

Índice

| | |
|--|-----------|
| 1. Introducción | 2 |
| 2. Predicciones de error cuadrático medio mínimo | 3 |
| 2.1. Predicciones lineales | 4 |
| 2.1.1. Breve recordatorio sobre las proyecciones ortogonales | 4 |
| 2.1.2. Las predicciones lineales son proyecciones ortogonales | 4 |
| 2.1.3. Proyección sobre el pasado de un proceso lineal causal, estacionario e invertible | 5 |
| 3. Cálculo de predicciones de procesos ARIMA | 6 |
| 3.1. Ejemplos | 7 |
| 3.1.1. ARMA(1,1) | 7 |
| 3.1.2. AR(1) | 8 |
| 3.1.3. MA(1) | 8 |
| 3.1.4. Paseo aleatorio sin deriva | 9 |
| 3.1.5. Paseo aleatorio con deriva | 9 |
| 3.1.6. IMA(1) | 9 |
| 3.1.7. IMA(1) estacional | 10 |
| 4. Varianza de las predicciones lineales | 10 |
| 4.1. Intervalos de confianza para las previsiones | 11 |
| 5. Adaptación de las predicciones a las nuevas observaciones | 12 |
| 6. Medidas del error de previsión | 12 |
| 7. Modelos con logaritmos | 13 |
| 8. Nota sobre subespacio sobre el que proyectamos al prever | 13 |

Lección 8. Predicción de modelos ARIMA

Marcos Bujosa

13 de noviembre de 2025

Resumen

Esta lección se centra en la predicción de modelos de series temporales univariantes.

- ([slides](#)) — ([html](#)) — ([pdf](#)) — ([mybinder](#))

Carga de algunos módulos de python y creación de directorios auxiliares

```
# Para trabajar con los datos y dibujarlos necesitamos cargar algunos módulos de python
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib as mpl
# definimos parámetros para mejorar los gráficos
mpl.rc('text', usetex=False)
import matplotlib.pyplot as plt # data visualization
import dataframe_image as dfi # export tables as .png
```

- Directorio auxiliar para albergar las figuras de la lección: para publicar la lección como pdf o página web, necesito los gráficos como ficheros `.png` alojados algún directorio específico:

```
imagenes_leccion = "./img/lecc07" # directorio para las imágenes de la lección
import os
os.makedirs(imagenes_leccion, exist_ok=True) # crea el directorio si no existe
```

Gráficos para las ACF, PACF y densidades espectrales teóricas

Cargamos las funciones auxiliares (véase la carpeta `src/`)

```
import warnings
warnings.filterwarnings("ignore", category=UserWarning)
%run -i ./src/analisis_armas.py
```

1. Introducción

Consideremos una serie temporal $\mathbf{y}_1^T = (y_1, \dots, y_T)$. En relación con la predicción de los valores futuros de la serie, hay tres aspectos a considerar :

- Necesitamos un método para hacer predicciones basadas en un modelo ajustado a las observaciones.

- Necesitamos evaluar la incertidumbre asociada a estas predicciones. Sin ello la utilidad de los pronósticos se ve comprometida.

Es común expresar la incertidumbre con un intervalo de confianza para la predicción; un rango que con alta probabilidad (como 0.95 o 0.99) incluye el valor futuro pronosticado.

- Además, la previsión es un proceso dinámico: al recibir nuevos datos, es necesario actualizar los pronósticos con la nueva información.

Los modelos lineales facilitan solventar estas tres necesidades.

2. Predicciones de error cuadrático medio mínimo

Queremos predecir $Y_{t+\ell}$ usando una función g de variables aleatorias con índice $n \leq t$, es decir, una función del conjunto $\mathbf{Y}_{|:t} = \{Y_n \mid n \leq t\}$.

Nos referiremos al conjunto $\mathbf{Y}_{|:t}$ como “*el pasado*” de \mathbf{Y} .

Para evaluar diferentes funciones de predicción, estableceremos como criterio *minimizar el error cuadrático medio* (MSE):

$$MSE(Y_{t+\ell}, g_t(\ell)) = E\left(Y_{t+\ell} - g_t(\ell)\right)^2.$$

Preferiremos aquel método $g_t(\ell)$ que minimiza la esperanza del cuadrado del error de predicción $Y_{t+\ell} - g_t(\ell)$; donde $g_t(\ell)$ es la función de predicción de $Y_{t+\ell}$, t es el índice de la última variable aleatoria considerada y ℓ representa el horizonte de predicción.

