Econometría I

Autocorrelación

Carlos Yanes | Departamento de Economía | 2023-04-19

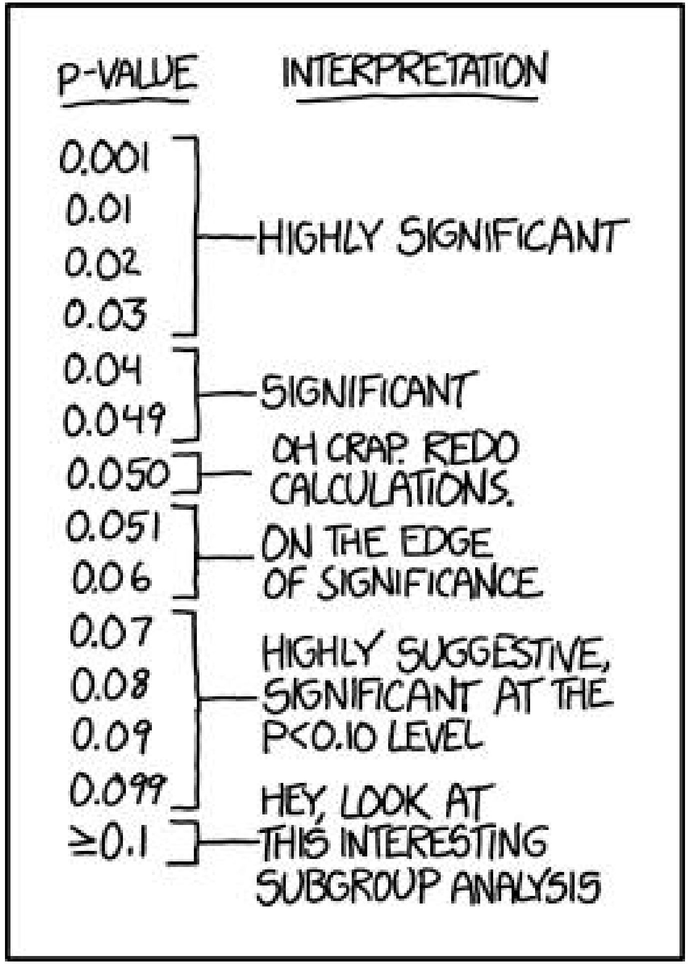
😅

Preguntas de la sesion anterior?


NYJAH HUSTON

Antes de empezar ❗ ...

Recordeis



P-VALUE
INTERPRETATION
0.001
0.01
0.02
0.03
HIGHLY SIGNIFICANT
0.04
0.049
SIGNIFICANT
0.050
OH CRAP. REDO CALCULATIONS.
0.051
0.06
ON THE EDGE OF SIGNIFICANCE
0.07
0.08
0.09
0.099
HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
≥0.1
HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS

Hey Girl,
Don't worry about Collinearity
I'm Independent of my Xs.
imgflip.com

Antes del arranque...

1. Este curso tiene siempre **teoría** por aprender.
2. Tendrá ventajas sobre los demás que hacen data science si conoce o sabe de esto.
3. Los últimos test que hemos estado aprendiendo no son nada **fáciles**, requieren de saber **interpretar** los resultados de estos y del como esta planteada la **prueba de hipótesis nula**.

Autocorrelación

Autocorrelación

⚠ Note un modelo de corte transversal:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

- Donde $i \to N$

Ahora uno de **series de tiempo**:

$$Y_i = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

- Donde $t \to T$

> Las **series de tiempo** a diferencia de los cortes transversales (diferentes por individuos) difieren o varian por periodos de tiempo.

Ejemplo

Nos enfocamos en una **unidad/individuo** *p.e:* Colombia

Lo observamos varias veces en el **tiempo** *p.e:* Colombia: 2015-2025

Autocorrelación: Series de Tiempo

Tiempo	Observaciones
2017	125
2018	451
2019	378
2020	391
2021	115

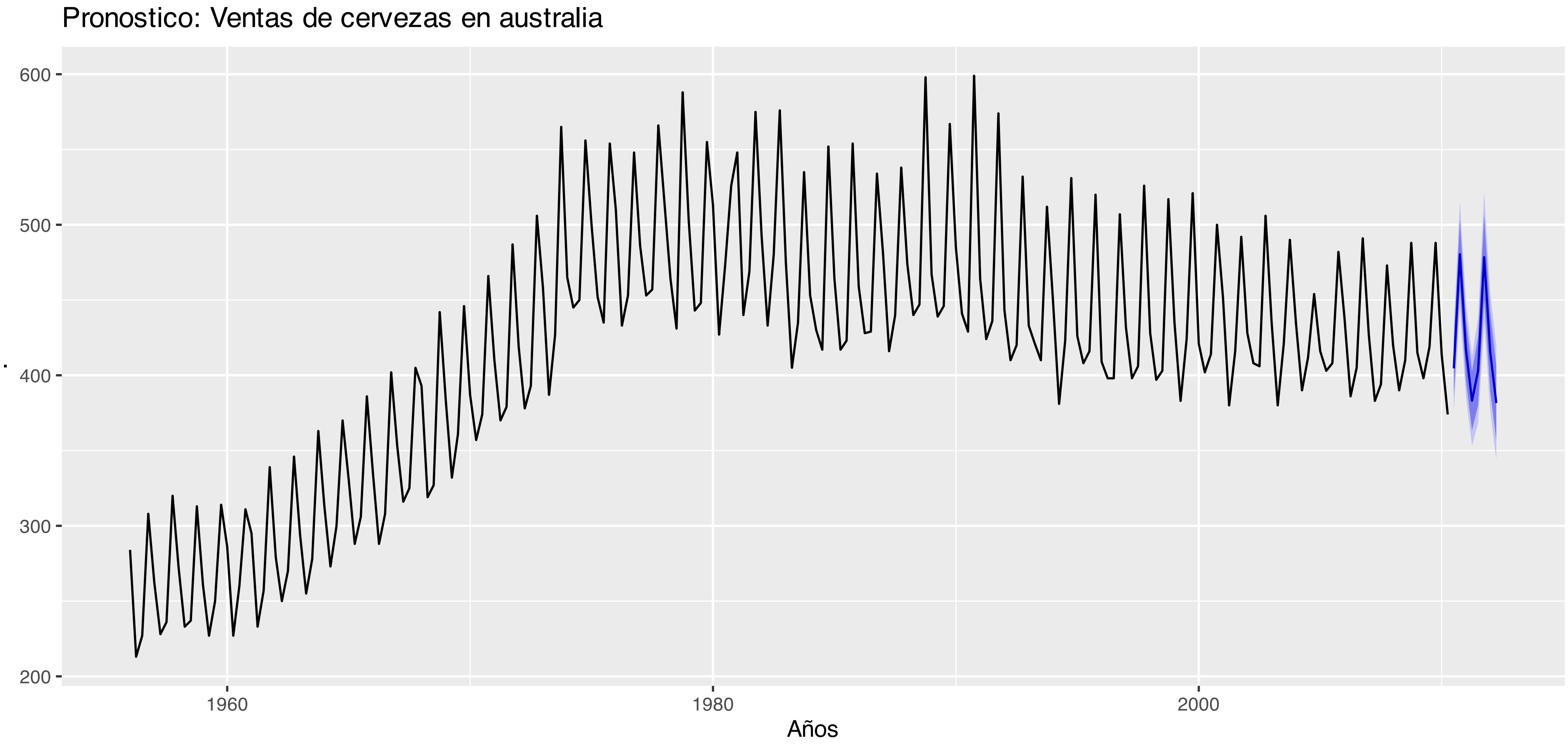
> En una **serie de tiempo** los reportes deben ir organizados por un **indice** que debe variar en el tiempo.

