

ANÁLISIS DE SERIES DE TIEMPO CON R

C. Vladimir Rodríguez Caballero

<https://github.com/Vlasmetrics7/BID-HACIENDA-2023>

- Profesor e investigador del ITAM
- Investigador internacional del CREATES, Aarhus University
- Áreas de investigación:
 - Modelación Econométrica
 - Pronóstico

Encuéntrame en:

vladimir.rodriguez@itam.mx

<http://c-vladimir-rodriguez-caballero.weebly.com/>

Parte I. Pasos iniciales

Recordando comandos y estructura de programación básica.

Modelos econométricos.

Análisis de diagnóstico.

Pronóstico y selección de modelos.

Parte II. Análisis de series de tiempo

Modelos básico de series de tiempo.

Estimación.

Metodología Box & Jenkins.

Estimación y pronóstico.

Parte III. Modelación con series económicas

Análisis de series de tiempo no estacionarias.

Ejecución de pruebas de raíces unitarias.

Rompimientos estructurales.

Análisis de cointegración.

Supuestos del curso

- Aunque es un curso introductorio de R, asumiré que ya tenemos cierta familiaridad con el lenguaje y con el manejo del ambiente de Rstudio.
- De igual forma, aunque no es un curso de Estadística o Econometría, asumiré que ya tenemos cierta familiaridad con los conceptos básicos como media, desviación estándar, cuantiles, regresión, etc...
- No es curso teórico por lo que no vamos a derivar los resultados sino a utilizar nuestro conocimiento de estadística y econometría y a ponerlo en práctica con R.
- A lo largo de estos seis días estaremos repasando conocimientos de econometría y series de tiempo y cómo llevar a cabo una etapa de modelaje en R.

Algunas referencias:

- ❑ Dalgaard, P. (2020). Introductory statistics with R. Springer
- ❑ Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2019). Introduction to Econometrics with R. University of Duisburg-Essen, 1-9.
- ❑ Hyndman, R. J. & Athanasopoulos, G. (2020) Forecasting: principles and practice, 3rd ed
- ❑ Kleiber, C., & Zeileis, A. (2008). Applied econometrics with R. Springer Science & Business Media.
 - Libres y pueden ser consultados online
 - Ciertos paquetes de R incluyen bases de datos
 - Códigos de R asociados

Algunos paquetes que usaremos:

```
install.packages(" ")
```

- ❑ AER: paquete del libro: Applied Econometrics with R
- ❑ tidyverse: Es una colección de paquetes de R diseñados para ciencia de datos.
<https://www.tidyverse.org/packages/>
- ❑ Ggally y ggplot2: Son un par de paquetes diseñados para crear gráficos técnicos.
- ❑ sugrants: Es un paquete que nos ayuda a crear gráficos espectaculares de series temporales.

Algunos paquetes que usaremos:

- ❑ tsibble: Es una colección de funciones para la realización de pronósticos.
- ❑ fable: Otro paquete diseñado para crear gráficas estadísticas
- ❑ sugrrants: Es un paquete que nos ayuda a crear gráficos espectaculares de series temporales.
- ❑ fpp3: Forecasting: Principles and Practice.

Es un paquete diseñado para análisis de series de tiempo.

<https://otexts.com/fpp3/>

- ❑ + Muchos otros paquetes especiales para estimar distintos modelos econométricos.

Estructura básica de RStudio



Retomando ideas:

RStudio interface showing code execution and environment details.