La función $g_t(\ell)$ que minimiza dicho criterio MSE es la **esperanza condicional** de $Y_{t+\ell}$:

$$E(Y_{t+\ell} \mid \mathbf{Y}_{|:t}).$$

Nota. La esperanza condicional es la proyección ortogonal sobre el espacio (de Hilbert) generado por el conjunto de todas las funciones (medibles) de $\mathbf{Y}_{|:t}$. Dicho conjunto es enorme y, entre otras muchas funciones, también contiene al conjunto de funciones de $\mathbf{Y}_{|:t}$ que son *lineales*.

Las predicciones resultantes se llaman predicciones de error cuadrático medio mínimo.

El problema de la esperanza condicional radica en que, según el caso, su cálculo puede ser muy difícil... o sencillamente inviable. Dicho de otro modo, determinar específicamente cuál es la transformación de $\mathbf{Y}_{|:t}$ que minimiza los errores de predicción puede ser imposible en la práctica.

Como solución alternativa podemos limitarnos a buscar predictores entre:

- las funciones de $\mathbf{Y}_{|:t}$ (i.e., funciones “del pasado”) que son *lineales*;
- y seleccionar aquella con menor error cuadrático medio (MSEL).

Al adoptar esta estrategia, aceptamos cometer mayores errores de predicción, $MSEL \geq MSE$, pues al excluir funciones no lineales, podríamos estar omitiendo la que genera los errores más pequeños.

Pregunta: ¿Cómo tiene que ser la esperanza condicional para que $MSEL$ y MSE sean iguales? ¿En qué caso ocurrirá?

2.1. Predicciones lineales

2.1.1. Breve recordatorio sobre las proyecciones ortogonales

Solo recordaremos que como

la proyección ortogonal de X sobre \mathcal{H} es el elemento de \mathcal{H} más *próximo* a X

- si $X \in \mathcal{H}$, la proyección de X sobre \mathcal{H} es el propio X .
- si Y es ortogonal a todo $X \in \mathcal{H}$, es decir, si $Y \perp \mathcal{H}$, la proyección de Y sobre \mathcal{H} es cero.

La proyección ortogonal de X sobre el espacio euclídeo \mathcal{H} se caracteriza porque la diferencia entre X y su proyección, \hat{X} , es perpendicular a \mathcal{H} .

En el contexto de la predicción lineal, dicha diferencia $X - \hat{X}$ recibirá el nombre de *error de predicción*.

Puede consultar algunos resultados sobre las proyecciones ortogonales en este [enlace](#).

2.1.2. Las predicciones lineales son proyecciones ortogonales

La forma general de un predictor *lineal* es

$$g_t(\ell) = b_{\ell 0}Y_t + b_{\ell 1}Y_{t-1} + b_{\ell 2}Y_{t-2} + \cdots = \mathbf{b}_\ell * (\mathbf{Y}_{|:t}) = \mathbf{b}_\ell(B)Y_t,$$

donde \mathbf{b}_ℓ es una serie formal de cuadrado sumable.

El error cuadrático medio del predictor lineal (MSEL) es mínimo cuando el error de predicción $(Y_{t+\ell} - g_t(\ell))$ es ortogonal a cada una de las variables aleatorias del pasado del proceso;¹ es decir, cuando

$$E\left((Y_{t+\ell} - \mathbf{b}_\ell * \mathbf{Y}_{|:t}) \cdot Y_n\right) = 0, \text{ para cualquier } n \leq t.$$

Consecuentemente, el mejor predictor lineal (que denotaremos con $\widehat{Y_{t+\ell|t}}$) es la proyección ortogonal de $Y_{t+\ell}$ sobre la *clausura*² del subespacio formado por las combinaciones lineales de 1 y las variables aleatorias $Y_t, Y_{t-1}, Y_{t-2}, \dots$

Denotando dicha clausura con \mathcal{H}_{Y_t} , podemos decir que el predictor lineal que minimiza los errores de previsión, $\widehat{Y_{t+\ell|t}}$, es la proyección ortogonal de $Y_{t+\ell}$ sobre \mathcal{H}_{Y_t} .

Es habitual referirse a los elementos de \mathcal{H}_{Y_t} como *variables observadas*;³ y a todo el espacio \mathcal{H}_{Y_t} como el *conjunto de información* empleado en la previsión.

¹Cuando el proceso \mathbf{Y} tiene media cero, esta condición es equivalente a que el error de previsión esté incorrelado con el pasado del proceso.