- Para series con frecuencia mas desagregadas *p.e:* (mes, trimestre, semestre) habría que usar en **R** la opción de `frequency`:

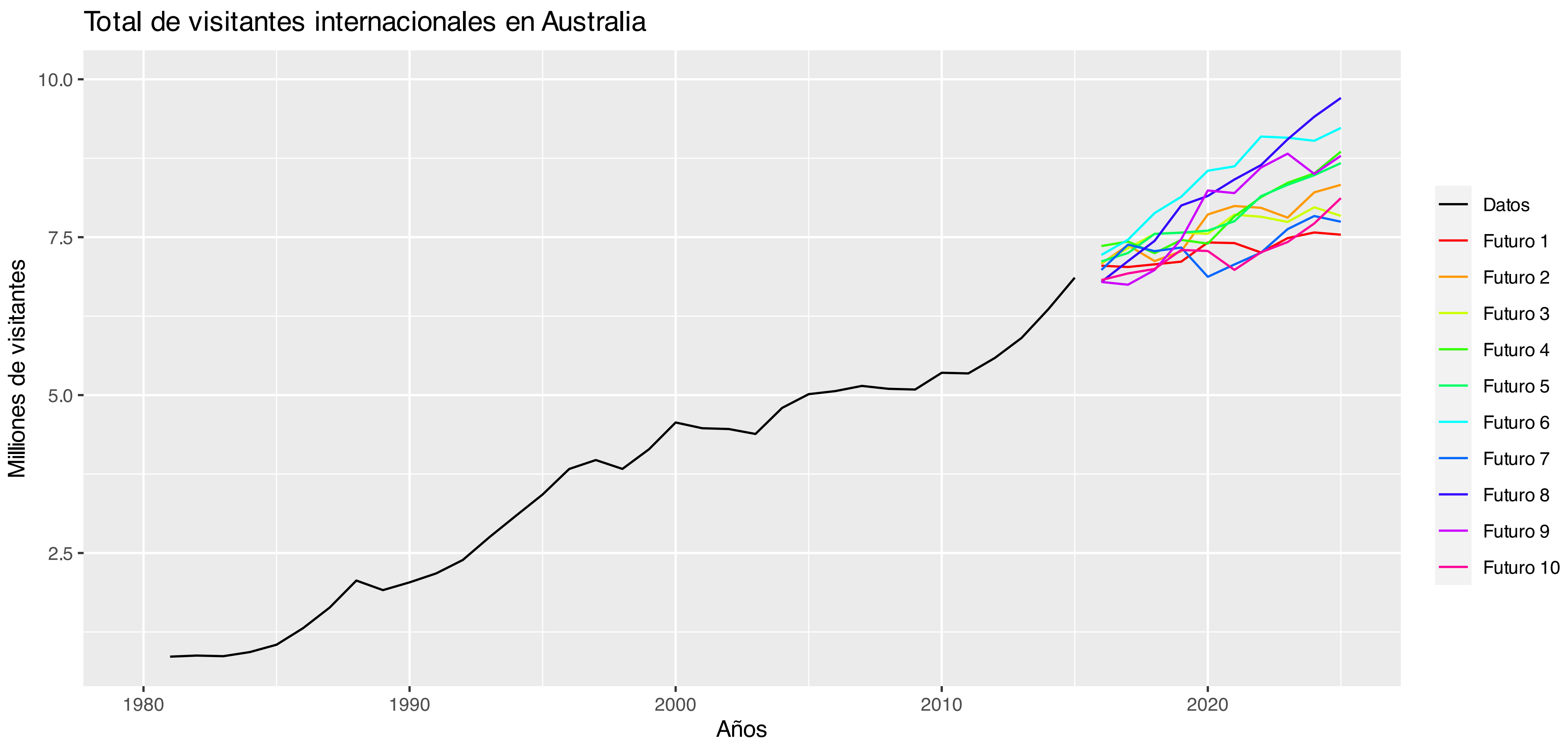
```
y <- ts(datos, frequency=12, start=c(2010, 1))
```

- *Donde 12 es la parte de frecuencia mensual y en la parte de `start` el año y mes donde inicia la serie.*