Environment Panel (Top Right):

Object	Details
CentroTemp	Large mts (74880 elements, 603.5 kb)
dataMAX	1440 obs. of 52 variables
dataMIN	1440 obs. of 52 variables
RangeTemp	Large mts (74880 elements, 603.5 kb)

Code Editor (Left):

```
13 library(ggplot2)
14 library(reshape2)
15 library(grid)
16 library(gridtext)
17 library("KFAS")
18 #####
19 setwd("C:/Users/crodriugab/Dropbox/0. CVRC [work]/B. Working papers and projects/3. Esther (climate)/Climate Data")
20 setwd("C:/Users/vlasm/Dropbox/0. CVRC [work]/B. Working papers and projects/3. Esther (climate)/Climate Data")
21
22 dataMAX = read_excel("Tempmaxspain.xlsx", sheet = "Hojal")
23 dataMIN = read_excel("Tempminspain.xlsx", sheet = "Hojal")
24 dataMIN = read_excel("Tempminspain.xlsx", sheet = "Hojal")
25
26 CentroTemp = (dataMAX+dataMIN)/2;
27 RangeTemp = log(dataMAX-dataMIN);
28
29 CentroTemp <- ts(CentroTemp,frequency=12,start=c(1901,1))
30 RangeTemp <- ts(RangeTemp,frequency=12,start=c(1901,1))
31
32 #####
33 ##### CENTER TEMP #####
34 ##### MULTIVARIADO #####
35 #####
36 #####
37
38
```

Console (Bottom Left):

```
combine
> library(ggplot2)
> library(reshape2)
Attaching package: 'reshape2'
The following object is masked from 'package:tidyr':
  smiths
> library(grid)
> library(gridtext)
> library("KFAS")
Please cite KFAS in publications by using:
  Jouni Helske (2017). KFAS: Exponential Family State Space Models in R. Journal of Statistical Software, 78(10), 1-39. doi:10.18637/jss.v078.i10.
> dataMAX = read_excel("Tempmaxspain.xlsx", sheet = "Hojal")
> dataMIN = read_excel("Tempminspain.xlsx", sheet = "Hojal")
> dataMIN = read_excel("Tempminspain.xlsx", sheet = "Hojal")
>
> CentroTemp = (dataMAX+dataMIN)/2;
> RangeTemp = log(dataMAX-dataMIN);
>
> CentroTemp <- ts(CentroTemp,frequency=12,start=c(1901,1))
> RangeTemp <- ts(RangeTemp,frequency=12,start=c(1901,1))
> plot.ts(CentroTemp[,1])
>
```

