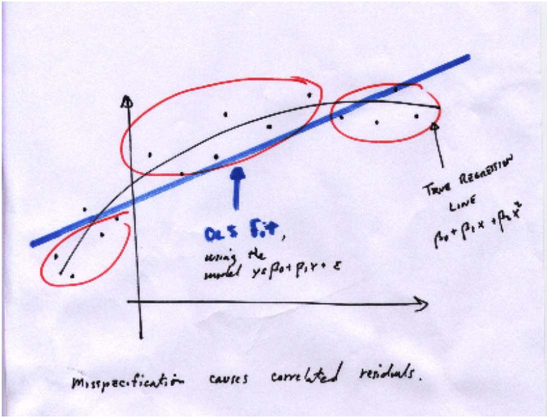


- errors d'especificació: forma funcional incorrecta

$$\text{fer la regressió } Y = \beta_1 + \beta_2 X + \nu, \text{ quan}$$

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + u$$

aquí tenim les següents pertorbacions



- l'efecte de teranyina
  - les empreses decideixen per l'oferta d'un producte segons el preu del mateix en el periode anterior

$$Q_t = \beta_1 + \beta_2 P_{t-1} + u$$

- transformació de les dades

el model $Y_t = \beta_1 + \beta_2 X_t + u_t$ en el periode anterior és

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1}$$

si fem una regressió utilitzant les primeres diferències Δ,

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t = \beta_2 \Delta X_t + \nu_t \tag{1}$$

les pertorbacions ν_t en l'equació (1) són correlacionades. Per què?

Exemples de pertorbacions correlacionades

Com podem generar una sèrie correlacionada, $u_t = u_{t-1} + v$

```
nulldata 100
series v=normal()*1
series u=0
series u=u(-1)+v
```

Implicacions a l'estimador MQO

Imaginem que el model

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

compleix tots els supòsits del MLG, menys el de pertorbacions no correlacionades

$$cov(u_t, u_s) = \sigma_{ts} \quad \text{(quan seria } cov(u_s, u_t)\text{?)}$$

Calculem l'esperança i la variància del $\hat{\beta}_{MQO}$

- $E(\hat{\beta}_{MQO}) = \beta$
- $var(\hat{\beta}_{MQO}) = \sigma^2 (X'X)^{-1} X' E(uu') X (X'X)^{-1}$

$$\Sigma = E(uu') = \begin{pmatrix} \sigma^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma^2 & \dots & \sigma_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma^2 \end{pmatrix}$$

- $var(\hat{\beta}_{MQO}) = \sigma^2(X'X)^{-1}X'\Sigma X(X'X)^{-1}$
- en canvi, quan $cov(u_t, u_s) = 0$, $var(\hat{\beta}_{MQO}) = \sigma^2(X'X)^{-1}$
- $\hat{\beta}_{MQO} \sim N(\beta, var(\hat{\beta}))$

Aleshores:

- $\frac{SQR}{T-k}(X'X)^{-1}$ no és un estimador de $var(\hat{\beta}_{MQO})$
- els mètodes habituals per fer contrastes no són aplicables

Problemes?

- fins aquest punt, no podem estimar la variància del $\hat{\beta}_{MQO}$
- no podem fer cap contraste

Els problemes anteriors apareixen per la matriu Σ. No tenim gens informació sobre les seves elements. Doncs,

- per poder estimar la variància del $\hat{\beta}_{MQO}$, tant com
- introduir el millor estimador sense biaix per un model amb autocorrelació,

necessitem un altre supòsit sobre el tipus d'autocorrelació de les pertorbacions

Introduïm unes sèries temporals amb les seves propietats

Sèries Temporals

Definició

Un procés estocàstic és una seqüència de variables aleatòries.