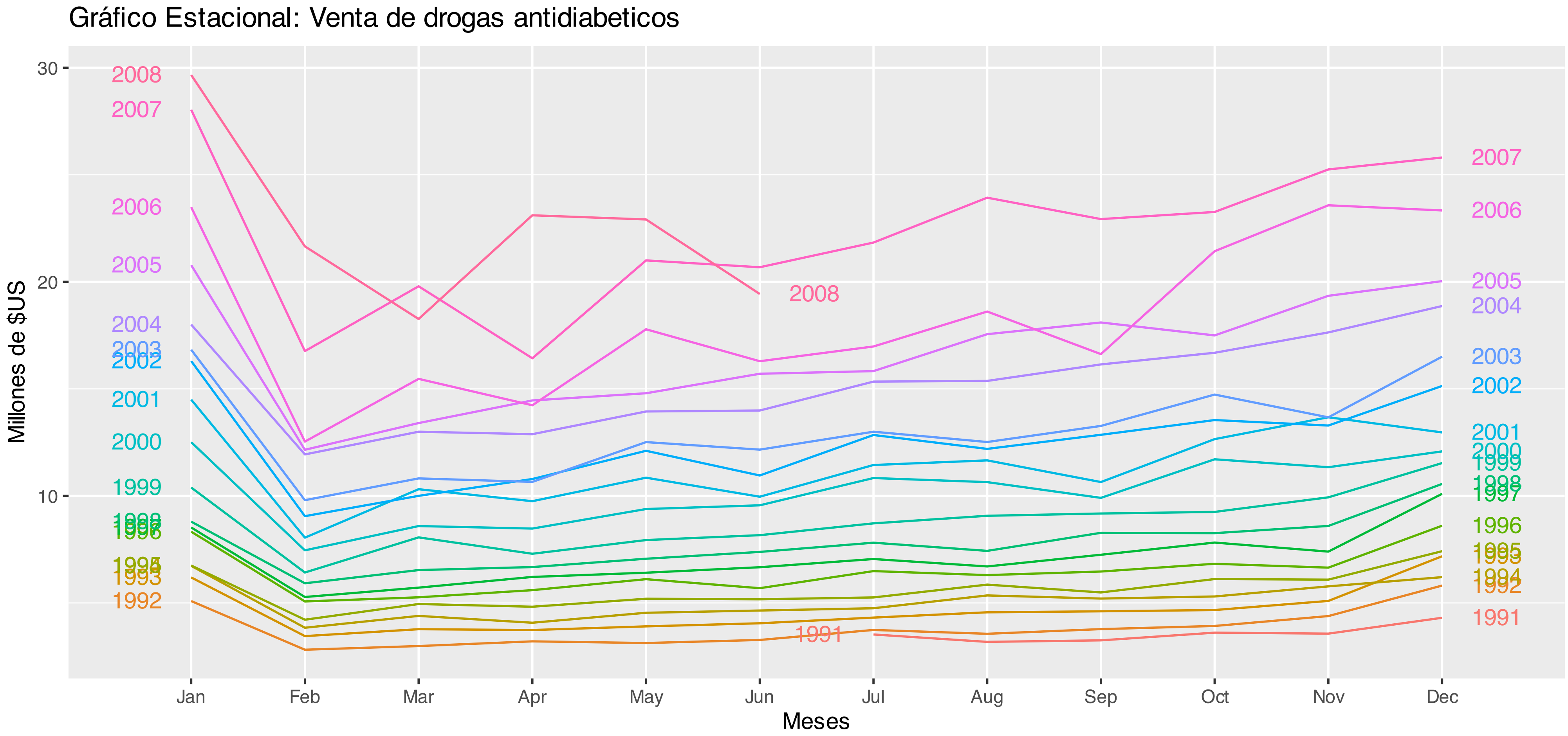
Autocorrelación: Series de Tiempo



Autocorrelación: Series de Tiempo



Autocorrelación: Series de Tiempo



Modelos MCO

Autocorrelación: Tipo de Modelos MCO

Nuestro modelo inicial puede ser:

$$\text{Consumo}_t = \beta_0 + \beta_1 \text{Ingreso}_t + u_t$$

O tal vez

$$\text{Consumo}_t = \beta_0 + \beta_1 \text{Ingreso}_t + \beta_2 \text{Ingreso}_{t-1} + u_t$$

Incluso puede ser

$$\text{Consumo}_t = \beta_0 + \beta_1 \text{Ingreso}_t + \beta_2 \text{Ingreso}_{t-1} + \beta_3 \text{Consumo}_{t-1} + u_t$$

- Las **series** pueden ser **estáticos**, un momento (t) en particular
- Y de otra forma **dinámicos**, que ya incluyen choques que afectan el futuro de la **serie**.

Autocorrelación: Tipo de Modelos MCO

Concepto

Autocorrelación: ocurre cuando nuestras *perturbaciones* están correlacionadas en el tiempo, *p.e*:
$\text{Cov}\left(u_t,\, u_j\right) \neq 0$ para $t \neq j$

Otra forma de pensar: Si el choque de la perturbación (t) se correlaciona con las perturbaciones "cercanas" en $(t-1)$ y $(t+1)$.

El término de **correlación serial** es el mismo que **Autocorrelación**

Operador rezago

Hay que estar atentos a lo que significa un **rezago**, esto es:

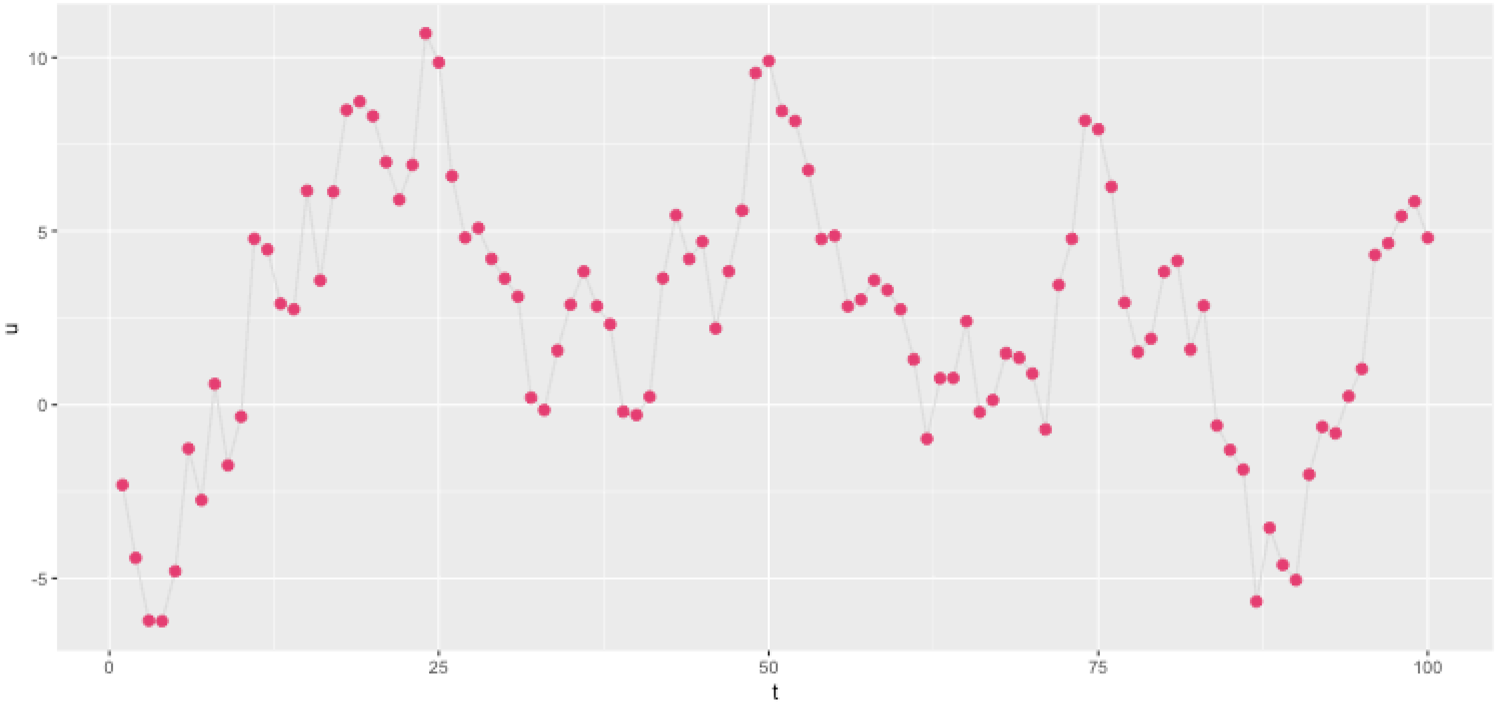
Tiempo	Observaciones	Rezago
2017	125	NA
2018	451	125
2019	378	451
2020	391	378
2021	115	391

Los **rezagos** o "lags" permiten medir los efectos temporales de una variable. Esto puede estar bien para (Y) pero mal para los residuos (u).

