ntic master School

UNIVERSIDAD COMPLUTENSE DE MADRID

Modelos de series temporales estacionarias. Procesos ARIMA

Minería de Datos y Modelización Predictiva

Máster Big Data, Data Science & Business Analytics
Universidad Complutense de Madrid
Curso 2024-2025

INTRODUCCIÓN

Dado su caracter temporal, la evolución de los valores de la serie sulen estar correlacionados con su comportamiento el pasado. Así, aparece el concepto de **autocorrelación**.

ESTIMADORES

En un proceso **estacionario** con valores observados $x_1, x_2, \ldots, x_T$ se tiene que el estimador de la **autocovarianza** de orden k viene dado por:

$$\gamma_k = \frac{1}{T} \sum_{t=k+1}^{T} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

Pudiendo estimar la **autocorrelación** de orden k como:

$$r_k = \frac{\gamma_k}{\gamma_0}$$

Estos coeficientes de correlación muestrales expresados en función del retardo forman la función de autocorrelación muestral ACF y su representación es el **correlograma**.

Estas cantidades miden la relación lineal entre las variables de la serie separadas por k posiciones.

Bajo los mismos supuestos de estaciorariedad, se puede calcular el error de estimación de los coeficientes de autocorrelación como:

$$S_{r_k} = \sqrt{\frac{1 + 2\sum_{j=1}^{k-1} r_j^2}{T}}$$

Distribución y contraste

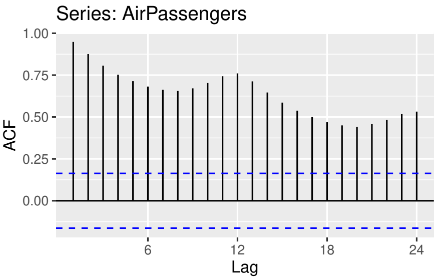
Dado que la distribución del coeficiente de correlación es asintóticamente Normal, podemos construir el intervalo de confianza aproximado bajo la hipótesis de **autocorrelaciones nulas**.

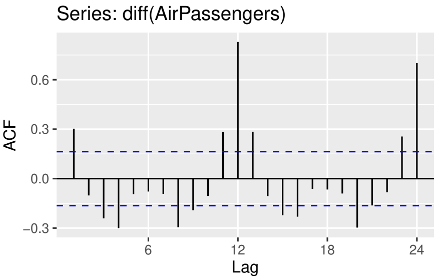
La función ACF es útil para identificar si la serie es o no estacionaria.

La observación y el análisis de los correlogramas nos sirve para detectar, la estacionariedad, el tipo de modelo y los retardos que son significativamente diferentes de cero.

- Decrecimineto progresivo: Indicador de serie **no estacionaria**
- Corte o decrecimiento rápido: Indicador de serie **estacionaria**

Esta función también nos va a indicar el tipo de modelo a ajustar y el orden.



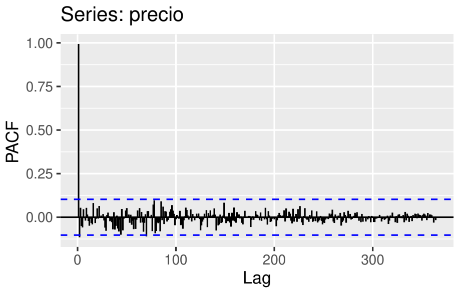


La función de autocorrelación simple mide las correlación con el retardo k. Es posible que esa correlación sea debida a la de los datos contenidos en el periodo retardado, es decir, x_t y x_{t-2} pueden tener correlación por medio de x_{t-1} y no por si mismas!

PACF

Representa las autocorrelaciones de orden k una vez eliminados los efectos de los retardos intermedios. **Autocorrelaciones parciales**.