- Utilitzem les lletres minúscules per les variables aleatòries, u, v, w

Definició

La realització d'una variable aletoria durant el temps representa una sèrie temporal

- L'index de cada variable representa el temps quan s'hagi realitzat, u_t, v_{t-1}, w_1

Si $\{u_t\}_t$ és una sèrie temporal, utilitzem la següent notació

- $E(u_t) = \mu_t$
- $var(u_t) = \gamma_{t,t}$
- $cov(u_t, u_s) = E\left[(u_t - E(u_t))(u_s - E(u_s))\right]^1 = \gamma_{t,s}$

Definició

Un procés estocàstic u és estacionari, si

- μ_t, $\gamma_{t,t}$ *i* $\gamma_{t,s}$ *existeixen per tots els t*
- μ_t *i* $\gamma_{t,t}$ *són independents del t*
- $\gamma_{t,s}$ *només depèn de* $|t-s|$ *però no del t*

Per un procés estacionari, escrivim, $\mu_t = \mu$, $\gamma_{t,t} = \gamma_0$ i $\gamma_{t,s} = \gamma_k$, on $k = |t-s|$ i definim la seva funció de correlació

- $corr(u_t, u_s) = \frac{cov(u_t,u_s)}{\sqrt{var(u_t)var(u_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} = \frac{\gamma_k}{\gamma_0} = \rho_k$

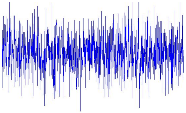
[1] recordeu que $E[(u_t - E(u_t))(u_s - E(u_s))] = E(u_t u_s) - E(u_t)E(u_s)$

Exemple d'un procés estacionari

El procés estocàstic ϵ_t es diu soroll blanc amb variància σ^2 si ϵ_t és un procés estacionari amb

- $E(\epsilon_t) = 0, \forall t$
- $var(\epsilon_t) = \sigma_\epsilon^2, \forall t$
- $cov(\epsilon_t, \epsilon_{t-k}) = 0, \forall k \neq 0$

Exemple d'una realització d'un soroll blanc



Quins problemes tindríem si un soroll blanc representés les pertorbacions d'un model?

Exercici 1

Analitzar si els processos són estacionaris, i en el cas positiu, determinar la esperança, variància i covariància:

1. $\nu_t = \epsilon_{1t} + t\epsilon_{2t}$, on ϵ_{1t} i ϵ_{2t} són sorolls blancs independents
2. $z_t = \beta_1 + \beta_2 t + \delta_t$, on δ_t és un soroll blanc
3. $\omega_t = z_t - z_{t-1}$, z_t de l'apartat anterior

Exercici 2

$y_t = z + \epsilon_t$, on ϵ_t és un soroll blanc amb variància σ_ϵ^2. z no canvia amb el temps, $E(z) = 0$, $var(z) = \sigma_z^2$ i no està correlacionada amb ϵ_t. V o F

1. y_t és estacionari
2. $\rho_k = corr(y_t, y_{t-k})$ disminueix amb k

1. Sèries Autoregressius, (AR)

Per una sèrie temporal u_t diem que és autoregressiu d'ordre p, AR(p) si

$$u_t = c + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \ldots + \phi_p u_{t-p} + \epsilon_t$$

on c, ϕ_1, ..., ϕ_p són constantes i $\{\epsilon_t\}$ és un soroll blanc. En particular, un procés u_t és AR(1),

$$u_t = c + \phi_1 u_{t-1} + \epsilon_t$$

El paràmetre ϕ es diu el coeficient de correlació.
Mirem a les propietats d'un procés AR(1)

Un procés u_t es pot escriure com

- $u_t = c + \phi u_{t-1} + \epsilon_t = c + \phi(c + \phi u_{t-2} + \epsilon_{t-1}) + \epsilon_t = c(1+\phi) + \phi^2 u_{t-2} + \epsilon_t + \phi\epsilon_{t-1}$
- si seguíssim fer les maiteixes substitucions $T-1$ vegades, tindríem
- $u_t = c(1 + \phi + \ldots + \phi^{T-1}) + \phi^T u_{t-T} + \epsilon_t + \phi\epsilon_{t-1} + \ldots + \phi\epsilon_{t-(T-1)}$
- que també es pot escriure com

$$u_t = c\sum_{i=0}^{T-1} \phi^i + \sum_{i=0}^{T-1} \phi^i \epsilon_{t-i} + \phi^T u_{t-T}$$