P.e.:,
{1, 2, 3, 4, 5, 6, 7, 8, 9} = x
{**?**, 1, 2, 3, 4, 5, 6, 7, 8} = lag(x)
{**?**, **?**, 1, 2, 3, 4, 5, 6, 7} = lag(x, 2)
{**?**, **?**, **?**, 1, 2, 3, 4, 5, 6} = lag(x, 3)

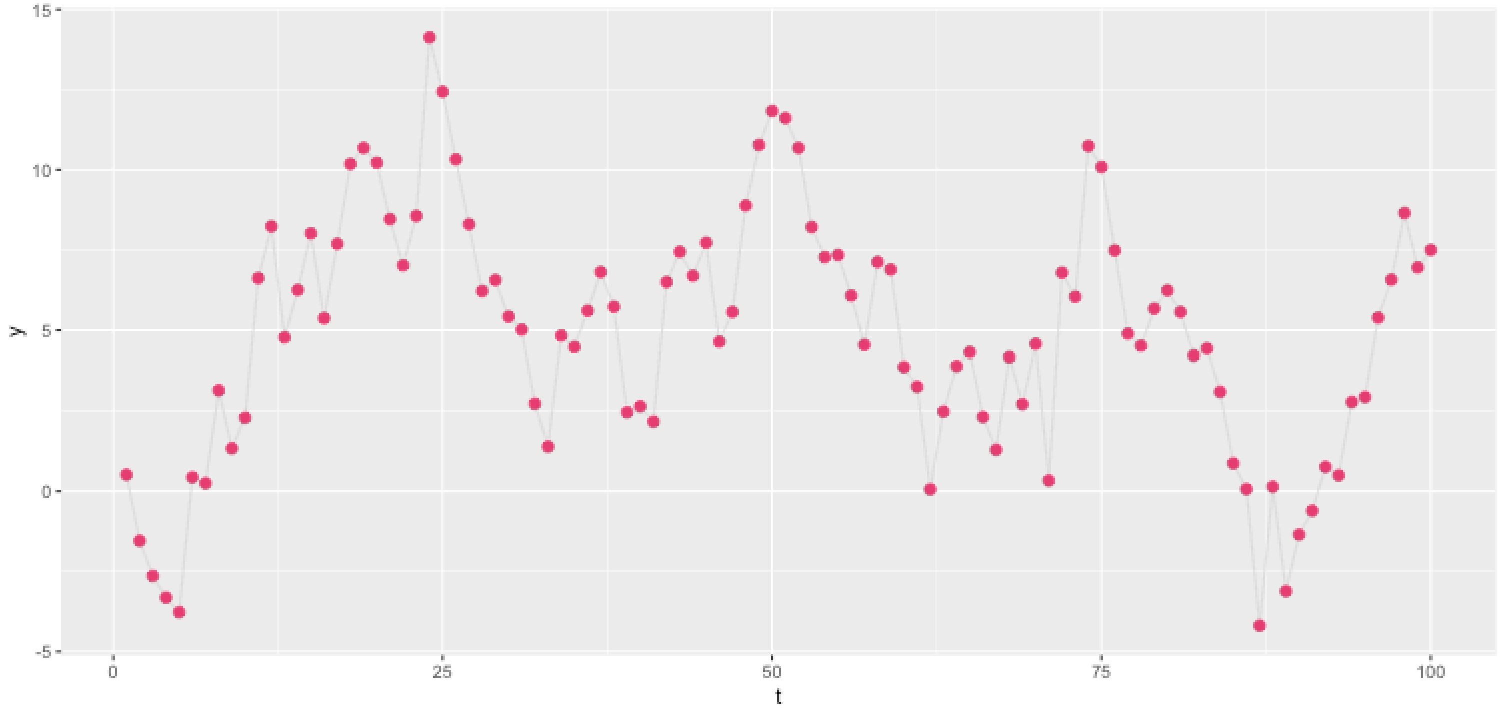
Autocorrelación

Autocorrelación Positiva: Residuos (u_t) en el tiempo



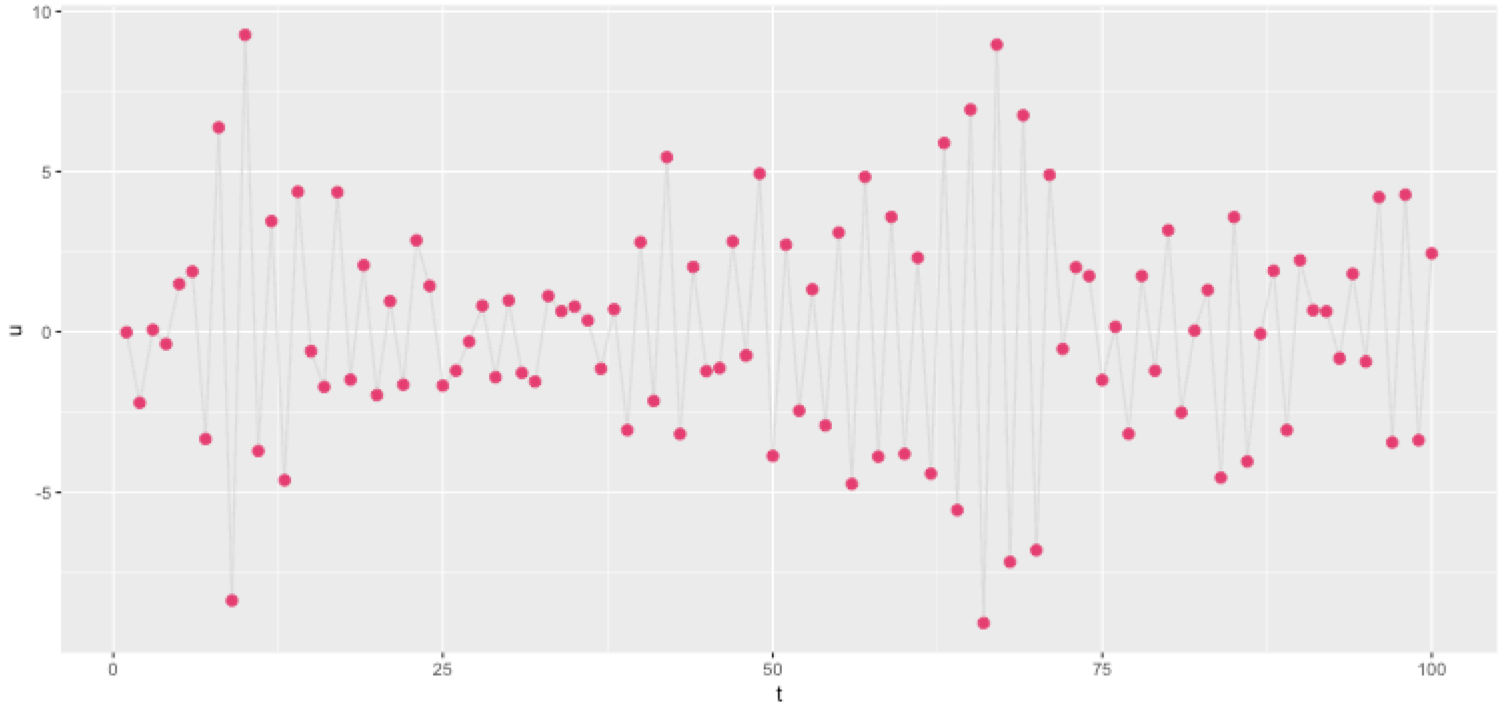
Autocorrelación

Autocorrelación Positiva: Dependiente (y_t) en el tiempo



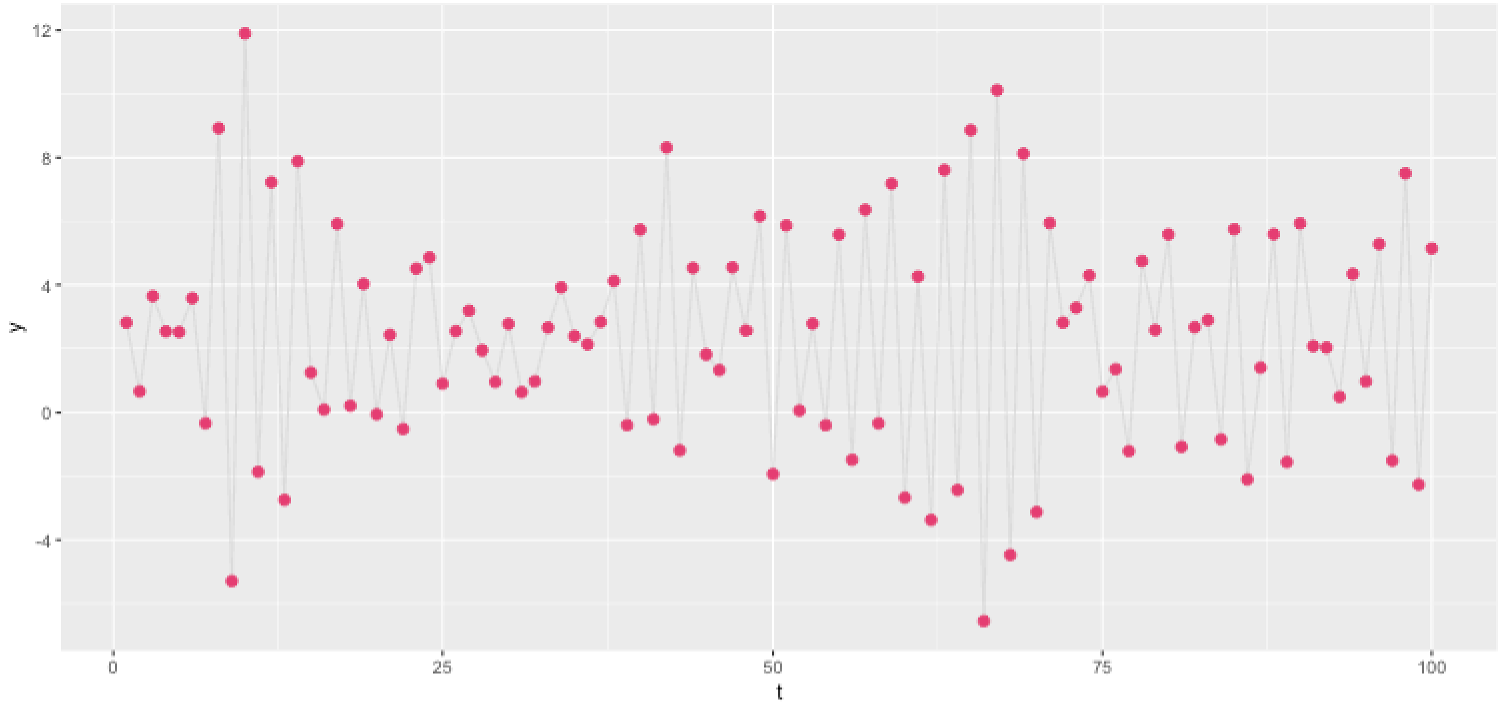
Autocorrelación

Autocorrelación Negativa: Residuos (u_t) en el tiempo



Autocorrelación

Autocorrelación Negativa: Dependiente (y_t) en el tiempo



Autocorrelación: Modelos Estáticos

Empecemos con un modelo muy común: un modelo de serie temporal estática cuyas perturbaciones presentan **autocorrelación de primer orden**, *conocido como* **AR(1)**:

$$\text{Consumo}_t = \beta_0 + \beta_1 \text{Ingreso}_t + u_t$$

Donde

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

Y el residuo ε_t es independiente e idénticamente distribuido (*i.i.d.*).

Para una **Autocorrelación de segundo-orden**, o **AR(2)**, puede ser

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

Autocorrelación: Modelos Estáticos