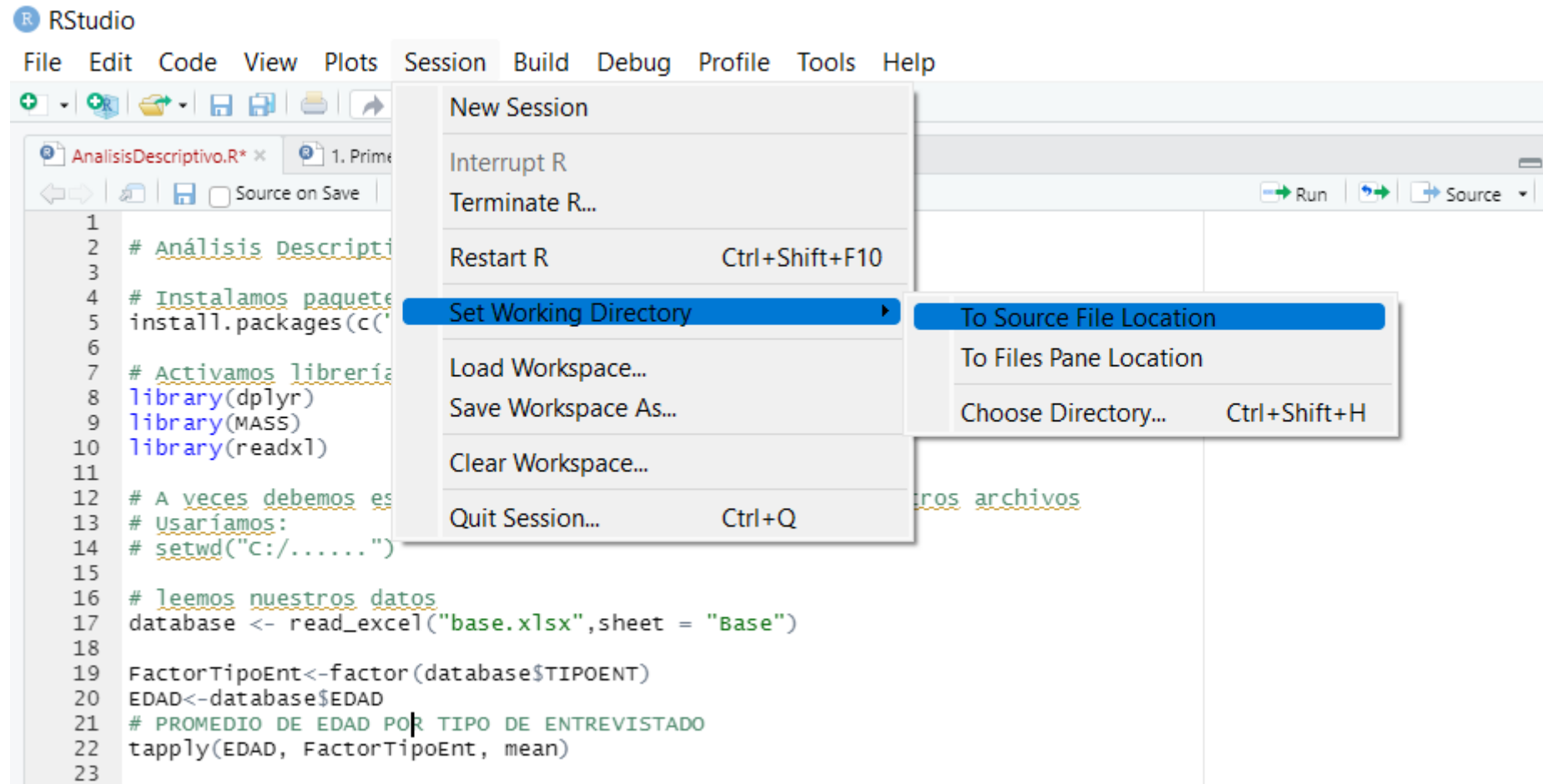
Plot (Bottom Right):

CentroTemp[,1]

Time

Retomando ideas:

A veces las direcciones de nuestras carpetas son complejas:
Este "truco" nos facilitará la vida



Una revisión de estadística usando R:

Recordemos que:

1. Debemos instalar los paquetes si es la primera vez que lo vamos a usar:
 - `install.packages(c("readxl","dplyr","MASS"))`
2. Debemos estar seguros de cargar las librerías respectivas:
 - `library(dplyr)`
 - `library(MASS)`
 - `library(readxl)`

Algunas bases de datos de interés

Algunas bases de datos internacionales que suelen usarse

Federal Reserve Bank of St. Louis's website (FRED)

<https://fred.stlouisfed.org/>

World Bank's website

<https://data.worldbank.org/>

Por ejemplo en la base de datos de la FRED:

- Inflation Rate: FRED Code: Consumer Price Index: Total All Items for the United States (CPALTT01USQ657N)

Algunas bases de datos internacionales que suelen usarse

Códigos a usar:

1. Base de datos.R

```
#install.packages('quantmod')
library(quantmod)

## Loading required package: xts
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
## Loading required package: TTR
## Version 0.4-0 included new data defaults. See ?getSymbols.

##?getSymbols
getSymbols('CPALTT01USQ657N',src = 'FRED')

## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.

## Warning in strptime(xx, f <- "%Y-%m-%d", tz = "GMT"): unknown timezone
## 'zone/tz/2018e.1.0/zoneinfo/America/Los_Angeles'
```