Si $|\phi| < 1$, quan $T \to \inf$, $\phi^T u_{t-T} \to 0$, doncs

$$u_t = c\sum_{i=0}^{\infty} \phi^i + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} = \frac{c}{1+\phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} \tag{2}$$

A partir de l'equació (2) podem derivar les expresions de l'esperança, variància i covariància del procés AR(1)

- $\mu_t = E(u_t) = E(\frac{c}{1+\phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}) = \frac{c}{1+\phi} + \sum_{i=0}^{\infty} \phi^i E(\epsilon_{t-i}) = \frac{c}{1+\phi}$
- $\gamma_{t,t} = var(u_t) = var(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}) = \sum_{i=0}^{\infty} \phi^{2i} var(\epsilon_{t-i})$
    - $= \sigma_\epsilon^2 \sum_{i=0}^{\infty} \phi^{2i} = \frac{\sigma_\epsilon^2}{1-\phi^2}$
- $\gamma_{t,t-k} = cov(u_t, u_{t-k}) = cov(\sum \phi^i \epsilon_{t-i}, \sum \phi^i \epsilon_{t-k-i})$

recordem que ϵ_t és soroll blanc, $cov(\epsilon_t, \epsilon_{t-k}) = 0$, per $k \neq 0$

- $cov(u_t, u_{t-k}) = cov(\phi^k \sum \phi^i \epsilon_{t-k-i}, \sum \phi^i \epsilon_{t-k-i}) = \phi^k var(u_{t-k}) = \phi^k \frac{\sigma_\epsilon^2}{1-\phi^2}$

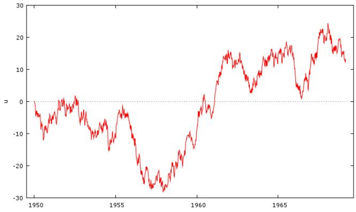
Doncs, per un procés AR(1), sabem que $\mu_t = \frac{c}{1+\phi}$, $\gamma_{t,t} = \frac{\sigma_\epsilon^2}{1+\phi^2}$ i $\gamma_{t,s} = \phi^{t-s} \frac{\sigma_\epsilon^2}{1+\phi^2}$

$\Rightarrow$ AR(1) és un procés estacionari si $|\phi| < 1$

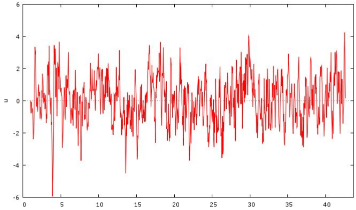
La funció de correlació és

$$\rho_k = corr(u_t, u_{t-k}) = \frac{\gamma_k}{\gamma_0} = \phi^k$$

Mirem al gràfic del procés AR(1) utilitzat abans, $u_t = 1 * u_{t-1} + \nu$



És estacionari? Si $\rho = 0.7$, el gràfic serà



2. Sèries Mitjana Mòbil, (MA)

Per una sèrie temporal u_t diem que és mitjana mòbil d'ordre q, MA(q) si

$$u_t = c + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

on c, θ_1, ..., θ_q són constantes i $\{\epsilon_t\}$ és un soroll blanc. En particular, un procés u_t és MA(1),

$$u_t = c + \epsilon_t - \theta \epsilon_{t-1}$$

Com en el cas amb AR(1), derivar les expresions de l'esperança, variància i covariància del procés MA(1)

