

Contents

1	Series temporales	2
2	Función de autocovarianzas y función de autocorrelación	5

Econometría Aplicada. Lección 5

Marcos Bujosa

August 19, 2024

En esta lección veremos algunas herramientas estadísticas.

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(latticeExtra) # para gráficos con doble eje vertical (doubleYScale)
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 Series temporales

Corresponden a mediciones de un mismo objeto a lo largo del tiempo. El índice indica el instante de cada medición. *El orden cronológico de los datos puede ser fundamental* para explicar cada uno de ellos.

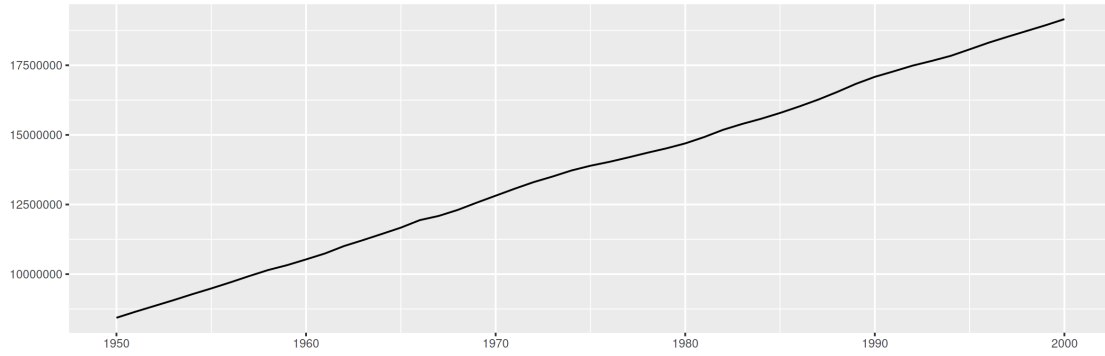
- El motivo es que con frecuencia la medición en un instante de tiempo está relacionada con otras mediciones próximas en el tiempo (*correlación serial*).
- Si es así, no podremos asumir que las variables aleatorias del proceso estocástico subyacente, $\mathbf{X} = (X_t \mid t \in \mathbb{Z})$, sean independientes entre sí.

Esto tiene importantes implicaciones en las técnicas de análisis a aplicar y el tipo de modelos a utilizar.

1. Población en Australia

```
PoblacionAustralia_ts = as.ts( read.zoo('datos/PoblacionAustralia.csv',
                                     header=TRUE,
                                     index.column = 1,
                                     sep=" ",
                                     FUN = as.yearmon))

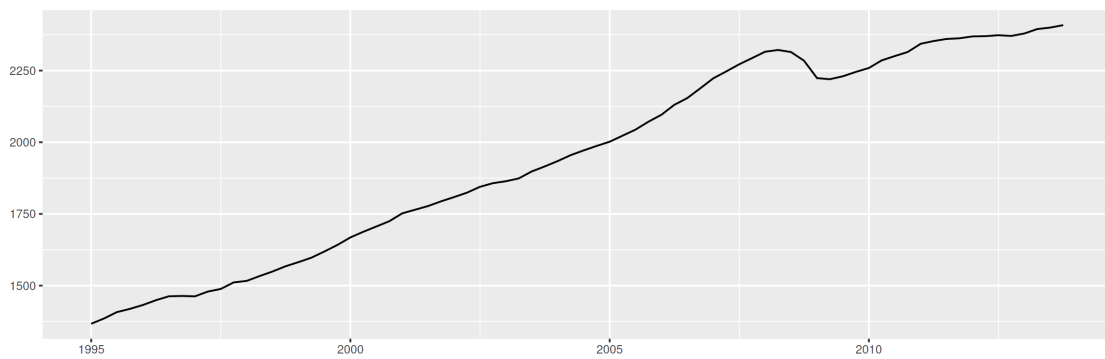
autoplot(PoblacionAustralia_ts)
```



2. PIB UEM

```
PIB_UEM_ts = as.ts( read.csv.zoo("datos/PIB_UEM.csv",
                                FUN = as.yearqtr,
                                format = "%YQ%q",
                                strip.white = TRUE))

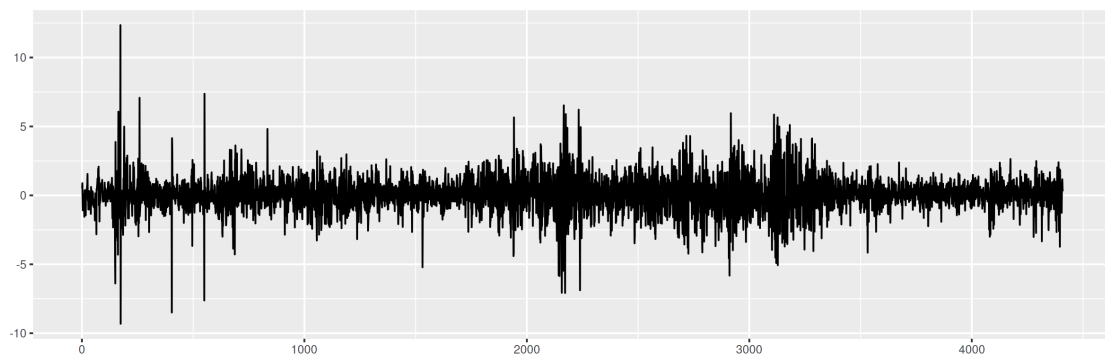
autoplot(PIB_UEM_ts)
```



