

# Econometria II

## Part 4: Autocorrelació - Part II

Riste Gjorgjiev  
[pareto.uab.es/~rgjorgjiev/econ\\_cat](http://pareto.uab.es/~rgjorgjiev/econ_cat)

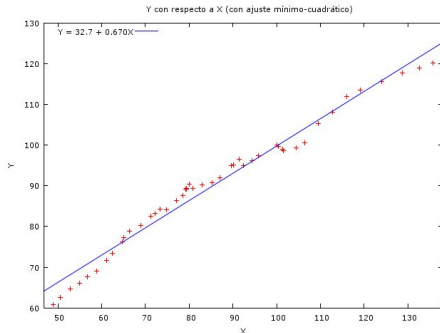
Lorenzo Burlon  
[idea.uab.es/lburlon/teaching\\_econometria2.html](http://idea.uab.es/lburlon/teaching_econometria2.html)

Universitat Autònoma de Barcelona

# Detecció de l'autocorrelació

## Un exemple de motivació

- El fitxer Table 12.4.txt conté les dades anuals dels EEUU sobre producció per hora ( $X$ ) i l'índex del salari real ( $Y$ )
- al següent gràfic hi ha la relació entre les dues variables



- alguna indicació per la relació?

Els resultats de la estimació del model lineal i log-lineal són

$$\widehat{Y} = \underset{(1.3940)}{32.7419} + \underset{(0.015671)}{0.670406} X$$

$$T = 46 \quad \bar{R}^2 = 0.9760 \quad F(1, 44) = 1830.2 \quad \hat{\sigma} = 2.3845$$

$$\widehat{\log(Y)} = \underset{(0.054709)}{1.60668} + \underset{(0.012353)}{0.652216} \log(X)$$

$$T = 46 \quad \bar{R}^2 = 0.9841 \quad F(1, 44) = 2787.8 \quad \hat{\sigma} = 0.022078$$

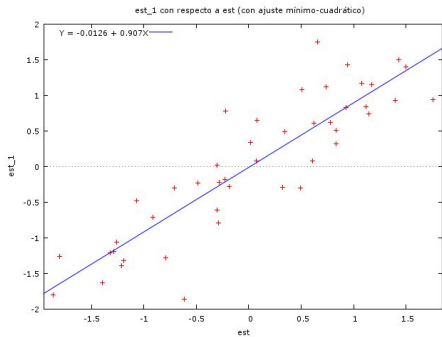
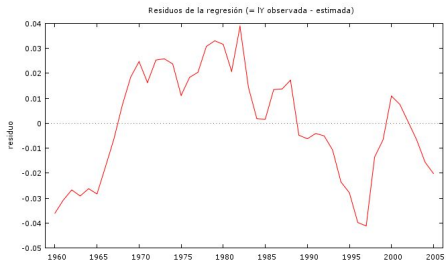
- per verificar la fiabilitat d'aquest resultats hem de contestar si les pertorbacions són correlacionats
- per tant hem d'introduir mètodes per detectar l'autocorelació

## 1. Mètode Gràfico

Per tenir una indicació per presència d'autocorrelació, podem

- fer el gràfic dels  $\hat{u}$  sobre el temps
  - en aquest cas  $\hat{u}$  són les estimacions de  $u$
- fer el gràfic de  $\hat{u}_t$  sobre  $\hat{u}_{t-1}$ 
  - en aquest cas sospitem AR(1)

Mirem als dos gràfic del exemple anterior



## 2. Prova de les ratxes

- mirem una altra vegada al gràfic



- ens capfiquem al signe dels residus
- una sèrie successiva que no canvia el signe es diu una ratxa
- el numero de residus en una ratxa és la longitud de la ratxa

Al gràfic de l'exemple

- hi ha 5 ratxes
- amb longitud 8, 21, 11, 3, 3

L'hipòtesi d'aquesta prova és

$$H_0 : \text{No hi ha autocorrelació}$$

- la decisió es pren utilitzant la distribució de la variable  $R$ -numero de ratxes
- $T$  - numero d'observacions
- $T_1, T_2$  - numero de símbols (+) i (-)