Un modelo/proceso **AR(p)** tiene una estructura de perturbaciones:

$$u_t = \sum_{j=1}^{p} \rho_j u_{t-j} + \varepsilon_t$$

Mostrando que las perturbaciones del presente (t) estan correlacionadas con sus rezagos p.

Recuerde que ρ es similar a un parámetro β solo que este se usa como si estuviera modelando a las perturbaciones (e_t)

Autocorrelación: Problemas

De forma similar a la **heterocedasticidad** tener **correlación serial** nos genera:

- MCO con estimadores **insesgados**.
- MCO los errores estandar $(S.E)$ de los $(\beta' s)$ son **sesgados**
- MCO es **ineficiente**

> La autocorrelación se vuelve más complicada con las variables dependiente rezagadas.

Autocorrelación: Modelo Dinámico

Considere un modelo de rezago distribuido de tal manera que:

$$\text{Consumo}_t = \beta_0 + \beta_1 \text{Ingreso}_t + \beta_2 \text{Consumo}_{t-1} + u_t$$

Donde

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

Problema:

Ambos Consumo_{t-1} (como regresor en el periodo t) y u_t (el residuo en t) dependen de u_{t-1}. *p.e.*, un regresor esta correlacionado con el residuo contemporaneo

P: Por qué este problema?
R./: Esto viola **Exogeneidad contemporanea**, *p.e.*, $\text{Cov}(x_t,\, u_t) \neq 0$.