Algunas bases de datos internacionales que suelen usarse

Penn World Tables

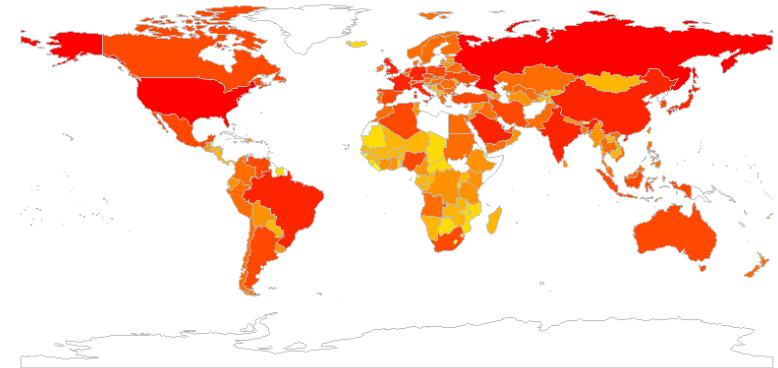
Los datos son bastante limpios y contienen medidas históricas de la actividad económica de 1950 a 2014 para 182 países

Se proporciona una lista completa de las variables de Penn World Table:

<https://cran.r-project.org/web/packages/pwt9/pwt9.pdf>

Códigos a usar:

2. Base de datos Penn.R



Modelos Econométricos

Análisis de Regresión

Metodología clásica:

- ① Planteamiento de la teoría o de la hipótesis.
- ② Especificación del modelo matemático de la teoría.
- ③ Especificación del modelo econométrico estadístico de la teoría.
- ④ Obtención de datos.
- ⑤ Estimación.
- ⑥ Pronóstico o predicción.
- ⑦ Control de política o evaluación.

Modelo de regresión múltiple

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_k x_{ki} + u_i.$$

- y_i es la variable explicada,
- x_{ki} , para $k = 1, 2, \dots, K$, son K variables explicativas,
- β_k , igual, para $k = 1, 2, \dots, K$ son los parámetros asociados a tales variables.
- Note que β_1 puede ser empleado como intercepto u ordenada en el origen; por lo mismo, si se desea tener un intercepto, x_{1i} quedaría conformado como un vector *de unos*, es decir: $x_{1i} = 1$ para todo $i = 1, \dots, n$ (a ese vector usualmente lo denotaremos ι), y, finalmente,
- u_i denota al término de error.

Algunos ejemplos populares

Tipo de modelo de regresión	Interpretación
$y = \beta_0 + \beta_1 x_1 + u$ <i>Lineal</i>	un cambio de una unidad en x_1 repercutirá en un cambio en el valor esperado de y por β_1
$y = \beta_0 x_1^{\beta_1} e^u$ <i>Exponencial o log-log</i>	β_1 mide la elasticidad de y con respecto a x (Cambio % en y antes un cambio % pequeño en x) Modelo linealizado: $\ln(y) = \beta_0 + \beta_1 \ln(x_1) + u$
$\ln(y) = \beta_0 + \beta_1 x_1 + u$ <i>semi-log</i>	un cambio de una unidad en x_1 repercutirá en un cambio en el valor esperado de y por $100 \times \beta_1$ %
$y = \beta_0 + \beta_1 \ln(x_1) + u$ <i>lin-log</i>	un cambio absoluto en y es debido a un cambio relativo en x por β_1
$y = \beta_0 + \beta_1 \left(\frac{1}{x_i} \right) + u_i$ <i>reciproco</i>	La relación entre y y x es inversa. Conforme aumenta x , y tenderá a disminuir.

En palabras llanas:

- 1 Forma funcional lineal de la relación,
- 2 Identificabilidad de los parámetros del modelo,
- 3 Valor esperado de la perturbación dada la información observada,
- 4 Varianzas y covarianzas de las perturbaciones dada la información observada,
- 5 Naturaleza de la muestra de los datos sobre las variables independientes, y
- 6 Distribución de probabilidad de la parte estocástica del modelo.