La función PACF es útil para identificar el orden de los modelos ARIMA



OPERADOR RETARDO

Se define el operador retardo aplicado a un dato temporal como el valor anterior correspondiente al orden retardado, así:

$$BX_t = x_{t-1} \qquad B^2X_t = x_{t-2} \qquad B^pX_t = x_{t-p}$$

DEFINICIÓN

El modelo autoregresivo es una generalización de la idea de regresión para representar la relación entre una variable de la serie y las anteriores.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$$

Haciendo uso del operador retardo, se tiene que,

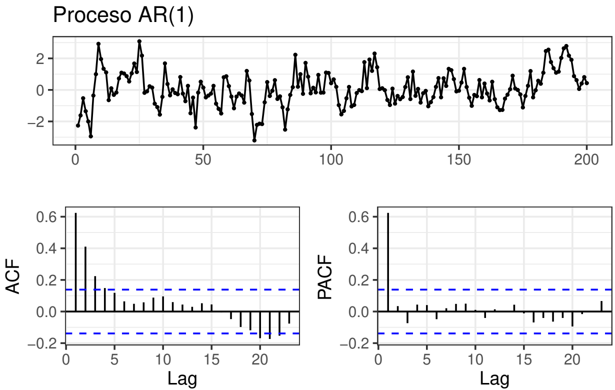
$$X_t = (\phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p)X_t + Z_t = \phi_p(B)X_t + Z_t$$

El polinomio característico $\phi_p(B)$ se relaciona con la estacionariedad de la serie. Así, el proceso será estacionario si las raíces de dicho polinomio tienen módulo mayor que 1.

El modelo más sencillo es el aquel que considera que el valor en el intante t depende del valor en el instante inmediatamente anterior $t-1$, es decir, considera un solo retardo.

$$X_t = \phi_1 X_{t-1} + Z_t \qquad \phi_1(B) X_t = Z_t$$

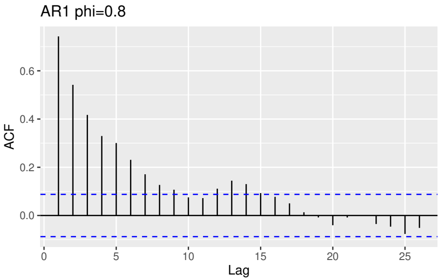
Siendo el proceso es estacionario si $|\phi_1| < 1$. La función de autocorrelación es $\gamma_k = \phi^k$, y decrece geométricamente.

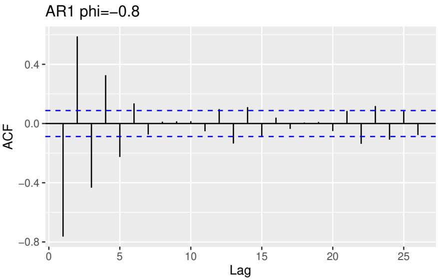


INSPECCIÓN GRÁFICA

Un proceso AR(1) puede tener aspectos distintos en fución del signo del parámetro ϕ_1.

- $\phi_1 > 0$ –> ACF positivo decretciente. PACF con un solo efecto positivo fuera de las bandas.
- $\phi_1 < 0$ –> ACF alternado decretciente. PACF con un solo efecto negativo fuera de las bandas.

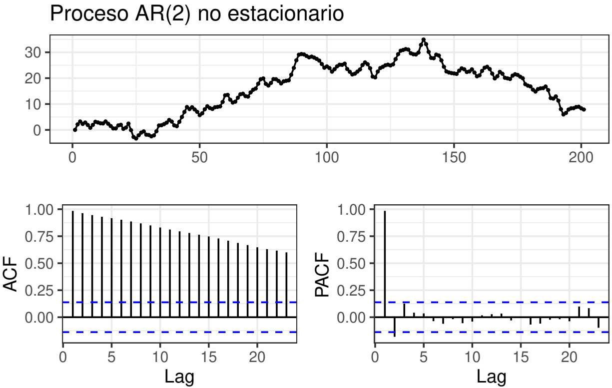




En este caso, se considera que el valor en el intante t depende del valor en los dos tiempos anteriores $t-1$ y $t-2$.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t \qquad \phi_2(B) X_t = Z_t$$

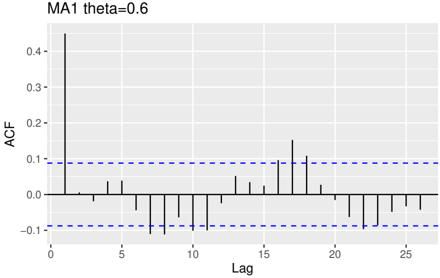
Siendo el proceso es estacionario si $1 - \phi_1 B - \phi_2 B^2$ tiene raíces fuera del círculo unidad.