Sota el supòsit de pertorbacions no correlacionats

- $E(R) = \frac{2T_1T_2}{T} + 1$
- $\sigma_R^2 = \text{var}(R) = \frac{2T_1T_2(2T_1T_2 - T)}{T^2(T-1)}$

- si  $R \in [E(R) - 1.96\sigma_R, E(R) + 1.96\sigma_R]$ 
  - $\Rightarrow$  no rebutgem l' $H_0$  al nivell de 95%

Tornem a l'exemple

- $T_1 = 24$ ,  $T_2 = 22$  i  $R = 5$ , però
- $E(R) = 24$ ,  $\sigma_R = 3.32$
- l'interval de confiança és
$$[24+1.96*3.32, 24-1.96*3.32]=[17.5, 30,5]$$
- veiem que  $R \notin [17.5, 30,5]$

$\Rightarrow$  Rebutgem l' $H_0$



### 3. Prova de Durbin-Watson (d)

Al quadre dels resultats d'una estimació amb Gretl

Modelo 1: MCO, usando las observaciones 1960–2005 ( $T = 46$ )  
Variable dependiente: lnY

	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	1.60668	0.0547085	29.3680	0.0000
lnX	0.652216	0.0123527	52.7996	0.0000
Media de la vble. dep.	4.490151	D.T. de la vble. dep.	0.175142	
Suma de cuad. residuos	0.021448	D.T. de la regresin	0.022078	
$R^2$	0.984462	$R^2$ corregido	0.984109	
$F(1, 44)$	2787.801	Valor p (de $F$ )	1.96e-41	
Log-verosimilitud	111.1567	Criterio de Akaike	-218.3134	
Criterio de Schwarz	-214.6561	Hannan-Quinn	-216.9433	
$\hat{\rho}$	0.867892	Durbin-Watson	0.217558	

hi ha l'estadístic  $d$ . Aquí expliquem com s'utilitza per detectar l'autocorrelació

- l'estadístic  $d$  es calcula  $d = \frac{\sum_2^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_1^T \hat{u}_t^2}$

Aquest estadístic s'utilitza per poder rebutjar o no la següent hipòtesi

$$H_0 : u_t \text{ és un procés AR}(1), u_t = \rho u_{t-1} + \epsilon_t$$

Per poder utilitzar-lo, és important que es verifiquin els supòsits

- 1  $u_i \sim N(0, \sigma^2)$
- 2 el model no conté retards d' $Y$  com una variable explicativa
- 3 no hi ha observacions que manquen

## la relació entre $d$ i $\rho$

- A part de la contesta sobre l' $H_0$ , l'estadístic  $d$  ens dona una altra informació
- segons la relació  $d \approx 2(1 - \hat{\rho})$ , una vegada tenint l'estadístic  $d$ , tenim el valor estimat de  $\rho$ , en el cas quan el necessitem

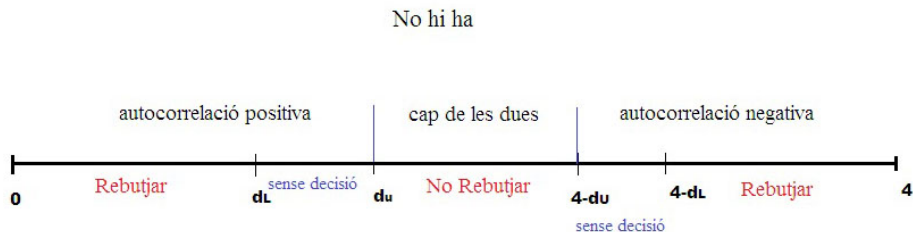
$$d = \frac{\sum(\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2} = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2 \sum \hat{u}_{t-1} \hat{u}_t}{\sum \hat{u}_t^2} \approx 2 \left( 1 - \frac{\sum \hat{u}_{t-1} \hat{u}_t}{\sum \hat{u}_t^2} \right)$$
$$= 2(1 - \hat{\rho}), \text{ perquè } \rho = \frac{E(u_t u_{t-1})}{\text{var}(u_t)}$$

- sabem que un AR(1) és estacionari quan  $|\rho| < 1$
- això implica que  $d \in [0, 4]$
- què vol dir quan  $d \approx 0$ ,  $d \approx 2$  i  $d \approx 4$ ?