¿Cómo estimamos un modelo de regresión múltiple en R?

Códigos a usar:

3. Modelo Solow Swan.R

Recuerde que en el modelo de Solow-Swan, los aumentos en el crecimiento de la población se relacionaron negativamente con la producción por unidad de trabajo efectivo. Podemos comprobar esta predicción trazando el crecimiento medio de la población frente al crecimiento medio de la producción per cápita para ver si surge algún patrón.

Regresión Múltiple: Modelo lineal de Taylor

Taylor(1993) propone una regla simple para establecer una política monetaria donde los tomadores de decisiones ajustan la *Federal fund rate*, i_t , en función de la inflación, π_t , y la diferencia entre la producción y su nivel potencial de largo plazo, y_t .

El modelo es:

$$i_t = \beta_0 + \beta_1 \pi_t + \beta_2 y_t + u_t.$$

Taylor sugiere los valores de $\beta_1 = 1,5$ y $\beta_2 = 0,5$ para satisfacer su intuición monetaria. Ambas variables son exógenas de acuerdo a Taylor.

El modelo es estimado usando datos de los Estados Unidos desde marzo 1987-diciembre 1999, para un total de $T=52$ observaciones.

Códigos a usar:

4. ModeloLinealTaylor.R

- Usaremos este ejemplo para mostrar cómo podemos realizar pruebas de hipótesis básicas en nuestro modelo

Modelos Econométricos

Pruebas de diagnóstico

- El método de Mínimos Cuadrados Ordinarios permanece como el caballo de batalla en econometría y se emplea de manera rutinaria en el análisis de datos.
- El cumplimiento de los supuestos permitirá propiedades deseables: no sesgo, consistencia, y eficiencia.
 - ① Correcta especificación.
 - ② Rango completo.
 - ③ Ortogonalidad
 - ④ Homocedasticidad.
 - ⑤ No autocorrelación.

Importante

Es muy importante nunca fiarse de una regresión cuyos supuestos no hayan sido revisados escrupulosamente: hay que aplicarle todas las pruebas habidas y por haber.

Las tres reglas de oro en econometría

¡Probar, Probar y Probar!

Professor Sir David Hendry.

De hecho acorde a este mismo autor, la elección de un modelo econométrico para un análisis empírico debe satisfacer los siguientes criterios.

- Ser aceptable con respecto a los datos, es decir que las predicciones hechas con base en el modelo debe ser aceptablemente buenas.
- Ser coherente con la teoría: el valor y el signo de los parámetros así como las variables incluidas deben ser los correctos.
- Tener regresores débilmente exógenos: variables explicativas y términos de error deben ser ortogonales.
- Mostrar constancia paramétrica: estabilidad de los parámetros.
- Exhibir coherencia en los datos: residuales asimilables a ruido-blanco (en caso contrario, probablemente estaríamos enfrentando un error de especificación *disfrazado de* autocorrelación, heteroscedasticidad, etc)
- Ser inclusivo (*encompassing* en inglés): debe ser el mejor modelo posible, pero al mismo tiempo debe satisfacer el principio de **PARSIMONIA**.

Debemos tener muy claro que cualquier rompimiento de los supuestos constituye un **error de especificación**.

No obstante, no todos los errores de especificación tienen las mismas consecuencias.

A continuación una síntesis de los tipos de errores posibles que hemos cubierto hasta ahora:

Código a usar en toda esta sección:

5. Análisis MCO.R

Consumption = $b_0 + b_1 \text{ income} + b_2 \text{ production} + b_3 \text{ unemployment} + b_4 \text{ saving} + u$

Problemas potenciales: 1) Multicolinealidad

Multicolinealidad perfecta

- Lo que no queremos es que exista una relación lineal perfecta entre las variables explicativas.
- En este caso diríamos que dichas variables son colineales perfectas o bien que hay un problema de colinealidad perfecta.
- En muchas ocasiones dicha relación obedece a problemas de construcción de las variables, mismos que pueden pasar desapercibidos por los econometristas incautos.
- En realidad, éste es un problema fácil de identificar y por lo mismo, fácil de corregir (se puede eliminar la variable que esté causando la multicolinealidad).