Los procesos **autoregresivos** representan series que tienen una *memoria larga*, ya que su función de autocorrelación **ACF decrece de forma exponencial** pero no se corta a partir de un determinado retardo.

PISTAS GRÁFICAS

Para representar series de *memoria corta* crearemos los modelos de **medias móviles** cuya identificación es sencilla porque su función de autocorrelación **ACF se corta** a partir de un determinado retardo. Estos procesos son una media de un número finito de innovaciones pasadas.



DEFINICIÓN

Este modelo asume la existencia de dependencia entre los términos del error de distintos retardos. Por tanto, es la suma de procesos estacionarios, cuyo resultado también lo es con independencia del valor del parámetro.

$$X_t = Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \ldots - -\theta_q Z_{t-q} = \theta_q(B) Z_t$$

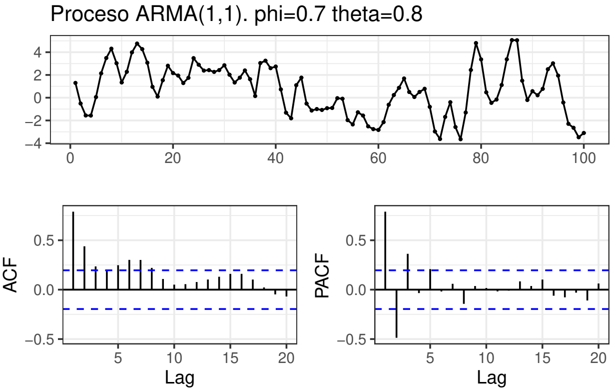
El proceso es invertible y tiene la propiedad de que el efecto de los valores pasados decrece con el tiempo si las raíces del polinomio característico son en módulo mayores que la unidad.

DUALIDAD AR-MA

La función de **autocorrelacion** de un proceso **MA(q)** tiene la misma “forma” que la función de **autocorrelación parcial** de un modelo **AR(q)**. Concluimos que existe una dualidad entre los modelos AR y MA, de manera que la FAS de un MA es como la FAP de un AR y viceversa.

Una extensión natural de los modelos anteriores es aquella que incuye términos autorregresivos y términos de medias móviles. Se representan por la ecuación:

$$(1 - \phi_1 B - .. - \phi_p B^p) X_t = (1 - \theta_1 B - .. - \theta_q B^q) Z_t$$



Como veíamos, existen ciertos patrones que caracterizan a los modelos AR y MA, y dichos patrones son contrapuestos:

- **AR**: Decrece ACF y se corta PACF (el corte indica el retardo de órden AR)
- **MA**: Se corta ACF (el corte indica el retardo de órden MA) y decrece PACF

Además, en general podemos intuir más cosas sobre la serie mediante el análisis de las funciones de autocorrelación simple y parcial.

Forma	Modelo
Serie exponencial decayendo a 0	Modelo Auto Regresivo (AR). Función pacf () que se utilizará para identificar el orden del modelo
Picos alternativos positivos y negativos, decayendo a 0	Modelo Auto Regresivo (AR). Función pacf () que se utilizará para identificar el orden del modelo
Uno o más picos en serie, resto todos son 0	Modelo de media móvil (MA), identifica el orden donde el gráfico se convierte en 0
Después de algunos retrasos en general la serie va decayendo.	Modelo mezclado AR & MA
La serie total es 0 o casi 0	Datos aleatorios
Valores medios a intervalos fijos	Necesitamos incluir el término AR de estacionalidad
Picos visibles que no decaen a 0	Series no son estacionarias

- Cuando el nivel de la serie no es constante en el tiempo, pudiendo en particular tener tendencia creciente o decreciente, diremos que la serie es no estacionaria en la media.
- Cuando la variabilidad o las autocorrelaciones se modifican en el tiempo, diremos que la serie es no estacionaria en la varianza o en las autocorrelaciones.
- Finalmente si la distribución de la variable varía con el tiempo, diremos que la serie es no estacionaria en distribución.

VENTAJAS DE ARIMA

Los procesos no estacionarios más importantes son los procesos integrados, que tienen la propiedad fundamental de que al diferenciarlos se obtienen procesos estacionarios.
Una propiedad importante que diferencia a los procesos integrados de los estacionarios es la forma en que desaparece la dependencia en el tiempo.