Autocorrelación: Modelo Dinámico

Para ver esto hay que escribir el modelo en t y en $t-1$:

$$\begin{aligned} \text{Consumo}_t &= \beta_0 + \beta_1 \text{Ingreso}_t + \beta_2 \text{Consumo}_{t-1} + u_t \\ \text{Consumo}_{t-1} &= \beta_0 + \beta_1 \text{Ingreso}_{t-1} + \beta_2 \text{Consumo}_{t-2} + u_{t-1} \end{aligned}$$

Observe ahora que $u_t = \rho u_{t-1} + \varepsilon_t$.

$$\text{Consumo}_t = \beta_0 + \beta_1 \text{Ingreso}_t + \beta_2 \text{Consumo}_{t-1} + \overbrace{(\rho u_{t-1} + \varepsilon_t)}^{u_t} \tag{1}$$

$$\text{Consumo}_{t-1} = \beta_0 + \beta_1 \text{Ingreso}_{t-1} + \beta_2 \text{Consumo}_{t-2} + u_{t-1} \tag{2}$$

En (1), podemos ver que u_t depende de (esta correlacionado) u_{t-1}.
En (2), miramos entonces que Consumo_{t-1}, un regresor en (1), también con covarianza con u_{t-1}.

$\therefore$ En síntesis, el modelo esta sesgado y no es **exogeno**

Autocorrelación: Test de corrección

- Tenemos a **Breush - Pagan**
- Uno ampliamente conocido como **Durbin-Watson**

Autocorrelación: Breush-Pagan

Paso 1: Estima su modelo estático $(y_t = \beta_0 + \beta_1 x_t + u_t)$ con MCO

```
modelo <- lm(y ~ x, data = base)
```

Paso 2: Agregamos residuos a la base

```
base$e <- residuals(modelo)
```

Paso 3: Estima la regresión del residuo con su rezago (**sin intercepto**)

```
mod_resid <- lm(e ~ -1 + lag(e), data = base)
```

Autocorrelación: Breush-Pagan

Paso 4: Observe el t estadístico del parámetro $(\hat{\rho})$ que es el coeficiente en el paso 3.

```
tidy(mod_resid)
```

term	estimate	std.error	statistic	p.value
lag(e)	0.973	0.0508	19.1	1.28e-17

Si la probabilidad del t valor es menor al p-value de 0.05, tendríamos que rechazar a H_0.

La hipótesis H_0 es $H_0 : \rho = 0$, *p.e.*, es decir, no hay autocorrelación.

Paso 5: Concluimos para este caso que **existe** suficiente evidencia estadística para rechazar H_0 y por ende tenemos autocorrelación.

Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado

Paso 1: Estimamos el modelo en rezago (1, 0) con MCO.

```
# Estimar el modelo
mco_est <- lm(
  y ~ lag(y) + x,
  data = base
)
# Resumen
tidy(mco_est)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-3.85e-10	2.78e-10	-1.38	0.179
lag(y)	1.02	1.21e-15	8.42e+14	0
x	-1.84e-16	1.97e-16	-0.934	0.359

Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado

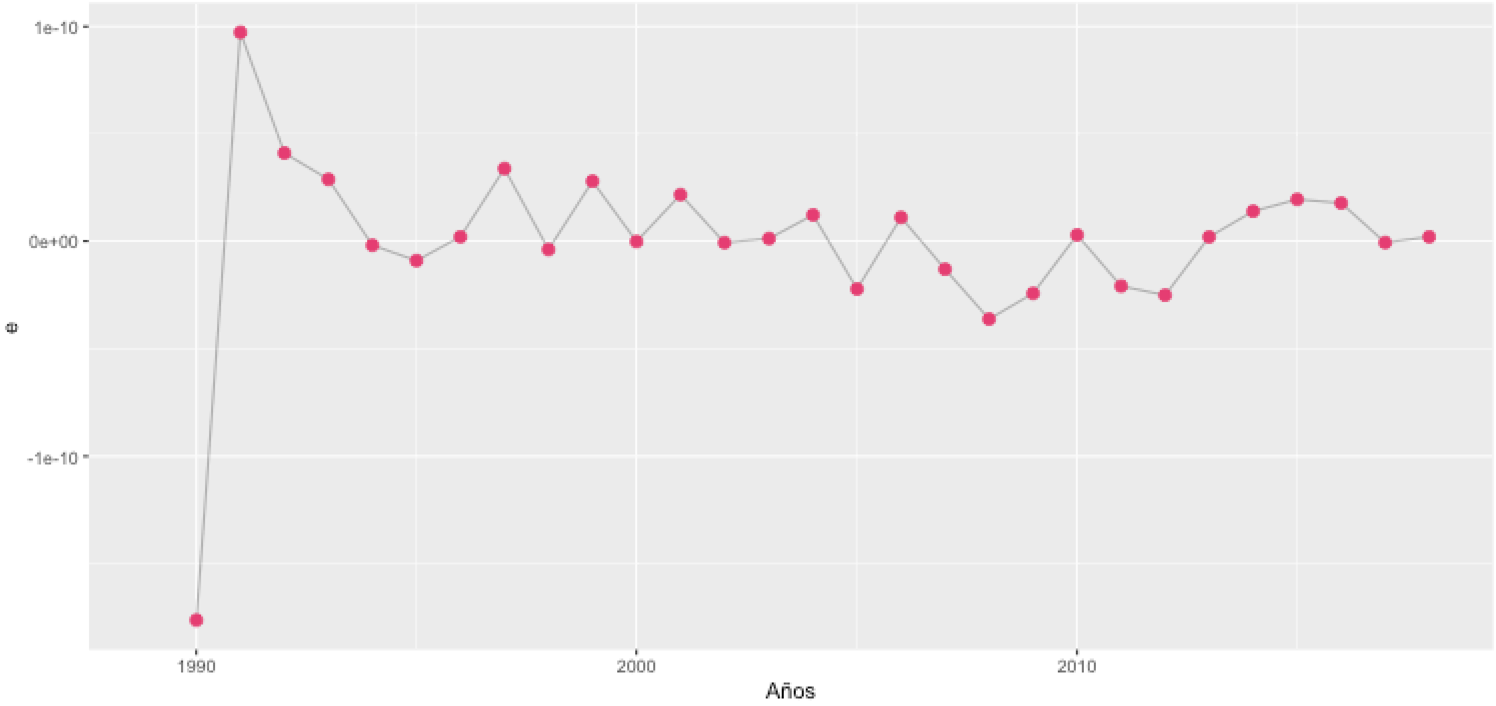
Paso 2: Guardamos los residuos de ese modelo

```
# Guardar residuos
base$ebg <- c(NA, residuals(mco_est))
```

Nota: Colocamos la parte de `NA` porque el primer rezago desaparece (*valor perdido*).

Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado



Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado

Paso 3: Regresión de los residuos con un intercepto, variables explicativas, y residuos rezagados.

