

# Índice

<b>1. Función de autocovarianzas y función de autocorrelación</b>	<b>2</b>
<b>2. La especificación del modelo se escoge según las características de los datos.</b>	<b>3</b>
<b>3. Herramientas para desvelar propiedades de una serie temporal</b>	<b>3</b>
3.1. Análisis gráfico . . . . .	3
3.1.1. Gráfico de la serie temporal . . . . .	3
3.1.2. Gráfico rango-media . . . . .	3
3.2. Determinación del orden de integración . . . . .	4
3.2.1. Análisis gráfico . . . . .	4
3.2.2. Contrastes formales sobre el orden de integración . . . . .	7
<b>4. Función de autocovarianzas y función de autocorrelación</b>	<b>7</b>
<b>5. Otras herramientas estadísticas</b>	<b>CómoHacerConR 8</b>
5.1. Estadísticos descriptivos . . . . .	8
5.2. Test de normalidad Jarque-Vera . . . . .	8

# Econometría Aplicada. Lección 6

Marcos Bujosa

28 de agosto de 2024

## Resumen

En esta lección veremos algunas herramientas estadísticas. y usaré la librería `tfarima` de José Luis Gallego siempre que pueda.

### Carga de algunas librerías de R

Primero cargamos la librería `tfarima` (Repositorio Cran: <https://cran.r-project.org/web/packages/tfarima/index.html>; repositorio GitHub: <https://github.com/gallegoj/tfarima>)

---

```
library(tfarima) # librería de José Luis Gallego para Time Series
library(readr)  # para leer ficheros CSV
library(ggplot2) # para el scatterplot (alternativamente library(tidyverse))
library(ggfortify) # para pintar series temporales
library(jtools)  # para representación resultados estimación
library(zoo)     # para generar objetos ts (time series)
```

---

y además fijamos los parámetros por defecto para las figuras en `png` del notebook

---

```
# fijamos el tamaño de las figuras que se generan en el notebook
options(repr.plot.width = 12, repr.plot.height = 4, repr.plot.res = 200)
```

---

## 1. Función de autocovarianzas y función de autocorrelación

- La secuencia  $(\gamma_k)$  con  $k \in \mathbb{Z}$  se denomina *función de autocovarianzas*
- La secuencia  $\{\rho_k\}$  con  $k \in \mathbb{Z}$ , donde

$$\rho_k = \frac{\text{Cov}(X_t, X_{t-k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

se denomina *función de autocorrelación* (ACF).

Debido a la estacionariedad, la correlación entre  $X_t$  y  $X_{t+k}$  no depende de  $t$ ; tan solo depende de la distancia temporal  $k$  entre ambas variables.

## 2. La especificación del modelo se escoge según las características de los datos.

¿Es la serie

- "*estacionaria en media*"

- (y si lo es, ¿cuál es su media?)
- (y si no lo es, ¿cómo cambia o evoluciona su media?)

- "*estacionaria en varianza*" (*homocedástica*)

- (y si lo es, ¿cuál es su varianza?)
- (y si es *heterocedástica*, ¿cómo cambia o evoluciona su varianza?)

¿Están sus valores correlados con su historia pasada (autocorrelados)?

¿Están correlados con los valores presentes o pasados de otras series?

Veamos algunas herramientas estadísticas para poder desvelar estas características.

## 3. Herramientas para desvelar propiedades de una serie temporal

### 3.1. Análisis gráfico

#### 3.1.1. Gráfico de la serie temporal

Representa sus valores en el eje vertical ( $y$ ) frente a una escala temporal en el horizontal ( $x$ ). Es útil para detectar visualmente:

- tendencias y/o estacionalidad
- cambios de variabilidad
- valores atípicos (*outliers*)
  - el 95 % aprox. de una muestra de valores generados por una distribución normal debería estar comprendido entre  $\mu \pm 2\sigma$
  - la probabilidad de que una variable normal genere un valor fuera de las bandas de  $\mu \pm 3\sigma$  es 0,0023  
(véase gráfico IBEX 35)

En es importante escalar y rotular adecuadamente los ejes y asegurar la comparabilidad entre series y gráficos distintos (si los hubiere).

#### 3.1.2. Gráfico rango-media

Cambios de variabilidad de una serie pueden evidenciarse en su gráfico temporal. Pero también suelen verse bien en un gráfico rango-media, donde se representa:

**en el eje  $x$**  un indicador del nivel de la serie calculado para distintas submuestras no solapadas (normalmente la media).