- La decisió per rebutjar o no l' $H_0$  depèn de dos valors crítics,  $d_L$  i  $d_U$
- $d_L$  i  $d_U$  depenen només de numero d'observacions  $T$  i numero de variables explicatives  $k$
- i es poden trobar al quadre de la distribució D-W

Observations		X variables,									
		1		2		3		4		5	
N	Prob	D-L	D-U	D-L	D-U	D-L	D-U	D-L	D-U	D-L	D-U
15	0.05	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
	0.01	0.81	1.07	0.7	1.25	0.59	1.46	0.49	1.70	0.39	1.96
20	0.05	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
	0.01	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
25	0.05	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
	0.01	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
30	0.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
	0.01	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
40	0.05	1.44	1.54	1.39	1.60	1.34	1.66	1.39	1.72	1.23	1.79
	0.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
50	0.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
	0.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
60	0.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
	0.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
80	0.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
	0.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
100	0.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78
	0.01	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

La decisió es pren segons el dibuix



- A l'exemple anterior,  $T=46$ ,  $k=2$  i hem vist que  $d=0.232$
- $d_L = 1.43$ ,  $d_U = 1.615$

⇒ Rebutgem que no hi ha autocorrelació positiva

- quin és el valor estimat del  $\rho$ ?

Hi ha dues hipòtesis alternatives que es poden contrastar utilitzant el DW

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

$$H_0 : \rho = 0$$

$$H_1 : \rho < 0$$

No Rebutgem

$$d > d_U$$

$$d^* > d_U$$

Rebutgem

$$d < d_L$$

$$d^* < d_L$$

No es pot decidir

$$d_L < d < d_U$$

$$d_L < d^* < d_U$$

## 4. Prova de Breusch-Godfrey (BG)

Per demostrar la prova de BG, utilitzem el model

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

Suposem que

$$u_t = \rho_1 u_{t-1} + \dots + \rho_p u_{t-p} + \epsilon_t$$

on  $\epsilon_t$  és un soroll blanc. L'hipòtesi per contestar és

$$H_0 : \rho_1 = \dots = \rho_p = 0$$

Els passos per construir l'estadístic són

1 estimar  $Y_t = \beta_1 + \beta_2 X_t + u_t$

2 guardar els residus  $\hat{u}_t$

3 fer la regressió auxiliar

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 u_{t-1} + \dots + \hat{\rho}_p \hat{u}_{t-p} + \epsilon_t$$

4 guardar el  $R^2$  de l'última regressió

5 sota l' $H_0$ ,  $(T - p)R^2 \sim \chi_p^2$

6 conclusió:

$$\text{si } (T - p)R^2 > \chi_{p,\alpha}^2 \Rightarrow \text{Rebutgem l}'H_0$$



## Exemple - continuació

Fem la prova de BG per l'exemple anterior per  $p = 1$ . Els resultats de la regressió auxiliar són

$$\widehat{\text{uhat2}} = \underset{(0.025709)}{0.0372990} - \underset{(0.0057902)}{0.00832148} IX + \underset{(0.068064)}{0.869587} \text{uhat21}$$

$$T = 45 \quad R^2 = 0.796691 \quad F(2, 42) = 82.291 \quad \hat{\sigma} = 0.0098684$$