```
# BG reg
bg_mod <- lm(
  ebg ~ lag(y) + x + lag(ebg) + lag(ebg, 2),
  data = base
)
```

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	0.000	0.000	1.768	0.0909	.
lag(y)	-0.000	0.000	-1.743	0.0953	.
x	0.000	0.000	1.702	0.1029	
lag(ebg)	0.094	0.197	0.479	0.6366	
lag(ebg, 2)	-0.039	0.087	-0.441	0.6633	

*Signif. codes: 0 <= '**' < 0.001 < '' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1*

Residual standard error: 1.731e-11 on 22 degrees of freedom

Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado

Paso 4: Prueba F (o LM) Para probar si $\rho_1 = \rho_2 = 0$.

Recuerde: Debemos usar la prueba F, nos permite moldear de esta forma las hipótesis (Mire: $\rho_1 = \rho_2 = 0$)

$$F_{q,\, n-p} = \frac{\left(\text{SRC}_r - \text{SRC}_{nr}\right) / q}{\text{SRC}_{nr} / \left(n - p\right)}$$

Donde q es el número de restricciones y p es el número de parámetros del modelo *no restringido* (incluyendo el intercepto).

Podemos usar la función de `waldtest()` del paquete `lmtest` de R para esto.

Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado

Paso 4: Prueba F (o LM) Para probar si $\rho_1 = \rho_2 = 0$.

```
# BG regresion
bg_mod <- lm(
  ebg ~ lag(y) + x + lag(ebg) + lag(ebg, 2),
  data = base
)
# Usamos el test
p_load(lmtest)
waldtest(bg_mod, c("lag(ebg)", "lag(ebg, 2)"))
```

Aquí, estamos implementando la opción de `waldtest` para probar

- La especificación del modelo `bg_mod` (nuestro **modelo no restringido**)
- Contra un modelo sin los términos `lag(ebg)` y `lag(ebg, 2)` (nuestro **modelo restringido**)

Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado

Paso 4: Prueba F (o LM) Para probar si $\rho_1 = \rho_2 = 0$.

```r
# BG regresion
bg_mod <- lm(
  ebg ~ lag(y) + x + lag(ebg) + lag(ebg, 2),
  data = base
)
# Usamos el test
p_load(lmtest)
waldtest(bg_mod, c("lag(ebg)", "lag(ebg, 2)"))
```

Res.Df	Df	F	Pr(>F)
22			
24	-2	0.399	0.676

Autocorrelación: Breush-Pagan

Ejemplo: Modelo rezagado

Paso 5: Conclusiones

Con un p-value de ~0.676, **No rechazamos la hipotesis nula.**

- No podemos rechazar que $\rho_1 = \rho_2 = 0$.
- No podemos rechazar la "no autocorrelación" en el modelo.

Sin embargo, **Vamos a testear también la autocorrelación de orden**: AR(2).

Podemos obtener diferentes respuestas con diferentes pruebas.

El p-value para el AR(1) es 0.0178—que sugiere una autocorrelación de primer orden *este modelo es para una econometría mas avanzada*.

Autocorrelación: Durbin-Watson

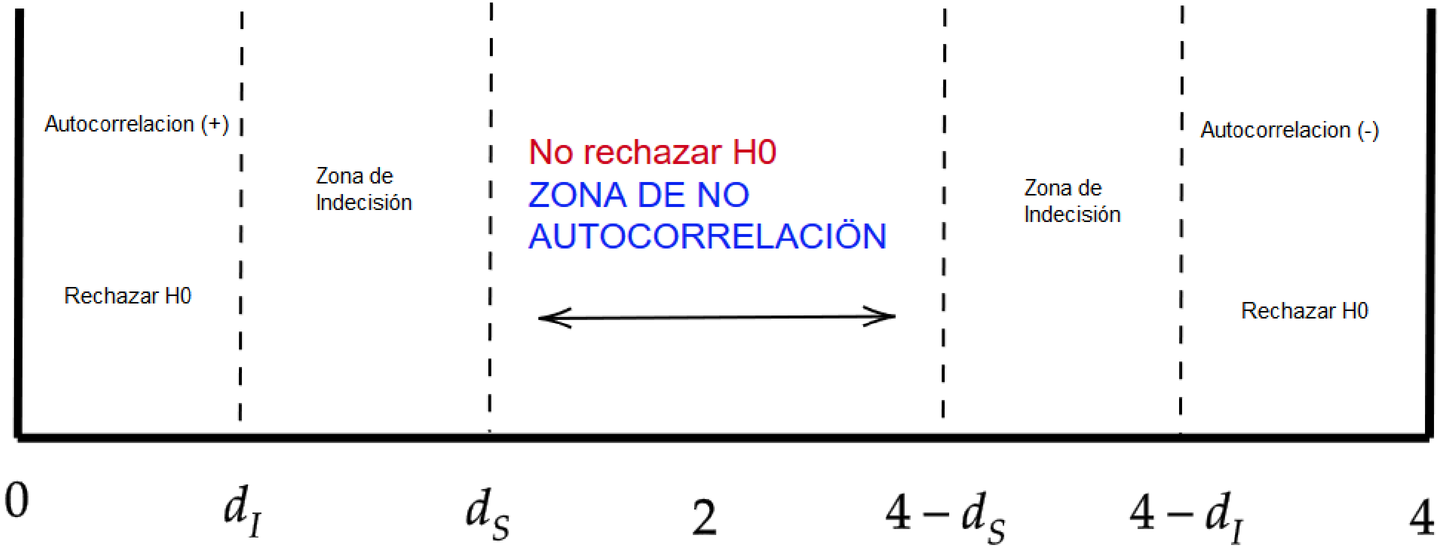
Es la mas conocida para encontrar autocorrelación.

- Asume que $\mu_t = \rho\mu_{t-1} + \epsilon_t$ y sigue una distribución asintótica.
- La formula del estadístico es:

$$d = \frac{\sum_{t=2}^{T} (\mu_t - \mu_{t-1})^2}{\sum_{t=1}^{T} \mu_t^2}$$

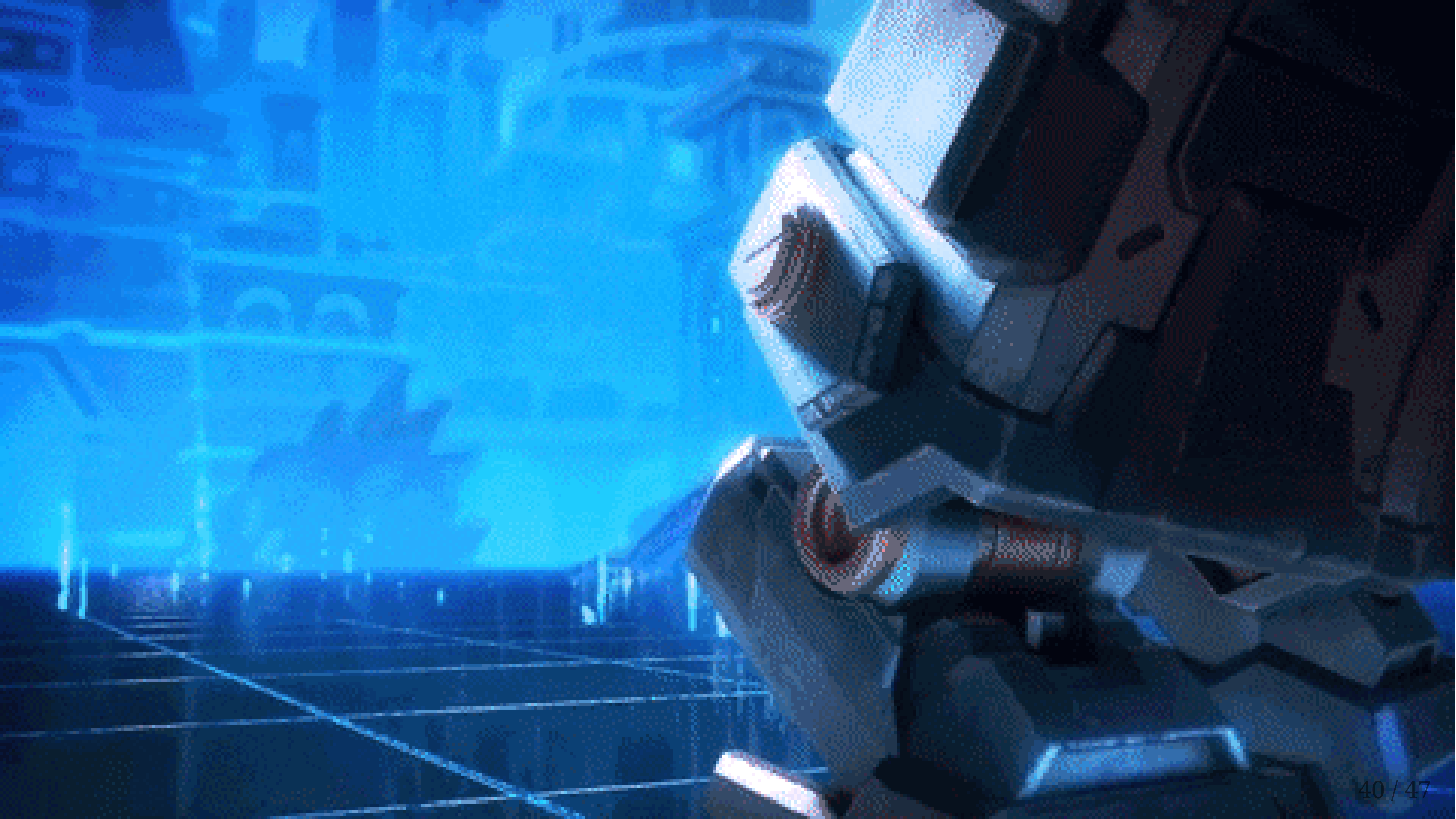
- El estadístico DW, se puede aproximar como: $d \approx 2(1 - \rho)$
- Como el coeficiente de correlación oscila entre -1 y 1, el estadístico DW se situara entre 0 y 4, es decir: $0 \leq d \leq 4$
- En el mejor de los casos, $d = 2$, nos indicaría que **NO** habría problema de autocorrelación.

Autocorrelación: Durbin-Watson



$$H_0 = \textit{No hay autocorrelación } (+)$$
$$H_0^* = \textit{No hay autocorrelación } (-)$$

Como corregir ? ...



Autocorrelación: Corrección Método Cochrane-Orcutt

Realizar una **transformación** del modelo original con un **modelo de diferencias** teniendo en cuenta el parámetro de la correlación del residuo.

- Hallar el parámetro ρ.

$$\epsilon_t = \rho\epsilon_{t-1} + v_t$$

- Se aplica la transformación de **primeras diferencias**.

$$Y_t - \rho Y_{t-1} = \beta_0(1 - \rho) + \beta_1(X_t - \rho X_{t-1}) + \mu_t$$

- De esta forma ya se ha corregido el problema.

Autocorrelación: Corrección Método Cochrane-Orcutt

Vamos a tener un modelo de:

$$\text{Ventas}_t = \beta_0 + \beta_1 \text{inventarios}_t + \epsilon_t$$

Paso 1: Estimar el modelo