²Sea $\overline{\text{sp}}\{1, Y_t, Y_{t-1}, \dots\}$ la clausura en la norma de L^2 del subespacio vectorial engendrado por la constante 1 y el “pasado” de \mathbf{Y} . Es decir, completamos el conjunto $\text{sp}\{1, Y_t, Y_{t-1}, \dots\}$ de las combinaciones lineales de la constante 1 y las variables $\{1, Y_t, Y_{t-1}, \dots\}$, con los límites de todas las sucesiones de Cauchy de elementos de $\text{sp}\{1, Y_t, Y_{t-1}, \dots\}$ (es decir, le añadimos la clausura). De esta manera toda sucesión de Cauchy de elementos del espacio converge en norma a algún elemento del propio espacio, pues hemos añadido los límites (así tendremos un espacio de Hilbert donde la proyección ortogonal está bien definida). Para abreviar, denotaremos $\overline{\text{sp}}\{1, Y_t, Y_{t-1}, \dots\}$ con \mathcal{H}_{Y_t} .

³Este nombre viene sugerido por el contexto de la previsión de datos reales. Pero es solo una interpretación del formalismo matemático. Fíjese que los objetos con los que estamos tratando son *variables aleatorias*, es decir, funciones matemáticas. Las funciones matemáticas son una construcción intelectual por lo que, en ese sentido, hablar de *variables aleatorias del pasado* es similar a hablar de *los argonautas del futuro* o *los unicornios del pasado*. Lo que es relevante es que dichas variables son elementos del espacio \mathcal{H}_{Y_t} sobre el que estamos proyectando las variables.

Cuando $\mathbf{Y} = \varphi * \mathbf{U}$ es un proceso estocástico lineal causal, estacionario e invertible, se cumple que $\mathcal{H}_{Y_t} = \mathcal{H}_{U_t}$ (véase la sección [Nota sobre subespacio sobre el que proyectamos al prever](#)).⁴

Restringir la búsqueda a funciones lineales implica que, si la esperanza condicional no es lineal respecto al pasado, el mejor predictor lineal será inferior a la esperanza condicional.

$$MSE \leq MSEL.$$

La igualdad se da solo cuando la esperanza condicional es lineal en $\mathbf{Y}_{|:t}$; y por tanto no ha sido excluida por la restricción. En resumen, $MSEL = MSE$ solo cuando la esperanza condicional $E(Y_{t+\ell} | \mathbf{Y}_{|:t})$ coincide con la proyección ortogonal de $Y_{t+\ell}$ sobre \mathcal{H}_{Y_t} (que es exactamente lo que ocurre cuando el proceso estocástico tiene distribución normal).

Los resultados anteriores, aplicables a procesos estacionarios infinitos, también son válidos para procesos no estacionarios si consideramos un tramo finito del pasado de \mathbf{Y} , es decir, si nuestra predicción depende de $\mathbf{Y}_{|1:T} = (Y_1, \dots, Y_T)$. En tal caso la discusión se centra en proyectar (o condicionar la esperanza⁵) sobre el conjunto finito de variables aleatorias del pasado $\mathbf{Y}_{|1:T}$. Así, tanto la esperanza como la varianza condicional serán finitas, lo que asegura que todo esté bien definido.

2.1.3. Proyección sobre el pasado de un proceso lineal causal, estacionario e invertible

Sea $\mathbf{Y} = \varphi * \mathbf{U}$ es un proceso estocástico lineal causal, estacionario e invertible.

Para $\ell > 0$,

- denominamos *previsión de $Y_{t+\ell}$ basada en la historia del proceso hasta Y_t* a la proyección ortogonal de $Y_{t+\ell}$ sobre \mathcal{H}_{Y_t} ; que denotamos con $\widehat{Y_{t+\ell}|t}$.
- Análogamente, la “*previsión $\widehat{U_{t+\ell}|t}$ basada en el pasado de \mathbf{Y}* ” es la proyección ortogonal de $U_{t+\ell}$ sobre $\mathcal{H}_{Y_t} = \mathcal{H}_{U_t}$.

Fíjese que $\widehat{U_{t+\ell}|t}$ siempre es cero; pues el ruido blanco es ortogonal a su pasado, $U_{t+\ell} \perp \mathcal{H}_{U_t}$, por ser incorrelado y con esperanza nula⁶

Sin embargo, para los índices $(t - k)$ donde $k \geq 0$:

- la proyección de Y_{t-k} y la de U_{t-k} sobre \mathcal{H}_{Y_t} son las propias Y_{t-k} y U_{t-k} respectivamente, ya que ambas variables aleatorias pertenecen al espacio $\mathcal{H}_{Y_t} = \mathcal{H}_{U_t}$ sobre el que se proyectan (i.e., son *variables observadas*).

Además para cualquier índice j :

- Con las secuencias constantes ($c_t = c$ para todo t) tenemos que $\widehat{c_j|t} = c$ dado que $1 \cdot c \in \mathcal{H}_{Y_t}$.