- $T = 46, p = 1, R^2 = 0.796691, \chi_1^2 = 3.8$
- $(T - 1)R^2 = 45 * 0.796691 = 35.8 > \chi_1^2 = 3.8$   
 $\Rightarrow$  rebutgem l'hipòtesi

## 1. Especificació incorrecta

- en l'exemple dels salari real i l'index de producció hem obtingut un valor del DW molt baix,  $d = 0.2176$
- una de les raons per presència d'autocorrelació era un model especificat incorrectament
- aquí mirem si hagim fet aquest tipus d'error en el mateix exemple
- l'elecció de la variable per afegir és difícil i necessita justificació econòmica
- mirem als resultats

$$\widehat{\ln Y} = 1.60668 + 0.652216 \ln X$$

(0.054709)      (0.012353)

$$T = 46 \quad \bar{R}^2 = 0.9841 \quad d = 0.2176 \quad \hat{\sigma} = 0.022078$$

- una possibilitat pot ser que el salari mostra tendència
  - Y està correlacionat amb el temps
- en aquest cas hauríem d'afegir-lo en el model

Modelo 2: MCO, usando las observaciones 1960–2005 ( $T = 46$ )  
 Variable dependiente: lnY

	Coefficiente	Desv. Tpica	Estadstico t	Valor p
const	0.120937	0.307035	0.3939	0.6956
lnX	1.02830	0.0775523	13.2594	0.0000
time	-0.00752822	0.00153941	-4.8903	0.0000
$\hat{\rho}$	0.800477	Durbin-Watson	0.449741	

- aleshores, no evitem el problema d'autocorrelació

- una altra possibilitat seria que la relació entre el salari i la productivitat no fos lineal
- podem afegir la variable dependent quadrada

Modelo 3: MCO, usando las observaciones 1960–2005 ( $T = 46$ )  
Variable dependiente:  $\ln Y$

	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	-1.78430	0.643853	-2.7713	0.0082
$\ln X$	2.19635	0.292692	7.5040	0.0000
$\ln X^2$	-0.175155	0.0331826	-5.2785	0.0000
$\hat{\rho}$	0.825425	Durbin-Watson	0.356051	

- arribem a la mateixa conclusió com abans
- doncs aquí, pot ser que el model no està especificat incorrectament, però
- només hi ha autocorrelació de la naturalesa de les variables i
- necessitem altres mesures per evitar-la

## 2. Correcció de l'autocorrelació: MQG i MQGF

- Hem vist que quan en el model

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

- $u$  és  $AR(1)$ ,  $u_t = \rho u_{t-1} + \epsilon_t$  i
- el valor de  $\rho$  és conegut,

la manera per evitar l'efecte de l'autocorrelació és utilitzar el  $\beta_{MQG}$

- $\hat{\beta}_{MQG} = (X^{*'} X^*)^{-1} X^{*'} Y^*$ , on
- $Y_t^* = Y_t - \rho Y_{t-1}$ ,  $X_t^* = X_t - \rho X_{t-1}$
- el  $\beta_{MQG}$  és aplicable només quan sabem el valor del  $\rho$
- en pràctica, aquest valor no és conegut i per tant
- hem de trobar una manera d'evitar l'autocorrelació

## Mètode de primeres diferències, $\Delta$

- en el model

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

- $u$  és  $AR(1)$ ,  $u_t = \rho u_{t-1} + \epsilon_t$
- si suposem que  $\rho = 1$  podem aplicar el mètode de primeres diferències
- $Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + u_t - u_{t-1}$  o
- $\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t$  ( $\Delta Y_t = \beta_2 \Delta X_t + \epsilon_t$ )
- nota: l'última equació no té constanta
- el model transformat compleix tots els supòsits del MLG
- es poden fer tots els tipus de contrastes
- el problema aquí és el supòsit  $\rho = 1$