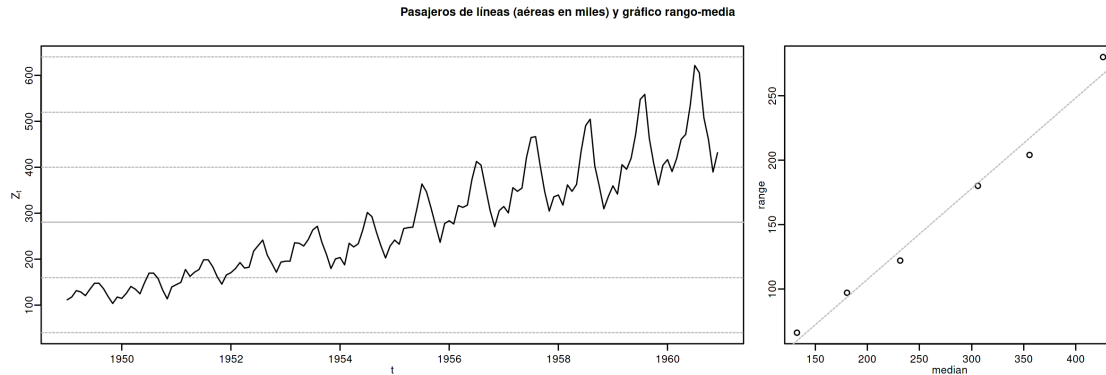
en el eje  $y$  un indicador de la dispersión de la serie calculado para las mismas submuestras (normalmente el rango.)

Veamos el gráfico de la serie de pasajeros de líneas aéreas junto a su gráfico de rango media:

---

```
Z <- AirPassengers
ide(Z, graphs = c("plot", "rm"), main="Pasajeros de líneas (aéreas en miles) y gráfico rango-media")
```

---



El gráfico de rango media a veces se acompaña de una regresión de la dispersión sobre los niveles para medir la relación nivel-dispersión.

### 3.2. Determinación del orden de integración

Decidir adecuadamente el orden de integración es crucial en el análisis de series temporales. Las herramientas utilizadas para tomar la decisión son

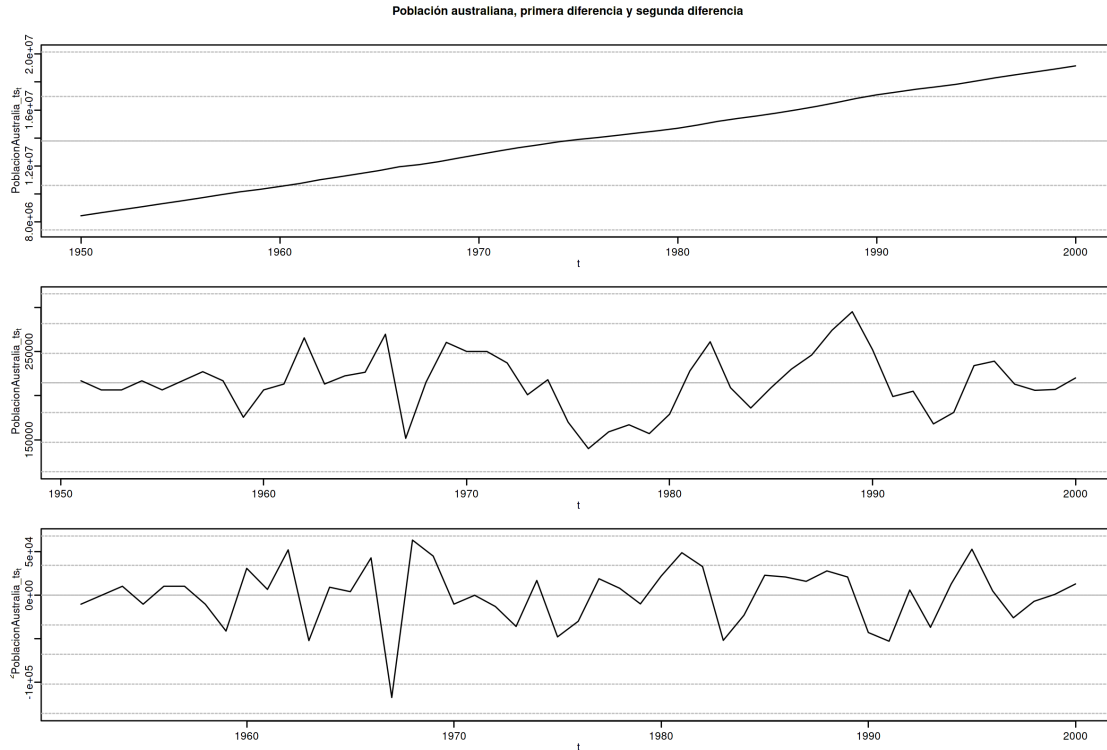
- el análisis gráfico
- los contrastes formales