```
library(dynlm) # Modelo dinámico
library(orcutt) # Corrección Autocorrelación
base_da<-read_excel("regauto.xlsx")
reg.dyn<-dynlm(ventas~inventarios, data = base_da)
reg.dyn %>% broom::tidy()
```

term	estimate	std.error	statistic	p.value
(Intercept)	-903	1.17e+03	-0.775	0.443
inventarios	0.643	0.00289	223	1.72e-63

Autocorrelación: Corrección Método Cochrane-Orcutt

Paso 2: Hallar estadistico D-Watson

```
dwtest(reg.dyn)
```

```
#> 
#>     Durbin-Watson test
#> 
#> data:  reg.dyn
#> DW = 1.3743, p-value = 0.01175
#> alternative hypothesis: true autocorrelation is greater than 0
```

Note que el estadístico DW, cae en la zona de **autocorrelación positiva**. Debemos proceder a su corrección

Autocorrelación: Corrección Método Cochrane-Orcutt

Paso 3: Corregir el problema

```
cochrane.orcutt(reg.dyn)
```

```
#> Cochrane-orcutt estimation for first order autocorrelation
#>
#> Call:
#> dynlm(formula = ventas ~ inventarios, data = base_da)
#>
#>  number of interaction: 3
#>  rho 0.311047
#>
#> Durbin-Watson statistic
#> (original):    1.37429 , p-value: 1.175e-02
#> (transformed): 2.04410 , p-value: 4.902e-01
#>
#>  coefficients:
#> (Intercept) inventarios
#> -966.166560    0.642991
```

> De tal manera que el modelo ya esta corregido

Autocorrelación: Corrección Método Cochrane-Orcutt

	(1)	(2)
(Intercept)	-902.827	-966.167
	(1165.121)	(1673.581)
inventarios	0.643 ***	0.643 ***
	(0.003)	(0.004)
N	42	42
R2	0.999	0.998
logLik	-414.513	
AIC	835.026	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Bibliografía

- Gujarati, D. N., & Porter, D. C. (2011). *Econometria Básica*. Ed. Porto Alegre: AMGH..
- Stock, J. H., Watson, M. W., & Larrión, R. S. (2012). *Introducción a la Econometría*.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Gracias por su atención!

Carlos Andres Yanes Guerra

cayanes@uninorte.edu.co

keynes37