- doncs per poder fer servir aquest mètode, hem de verificar que  $\rho = 1$
- en el cas quan  $\rho = 1$ , amb aquesta transformació evitem
  - el problema d'autocorrelació
  - fem la sèrie de les pertorbacions estacionària
- per verificar que  $\rho = 1$ , hem de fer una prova addicional
- per comprovar que  $\rho = 1$ , que ara és la nova hipòtesi, utilitzem l'estadístic

$$g = \frac{\sum_2^T \hat{e}_t^2}{\sum_1^T \hat{u}_t^2}$$

- on  $\hat{u}$  són els residus del model inicial i
- $\hat{e}$  són els residus del model transformat
- la conclusió: la transformació serà vàlida quan  $g < d_L$
- $d_L$  és el valor crític del DW per una variable i T-1 observacions<sup>1</sup>
- fem aquesta prova per l'exemple anterior (a la classe)

<sup>1</sup>aquest criteri es caviaran quan es tracta d'un model amb més que una variable

## $\rho$ estimada del DW

- en el cas quan la correcció de les primeres diferències no és aplicable
  - quan  $\rho \neq 1$
- la manera per estimar  $\rho$  és utilitzar l'estadístic DW de la regressió inicial
- la relació per aprofitar és  $d = 2(1 - \hat{\rho})$  o  $\hat{\rho} = 1 - d/2$
- una vegada tenint  $\hat{\rho}$ , hem de fer la mateixa transformació com el MQG

$$Y_t^* = Y_t - \hat{\rho}Y_{t-1}, \quad X_t^* = X_t - \hat{\rho}X_{t-1} \quad u_t^* = u_t - \hat{\rho}u_{t-1}$$

- i estimar el model  $Y_t^* = \beta_1^* + \beta_2 X_t^* + u_t^*$  amb
- $\hat{\beta}_{MQGF} = (X^{*'} X^*)^{-1} X^{*'} Y^*$

### Definició

*L'últim estimador dels paràmetres  $\beta$  es diu Minims Quadrats Generalitzats Factibles,  $\hat{\beta}_{MQGF}$*



## $\rho$ basat als $\hat{u}$

- si el supòsit de  $u_t = \rho u_{t-1} + \epsilon$  és correcte
- podem estimar el valor de  $\rho$  regressant el  $\hat{u}_t$  sobre  $\hat{u}_{t-1}$ , és a dir
- estimar el model

$$u_t = \rho u_{t-1} + \nu_t$$

- després fem la transformació com al cas anterior
- demostració per l'exemple (a la classe)

### 3. L'estimador de Newey-West

- en els casos quan al model hi ha autocorrelació de forma desconeguda
- Newey i West han proposat una manera d'estimar  $var(\hat{\beta}_{MCO})$
- aquesta matriu estimada,  $\hat{var}_{NW}(\hat{\beta}_{MCO})$ , ens pot servir per fer tots els tipus de contrastes coneguts
- la deducció d'aquest estimador és bastant complicada i ens només veiem com es pot utilitzar en pràctica
- tornem a l'exemple anterior
- per tenir els valors estimats de  $\hat{var}_{NW}(\hat{\beta}_{MCO})$  hem de fer
  - tenir sèries temporals
  - introduir les variables dependent i independent
  - a la mateixa pantalla escullir *desviacions típiques robust*
  - escullir **HAC** on hi ha dades de dèries temporals
- podem utilitzar els valor estimats de les desviacions típiques

# Exercici 1

Es vol analitzar la relació entre el número d'ambientadors venuts per una empresa,  $F_t$ , número de punts de distribució,  $P_t$  i la temperatura mitjà de l'àrea on és l'empresa,  $E_t$ . Utilitzant 50 observacions mensuals s'han obtingut els següents resultats:

$$(1) : F_t = 3.27 + 0.43P_t + res; \quad DW = 0.8846;$$

$$(2) : F_t = \hat{\delta}_1 + 0.51F_{t-1} + \hat{\delta}_3P_t + \hat{\delta}_4P_{t-1} + res;$$

$$(3) : F_t^* = 1.25 + 0.455P_t^* + res; \quad \hat{\sigma} = 1.7;$$

(0.49)    (0.043)