#### 3.2.1. Análisis gráfico

---

```
options(repr.plot.width = 12, repr.plot.height = 8, repr.plot.res = 200)
ide(PoblacionAustralia_ts,
    graphs = c("plot"),
    transf = list(list(bc = F), list(bc = F, d = 1), list(bc = F, d = 2)),
    main="Población australiana, primera diferencia y segunda diferencia" )
```

---



La serie de población  $y$  tiene una clara tendencia creciente (primer gráfico), que desaparece al tomar una diferencia ordinaria,

$$\nabla y = (1 - B) * y$$

(segundo gráfico). Basta con tomar una primera diferencia de la serie de población para obtener una nueva serie que se asemeja a la realización de un proceso estacionario.

No obstante, ¿qué pasa si tomamos una segunda diferencia ordinaria?

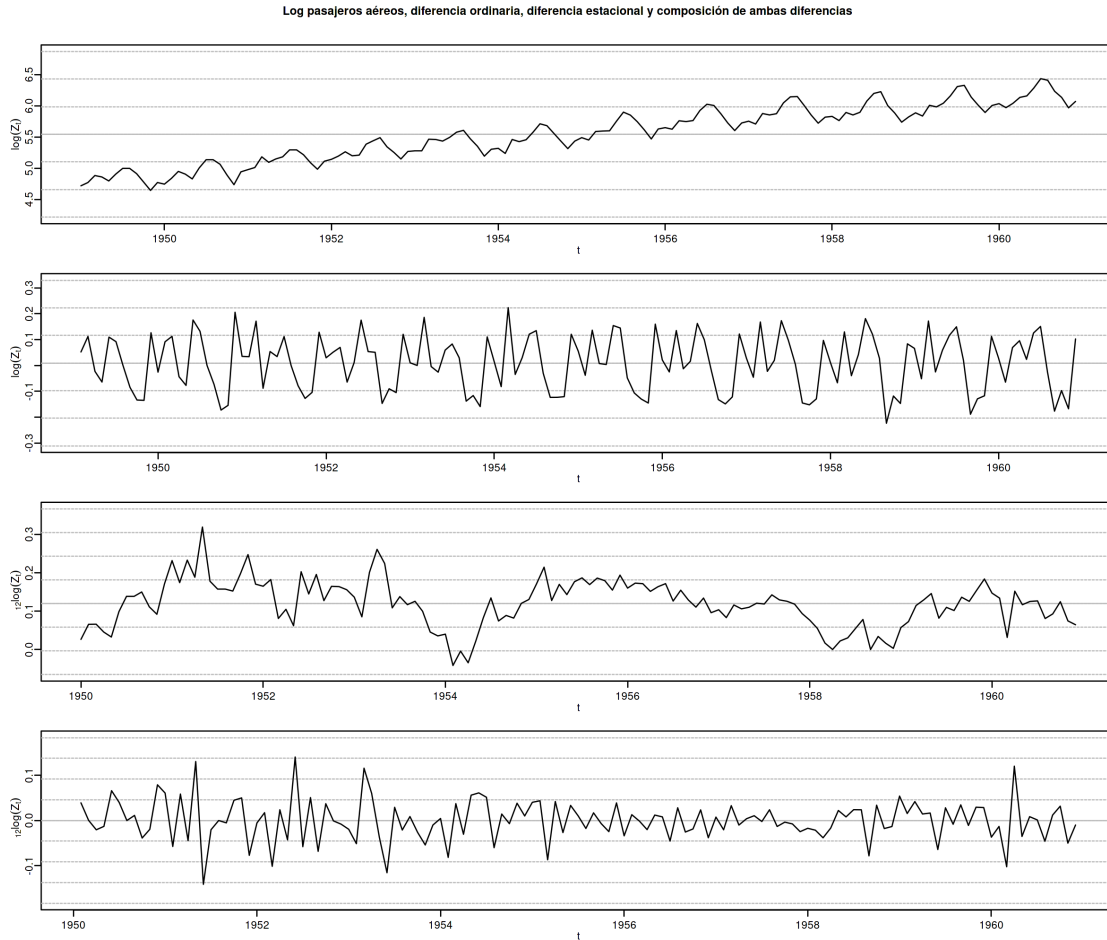
$$\nabla \nabla y = \nabla^2 y = (1 - B)^2 * y$$