Frecuentemente las series económicas no son estacionarias pero sus diferencias relativas, o las diferencias cuando medimos la variable en logaritmos, son estacionarias.

PROCESO INTEGRADO

Se dice que un proceso es **integrado de orde 1** cuando $\Delta X_t = X_t - X_{t-1}$ es estacionaria. En general, se dice que un proceso es **integrado de orde d** cuando hay que aplicar la diferenciación de orden d para conseguir estacionariedad.

FORMULACIÓN DE MODELOS ARIMA

Se puede formular el modelo ARIMA(p,d,q) como la siguiente expresión,

$$(1 - \phi_1 B - \ldots - \phi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B - \ldots - \theta_q B^q) Z_t$$

Se dice que un proceso es **integrado de orde 1** cuando $\Delta X_t = X_t - X_{t-1}$ lo es. Es decir, al diferenciar la serie se obtiene una serie estacionaria.

DIFERENCIACIÓN ESTACIONAL

En el tema de métodos descriptivos vimos que podíamos eliminar la estacionalidad mediante diferencias con los índices estacionales. Podemos **convertir** una serie con **estacionalidad** en **estacionaria** mediante las **diferencias de orden s**, siendo s el periodo de la serie.

Definimos el operador diferencia de periodo s o diferencia estacional de orden 1 como,

$$\Delta^s X_t = X_t - X_{t-s} = (1 - B^s)X_t$$

DEFINICIÓN

El modelo ARIMA estacional general $ARIMA(p, d, q)(P, D, Q)_s$ tiene la expresión:

$$(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B^s)^D (1 - B)^d X_t =$$

$$(1\text{-}\Theta_1 B^s - \dots - \Theta_Q B^{Qs})(1 - \theta_1 B - \dots - \theta_q B^q)Z_t$$

- $ARIMA(1,0,0)(1,0,0)_{12}$

$$(1-\Phi_1 B^{12})(1-\phi_1 B)X_t = Z_t$$

- $ARIMA(1,0,0)(0,0,1)_{12}$

$$(1-\phi_1 B)X_t = (1-\Theta_1 B^{12})Z_t$$

- $ARIMA(2,0,0)(1,0,0)_{12}$

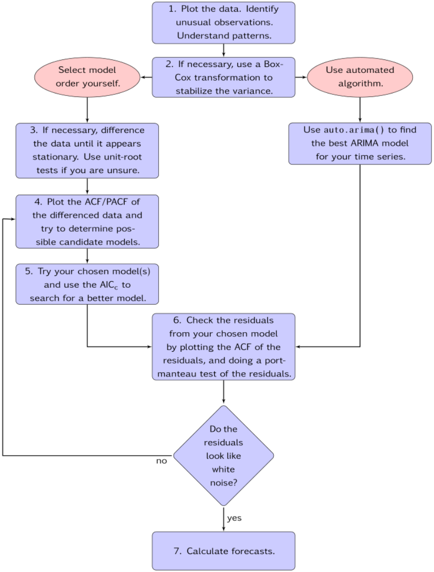
$$(1-\Phi_1 B^{12})(1-\phi_1 B-\phi_2 B^2)X_t = Z_t$$

- $ARIMA(1,1,0)(0,1,1)_{12}$

$$(1-\phi_1 B)(1-B^{12})(1-B)X_t = (1-\Theta_1 B^{12})Z_t$$

PROCESO DE ESTUDIO DE UNA SERIE TEMPORAL

- **Paso 1.** Identificación del modelo: Utilizamos los datos históricos de la serie para encontrar el modelo apropiado.
  - Representación. Descomposición. Diferenciación regular o estacional.
  - Funciones de autocorrelación.
- **Paso 2.** Estimación: Estimamos los parámetros del modelo escogido utilizando los datos históricos.
  - Manual
  - Automático
- **Paso 3.** Pruebas del modelo: Realizamos distintos contrastes para decidir si el modelo construido es adecuado. Si no lo es volveríamos al paso 1.
  - Gráficos de residuos. Ljung-Box test.
  - Errores de predicción. RMSE (Root Mean Square Error)/MAE (Mean Absolute Error)
- **Paso 4.** Predicción: Una vez que el modelo ha sido construido y comprobada su adecuación lo utilizamos para hacer predicciones.



Fuente: `https://otexts.com/fpp2/`