- $\mu_t = E(u_t) = E(c + \epsilon_t - \theta\epsilon_{t-1}) = c + E(\epsilon_t) - \theta E(\epsilon_{t-1}) = c$
- $\gamma_{t,t} = var(u_t) = var(c + \epsilon_t - \theta\epsilon_{t-1}) = var(\epsilon_t - \theta\epsilon_{t-1})$
  $= var(\epsilon_t) + var(\theta\epsilon_{t-1}) - 2cov(\epsilon_t, \theta\epsilon_{t-1}) = (1 + \theta^2)\sigma_\epsilon^2$
- $\gamma_{t,t-1} = cov(u_t, u_{t-1}) = cov(c + \epsilon_t - \theta\epsilon_{t-1}, c + \epsilon_{t-1} - \theta\epsilon_{t-2})$
  $= \ldots = -\theta\sigma_\epsilon^2$
- $\gamma_{t,t-k} = cov(u_t, u_{t-k}) = cov(c + \epsilon_t - \theta\epsilon_{t-1}, c + \epsilon_{t-k} - \theta\epsilon_{t-k-1}) = 0$ per $k \geq 2$

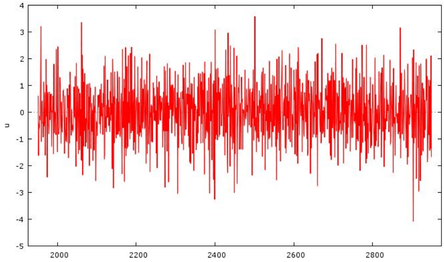
Aquí també veiem que μ_t, $\gamma_{t,t}$ i $\gamma_{t,s}$ existeixen, i a més no depenen de t

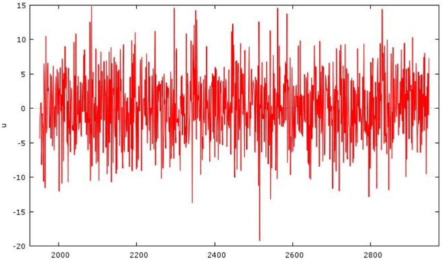
$\Rightarrow$ MA(1) és un procés estacionari

La funció de correlació és

$$\rho_1 = corr(u_t, u_{t-1}) = -\frac{\theta}{1+\theta^2}, \text{ i } \rho_k = 0, \text{ per } k \geq 2$$

Mirem al gràfic de dos MA(1) procesos. Al primer $\theta = 0.5$ i al segon $\theta = 5$





- l'esperança dels dos és igual a zero
- la variància del primer és $(1 + 0.5^2) * 1$
- la variància del segon és $(1 + 5^2) * 1$

2. Processos mixts, (ARMA)

Per una sèrie temporal u_t diem que és un procés, ARMA(p,q) si

$$u_t = c + \phi_1 u_{t-1} + \ldots + \phi_p u_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q}$$

En particular, un procés u_t és ARMA(1,1) si

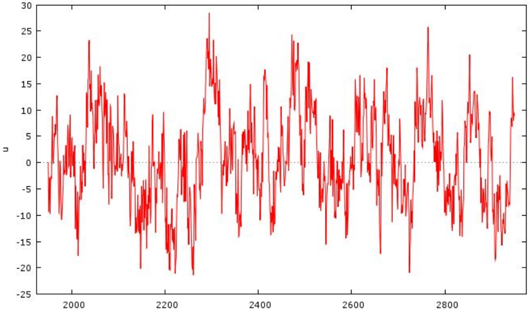
$$u_t = c + \phi u_{t-1} + \epsilon_t - \theta \epsilon_{t-1}$$

Sense donar els detalls de la derivació, aquí tenim els valors del μ i de les γ_i

Si $|\phi| < 1$, el procés ARMA(1,1)

- és estacionari
- $\mu = \frac{c}{1-\phi}$
- $\gamma_0 = \frac{\sigma_\epsilon^2(\theta^2 - 2\phi\theta + 1)}{1-\phi^2}$
- $\gamma_1 = \sigma_\epsilon^2 \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}$
- $\gamma_k = \phi\gamma_{k-1}$, per $k \geq 2$