(segundo gráfico). Pues que la serie obtenida también es estacionaria, pero ojo, es un grave error tomar más diferencias de las necesarias al modelizar los datos. Se debe tomar el mínimo número de transformaciones que arrojen una serie “estacionaria” (recuerde que decir que una serie temporal es *estacionaria* es un abuso del lenguaje).

---

```
options(repr.plot.width = 12, repr.plot.height = 10, repr.plot.res = 200)
ide(Z,
  graphs = c("plot"),
  transf = list(list(bc=T), list(bc=T, d=1), list(bc=T, D=1), list(bc=T, D=1, d=1)),
  main = "Log pasajeros aéreos, diferencia ordinaria, diferencia estacional y composición de ambas diferencias" )
```

---



Como ya vimos, la serie pasajeros en logaritmos tiene tendencia y estacionalidad muy evidentes. No basta con tomar solo una diferencia ordinaria

$$\nabla \mathbf{y} = (1 - B) * \mathbf{y};$$

pues el resultado muestra una pauta estacional. Ni tampoco basta con tomar solo una diferencia estacional

$$\nabla_{12}(\mathbf{y}) = (1 - B^{12}) * \mathbf{y};$$

pues resulta una serie que “deambula”, i.e., que no es “*estacionaria*” en media.

Tomar una diferencia ordinaria y otra estacional

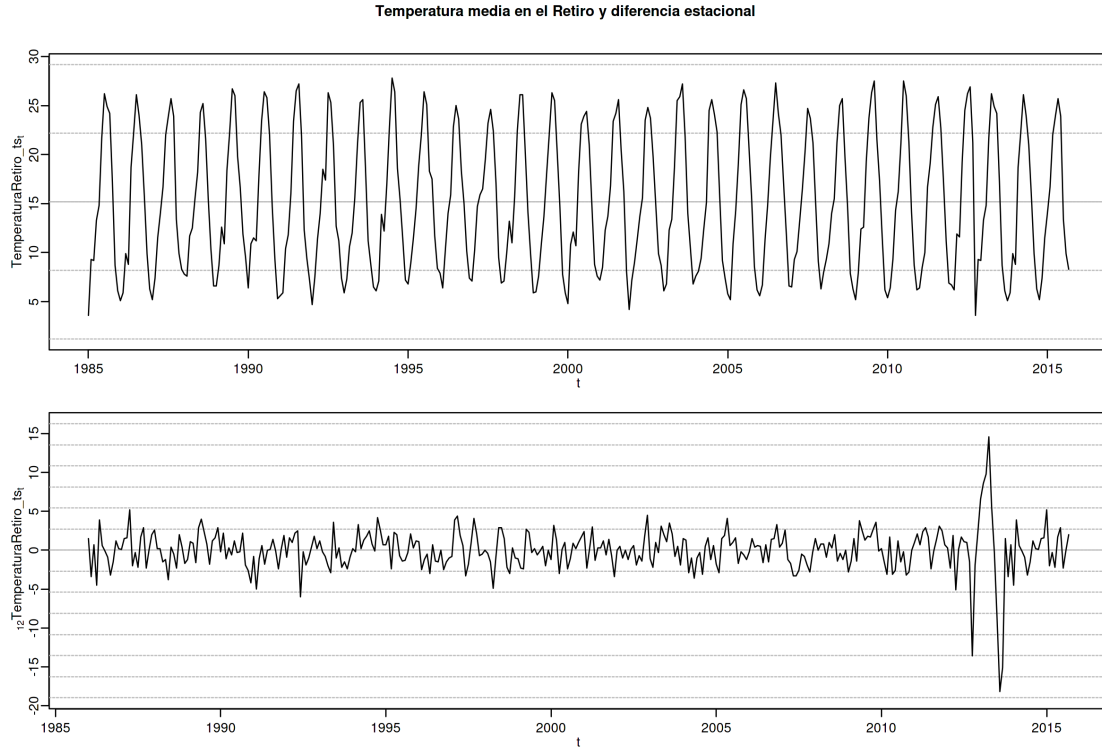
$$\nabla \nabla_{12}(\mathbf{y}) = (1 - B) * (1 - B^{12}) * \mathbf{y}$$

arroja una serie que sí parece ser “*estacionaria*”.