Multicolinealidad imperfecta

Multicolinealidad

La multicolinealidad infla las varianzas de los parámetros estimados y reduce su significancia estadística de manera espuria.

Problemas potenciales: 1) Multicolinealidad

Resumen de consecuencias prácticas de la Multicolinealidad (imperfecta)

- 1 Aunque los estimadores de MCO son MELI, presentan varianzas y covarianzas grandes que dificultan la estimación precisa.
- 2 Los intervalos de confianza tienden a ser mucho más amplios, lo cual propicia una aceptación más fácil de $H_0 : \hat{\beta} = 0$.
- 3 La razón t de uno o más coeficientes tiende a ser estadísticamente no significativa.
- 4 Aunque la razón t de uno o más coeficientes sea estadísticamente no significativa, nuestra R^2 puede ser muy alta.
- 5 Los estimadores de MCO y sus errores estándar son sensibles a pequeños cambios en los datos.

Detección mediante el número de condición

La regla heurística para interpretar ambas medidas es sencilla:

- 1 si el NC está entre 100 y 1,000, considere que existe un problema de multicolinealidad que se caracteriza como moderado a fuerte;
- 2 si el NC es superior a 1,000, entonces se considera que el problema de multicolinealidad es severo.
- 3 De manera equivalente, si el IC está entre 10 y 30, hay multicolinealidad que va de moderada a fuerte y si es superior a 30, entonces hay multicolinealidad severa.

Problemas potenciales con el término de error: 2) Normalidad

El término de error no se distribuye en realidad $iid\mathcal{N}(0, \sigma^2)$, sino más bien $iid(0, \sigma^2)$.

No-Normalidad

Cuando la normalidad no viene *incorporada* desde un principio a través del término de error pero éste aún mantiene las propiedades de homoscedasticidad, esperanza nula e independencia, los estimadores de MCO siguen siendo MELI, sólo que ahora la inferencia estadística sólo es válida asintóticamente. Vale la pena mencionar que existen pruebas para estudiar la normalidad de los residuales estimados. Una de las más socorridas es la prueba de normalidad de **Jarque-Bera**, la prueba de **Kolmogorov-Smirnov** y el **Gráfico Cuantil-Cuantil**.

Problemas potenciales con el término de error: 3) Homocedasticidad y No autocorrelación

El término de error (i) no es homoscedástico, o; (ii) está serialmente correlacionado (no es independiente).

Heterocedasticidad y/o autocorrelación

Al romperse los supuestos relativos a la forma de la matriz de covarianzas del término de error, los estimadores de MCO permanecen insesgados y consistentes, sólo que ya no son los más eficientes; ya no son MELI. Vimos que:

- existen muchas pruebas para estudiar si estos supuestos se cumplen.

Heterocedasticidad:

H0: Homocedasticidad vs Ha: Heterocedasticidad

la Prueba de Goldfeld-Quandt (en desuso),
la Prueba Breusch-Pagan y
la Prueba de White.

Autocorrelación:

H0: No autocorr vs Ha: autocorr serial

la Durbin Watson,
la Breuch-Godfrey, e
intuitivamente el El correlograma y la Q de Ljung-Box.

Problemas potenciales con el término de error: 3) Homocedasticidad y No autocorrelación

Formas de solucionar este problema de especificación:

1) Mínimos Cuadrados Generalizados

2) Mínimos Cuadrados Generalizados factibles

3) Análisis de series temporales (siguiente tema)

4) Matrices de covarianzas robustas

- White (estimación robusta ante heteroscedasticidad únicamente)
- Newey y West (estimación robusta ante ambos fenómenos, heteroscedasticidad y autocorrelación)