Aquí hi ha el gràfic de la sèrie $u_t = 0.9u_{t-1} + \nu + 5\nu_{t-1}$, on $\nu \sim N(0,1)$



series v=normal()*1
series u=0
series u=0.9*u(-1)+v-5*v(-1)

Example: Estimació amb pertorbacions AR(1)

Per simplificar el problema, imaginem model amb una variable explicativa X

$$Y = \beta_1 + \beta_2 X + u$$

- u és $AR(1)$, $u_t = \rho u_{t-1} + \epsilon_t$
- el valor de ρ és conegut
- per mantenir el supòsit $E(u_i) = 0$, posem $c = 0$

Aquí corregim el model per evitar el problema d'autocorrelació

- Multiplìquem la equació retardada de

$$Y_t = \beta_1 + \beta_2 X_t + u_t \text{ amb } \rho$$

- $\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1}$
- restem la segona de la primera

$$Y_t - \rho Y_{t-1} = \beta_1 - \rho\beta_1 + \beta_2 X_t - \rho\beta_2 X_{t-1} + u_t - \rho u_{t-1}$$

- que també es pot escriure com

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \underbrace{u_t - \rho u_{t-1}}_{=\epsilon_t}$$

- ara tenim un model transformat, que a més està corregit per l'autocorrelació

$$Y_t^* = \beta_1^* + \beta_2 X_t^* + \epsilon_t$$

- $Y_t^* = Y_t - \rho Y_{t-1}, \qquad X_t^* = X_t - \rho X_{t-1}$
- ϵ_t és el soroll blanc del procés AR(1)
- i atenció $\beta_1^* = \beta_1(1-\rho)$

El model transformat compleix tots els supòsits del MLG

- $\hat{\beta} = (X^{*'}X^*)^{-1}X^{*'}Y^*$ és el millor estimador lineal sense biaix pels β_1^* i β_2
- $\hat{\beta}_1 = \hat{\beta}_1^*/(1-\rho 2)$

Definició

Aquest estimador dels paràmetres β es diu Minims Quadrats Generalitzats, $\hat{\beta}_{MQG}$

Nota: Feient la transformació introduida, perdem una observació

- Aquella es pot recuperar posant $Y_1^* = Y_1\sqrt{1-\rho^2}$, i $X_1^* = X_1\sqrt{1-\rho^2}$

MQO amb autocorrelació

- Fem un experiment per veure l'efecte d'utilitzar el MQO a un model amb les pertorbacions AR(1)
- generem les dades
- fem el MQO
  - amb pertorbacions AR(1)
  - amb pertorbacions sense autocorrelació
- comparem les variàncies estimades

- Generem les dos tipus de les pertorbecions segons
- $e \sim N(0,1)$ i $u_t = 0.7u_{t-1} + e_t$
- els valors del X són ordenats
    - els valors del X són ordenats, perque X tingui una correlació mostral positiva ($r_X > 0$)
    - el coeficient de correlació és positiu, $\rho > 0$

En aquest cas, els resultats són Tenim les dades a l'arxiu ex.gdt

$$\widehat{\mathrm{y1}} = \underset{(1.8913)}{10.4306} - \underset{(0.30481)}{0.504408}\,\mathrm{x}$$

$$T = 10 \quad \bar{R}^2 = 0.1619 \quad F(1,8) = 2.7384 \quad \hat{\sigma} = 2.7686$$

$$\widehat{\mathrm{y2}} = \underset{(1.9038)}{5.80747} - \underset{(0.30683)}{0.213991}\,\mathrm{x}$$

$$T = 10 \quad \bar{R}^2 = -0.0605 \quad F(1,8) = 0.48640 \quad \hat{\sigma} = 2.7869$$

Conclusó: El MQO subestima la variància, $\frac{\sum \hat{u}_2}{T-k} < var(\hat{\beta})$ quan les pertorbacions són AR(1), $r_X > 0$ i $\rho > 0$