3. Rendimiento porcentual diario del IBEX 35

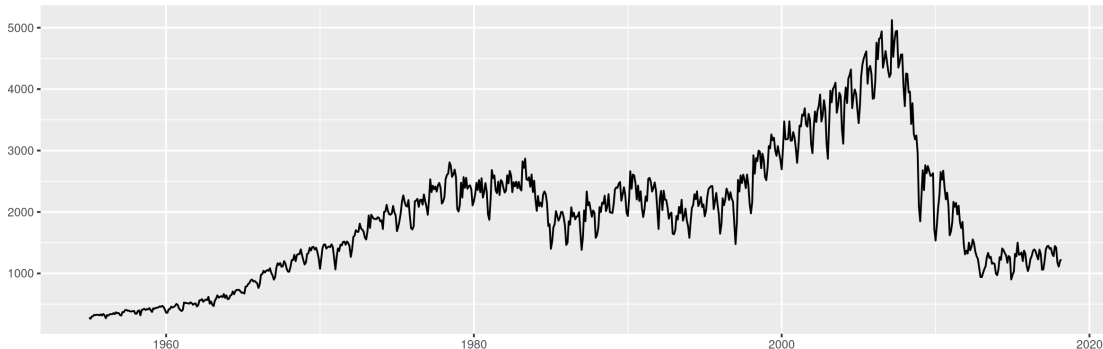
```
IBEX35_ts = as.ts( read.csv.zoo("datos/IBEX35.csv",
                                strip.white = TRUE))

autoplot(IBEX35_ts)
```



4. Producción de cemento

```
ProduccionCemento_ts = as.ts( read.csv.zoo("datos/ProduccionCemento.csv",  
                                     FUN = as.yearmon,  
                                     format = "%Ym",  
                                     strip.white = TRUE))  
  
autoplot(ProduccionCemento_ts)
```



Correlación serial vs muestreo aleatorio simple

- Con datos de sección cruzada se asume muestreo aleatorio simple
 - i.e., los datos son realizaciones de variables aleatorias i.i.d.
- Con series temporales dicha asunción resulta generalmente errónea
 - con frecuencia el valor esperado parece cambiar con t
 - con frecuencia hay dependencia temporal correlación serial

Ejemplo: dada la evolución de los datos, no parece aceptable asumir que $ProdCemento_{1960M01}$ tiene la misma distribución que $ProdCemento_{2000M04}$

Veamos por qué esto genera dificultades...

Consideremos el proceso estocástico

$$\mathbf{X} = (X_t \mid t = 0, \pm 1, \pm 2, \dots).$$

Caracterizar su distribución conjunta (todos los momentos) es demasiado ambicioso.

Tentativamente vamos a fijarnos únicamente en los dos primeros momentos:

$$E(X_t) = \mu_t \quad \text{y} \quad Cov(X_t, X_k) = E[(X_t - \mu_t)(X_k - \mu_k)] = \lambda_{t,k}$$

(si $k = t$ entonces $\lambda_{t,t} = Var(X_t) = \sigma_t^2$).

Si las X_t fueran gaussianas, conocer estos *parámetros* bastaría para caracterizar la distribución conjunta. Pero aún así...

- necesitaríamos una muestra suficiente de cada X_t para estimar los parámetros
 - y en una serie temporal tenemos una única realización para cada X_t
- y además hay más parámetros que variables aleatorias

Casos que simplifican el escenario

- Cuando el proceso es **débilmente estacionario** (Lecc01) se reduce drásticamente el número de parámetros.

$$E(X_t) = \mu \quad (1)$$

$$Cov(X_t, X_{t-k}) = \gamma_k \quad (2)$$

- Si además pudiéramos asumir que el proceso es i.i.d. podríamos interpretar la serie temporal como una realización de un muestreo aleatorio simple (lo que habilita la inferencia estadística).

El desafío para el analista es (y nótese el abuso de lenguaje)

primero transformar los datos para lograr que sean "**estacionarios**"

- (Lo vimos en la lección 1))

después transformar los datos estacionarios en una secuencia de "**datos i.i.d**"

- (es lo que nos falta por ver)

2 Función de autocovarianzas y función de autocorrelación

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

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

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.

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

$$\phi(B)$$