$$(4) : F_t = 2.36 + 0.47P_t + 0.49E_t + \hat{u}_t; \quad DW = 1.77;$$

(4.4)    (19.8)    (4.48)

$$(5) : \hat{u}_t = \hat{\gamma}_1 + \hat{\gamma}_2P_t + \hat{\gamma}_3E_t + \hat{\gamma}_4\hat{u}_{t-1} + \hat{\gamma}_5\hat{u}_{t-2} + \hat{\gamma}_6\hat{u}_{t-3} + res; \quad R^2 = 0.12;$$

- els valors entre parèntesis en (3) són errors estandard i
- de l'equació (4) els estadístics  $t$
- la variable dependent de (5) és el residu de (4)
- les variables de (3) s'han fet segons  $F_t^* = F_t - 0.51F_{t-1}$  i  $P_t^* = P_t - 0.51P_{t-1}$

## Preguntes:

- 1 L'investigador A formula el model

$$F_t = \beta_1 + \beta_2 P_t + u_t$$

Compleix aquest model les hipòtesi de pertorbacions no correlacionades?

- 2 segons els resultats de l'apartat anterior, obtenir estimacions eficients de  $\beta_1$ ,  $\beta_2$  i  $\sigma_u^2$
- 3 contrastar la significança de  $P_t$ . Com podríem der aquest contrast si l'única informació fos que  $u$  és estocastic?
- 4 L'investigador B afeigeix  $F_t$  com una variable dependent, constant,  $P_t$  i  $E_t$  com independents. Té autocorrelació de primer ordre aquest model?
- 5 L'investigador A declara que el model del B no presenta autocorrelació de primer ordre, però si del segon o tercer. Contrastar aquesta afirmació

## Exercici 2

Un grup d'investigadors fa un estudi sobre el preu del peix. Creuen que el preu  $P_t$  es pot explicar d'una mesura de l'altura de les ones del dia anterior  $ola_t$  i una variable fictícia  $vie_t$  que té valor 1 si el dia és divendres i 0 al contrari. El grup estima el model

$$P_t = \beta_1 + \beta_2 ola_t + \beta_3 vie_t + u_t$$

Utilitzant 95 observacions diàries s'han obtingut els resultats

$$(2): P_t = 3.13 + 1.48ola_t + 1.46vie_t + e_t \quad DW = 3.38$$

(0.55) (0.09) (0.43)  
[0.48] [0.09] [0.40]

$$(3): P_t^\bullet = 3.19 + 1.51ola_t^\bullet + 0.91vie_t^\bullet + e_t^\bullet$$

(0.28) (0.05) (0.35)

$$(4): P_t^+ = 3.75 + 1.37ola_t^+ + 1.28vie_t^+ + e_t^+$$

(0.93) (0.14) (0.53)

$$(5): e_t^\bullet = -0.17 - 0.15e_{t-1}^\bullet + 0.02ola_t^\bullet + 0.14vie_t^\bullet + res \quad R^2 = 0.02$$

$$(6): e_t^+ = -0.37 - 0.87e_{t-1}^+ + 0.16ola_t^+ - 0.01vie_t^+ + res \quad R^2 = 0.72$$

on

- els valor entre ( ) són els ee de MQO i entre [ ] de Newey-West
- les variables en (3) són  $P_t^\bullet = P_t + 0.69P_{t-1}\dots$  i
- i de la (4) són  $P_t^+ = P_t - 0.69P_{t-1}\dots$

Preguntes:

- 1 contrastar l'hipòtesi d'ausència d'autocorrelació en el model
- 2 utilitzant l'informació de (2), podem concluir que el preu del dividend és diferent del preu dels altres dies
- 3 un altre grup d'investigadors suggereix que  $u_t$  és un AR(1). Com podríem fer el contraste de l'apartat anterior
- 4 podem concluir que  $u_t$  de veritat és un AR(1)?