---

```
options(repr.plot.width = 12, repr.plot.height = 8, repr.plot.res = 200)
TemperaturaRetiro_ts=ts(read.csv("datos/Retiro.txt"),start=c(1985, 1), end=c(2015,9), frequency=12)
ide(TemperaturaRetiro_ts,
  graphs = c("plot"),
  transf = list(list(), list(D = 1)),
  main="Temperatura media en el Retiro y diferencia estacional" )
```

---



En el caso de la serie de temperaturas en el Parque del Retiro, parece que es suficiente con tomar solo una diferencia estacional.

### 3.2.2. Contrastes formales sobre el orden de integración

#### 1. Test de Dickey-Fuller (DF)

$H_0$  la serie es  $I(1)$

$H_1$  la serie es  $I(0)$ .

Consideremos el modelo

$$y_t = \rho y_{t-1} + u_t,$$

donde  $y_t$  es la variable de interés,  $\rho$  es un coeficiente, y  $u_t$  es un proceso de ruido blanco. Una raíz unitaria estará presente si  $\rho = 1$ . En tal caso el modelo será no-estacionario.

El modelo de regresión se puede escribir como

$$\nabla y_t = (\rho - 1)y_{t-1} + u_t = \delta y_{t-1} + u_t$$

#### 2. Test de Dickey-Fuller aumentado (ADF)

## 4. Función de autocovarianzas y función de autocorrelación

- La secuencia  $(\gamma_k)$  con  $k \in \mathbb{Z}$  se denomina *función de autocovarianzas*

- La secuencia  $\{\rho_k\}$  con  $k \in \mathbb{Z}$ , donde

$$\rho_k = \frac{\text{Cov}(X_t, X_{t-k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

se denomina *función de autocorrelación* (ACF).

Debido a la estacionariedad, la correlación entre  $X_t$  y  $X_{t+k}$  no depende de  $t$ ; tan solo depende de la distancia temporal  $k$  entre ambas variables.

$$\phi(B) : \phi$$

$$\phi(B)$$

## 5. Otras herramientas estadísticas

## CómoHacerConR

### 5.1. Estadísticos descriptivos

---

```
library(pastecs)      # resumen estadísticos descriptivos
# https://cran.r-project.org/web/packages/pastecs/index.html (stat.desc)
library(knitr)        # presentación de tabla resumen
# https://cran.r-project.org/web/packages/knitr/index.html (kable)
# https://bookdown.org/yihui/rmarkdown-cookbook/kable.html

# estadísticos principales y test de normalidad
kable(stat.desc(Z, basic=FALSE, norm=TRUE), 'rst')
```

---

```
=====
\                                     x
=====
median      265.5000000
mean        280.2986111
SE.mean      9.9971931
CI.mean.0.95 19.7613736
var          14391.9172009
std.dev      119.9663169
coef.var     0.4279947
skewness     0.5710676
skew.2SE     1.4132515
kurtosis     -0.4298441
kurt.2SE     -0.5353818
normtest.W   0.9519577
normtest.p   0.0000683
=====
```

### 5.2. Test de normalidad Jarque-Vera

[Jarque-Vera test \(Wikipedia\)](#)

Podemos calcularlo con la librería [momments](#):



---

```
#install.packages("moments")
library(moments)
# Perform the Jarque-Bera test
jb_test <- jarque.test(as.numeric(Z))
# Print the test result
print(jb_test)
```

---

### Jarque-Bera Normality Test

```
data:  as.numeric(Z)
JB = 8.9225, p-value = 0.01155
alternative hypothesis: greater
```

Otra librería alternativa para calcularlo: [tseries](#)

---

```
library(tseries)
# Perform the Jarque-Bera test
jb_test <- jarque.bera.test(Z)
# Print the test result
print(jb_test)
```

---

Registered S3 method overwritten by 'quantmod':

```
method      from
as.zoo.data.frame zoo
```

### Jarque Bera Test

```
data:  Z
X-squared = 8.9225, df = 2, p-value = 0.01155
```