⁴donde \mathcal{H}_{U_t} denota la clausura en la norma de L^2 del subespacio generado por la constante 1 y el pasado del proceso de ruido blanco \mathbf{U}

⁵entonces estaremos proyectado ortogonalmente sobre funciones medibles de un conjunto finito de variables aleatorias: $\{Y_1, \dots, Y_T\}$.

⁶Para una discusión sobre correlación y ortogonalidad, léase la sección 3.1.1 de los apuntes <https://github.com/mbujosab/Ectr/blob/master/apuntes-Ectr-print.pdf>

Tras este repaso de las proyecciones ortogonales, ya podemos ver cómo calcular predicciones para los modelos ARIMA causales.

3. Cálculo de predicciones de procesos ARIMA

Sea \mathbf{Y} un proceso $\text{ARIMA}(p, d, q) \times (P, D, Q)_S$

$$\phi_p(\mathbf{B}) \Phi_P(\mathbf{B}^S) \nabla^d \nabla_S^D Y_t = c + \theta_q(\mathbf{B}) \Theta_q(\mathbf{B}^S) U_t; \quad t \in \mathbb{N}.$$

donde c puede ser distinto de cero para permitir procesos con esperanza no nula.

Denotemos con φ al polinomio de grado $p + P + d + D$ resultante del producto convolución de los polinomios de la parte AR:

$$\varphi(z) = \phi_p(z) * \Phi_P(z^S) * \nabla^d * \nabla_S^D.$$

Y denotemos con ϑ al polinomio de grado $q + Q$ resultante del producto convolución de los polinomios de la parte MA:

$$\vartheta(z) = \theta_q(z) * \Theta_q(z^S).$$

Entonces podemos reescribir el modelo ARIMA de una manera mucho más compacta:

$$\varphi(\mathbf{B}) Y_t = c + \vartheta(\mathbf{B}) U_t.$$

Debido a la simple estructura de los modelos ARIMA, tenemos que

$$Y_t = c + \sum_{k=1}^p \varphi_k Y_{t-k} + U_t + \sum_{k=1}^q -\vartheta_k U_{t-k}. \quad (1)$$

De (1), para $\ell = 0, \pm 1, \pm 2, \dots$ y t arbitrario

$$Y_{t+\ell} = c + \sum_{k=1}^p \varphi_k Y_{t+\ell-k} + U_{t+\ell} + \sum_{k=1}^q -\vartheta_k U_{t+\ell-k}, \quad (2)$$

Dividiendo los sumatorios de (2) en dos partes: por un lado la suma desde $k = 1$ hasta $\ell - 1$, y por otro la suma del resto (de ℓ en adelante); tenemos que⁷

$$Y_{t+\ell} = c + \sum_{k=1}^{\ell-1} \varphi_k Y_{t+\ell-k} + \sum_{k=\ell}^p \varphi_k Y_{t+\ell-k} + U_{t+\ell} + \sum_{k=1}^{\ell-1} -\vartheta_k U_{t+\ell-k} + \sum_{k=\ell}^q -\vartheta_k U_{t+\ell-k}, \quad (3)$$

Indicar la relación de cada elemento con el subespacio \mathcal{H}_{Y_t} nos ayudará a calcular las previsiones:

$$Y_{t+\ell} = \underbrace{c}_{\in \mathcal{H}_{Y_t}} + \underbrace{\sum_{k=1}^{\ell-1} \varphi_k Y_{t+\ell-k}}_{\notin \mathcal{H}_{Y_t} \text{ pues } k < \ell} + \underbrace{\sum_{k=\ell}^p \varphi_k Y_{t+\ell-k}}_{\in \mathcal{H}_{Y_t} \text{ pues } k \geq \ell} + \underbrace{U_{t+\ell}}_{\perp \mathcal{H}_{Y_t}} + \underbrace{\sum_{k=1}^{\ell-1} -\vartheta_k U_{t+\ell-k}}_{\perp \mathcal{H}_{Y_t} \text{ pues } k < \ell} + \underbrace{\sum_{k=\ell}^q -\vartheta_k U_{t+\ell-k}}_{\in \mathcal{H}_{Y_t} \text{ pues } k \geq \ell}; \quad \text{con } \ell \geq 1. \quad (4)$$

⁷Recuerde que en la notación con sumatorios se asume que si el limite inferior del índice (el que aparece abajo del sumatorio) es mayor que el valor que aparece arriba, el sumatorio es cero: si $m > n$ entonces $\sum_{i=m}^n a_i = 0$.

Por linealidad de la proyección ortogonal, el predictor ℓ -pasos adelante de Y_t es

$$\widehat{Y_{t+\ell|t}} = c + \sum_{k=1}^{\ell-1} \varphi_k \widehat{Y_{t+\ell-k|t}} + \sum_{k=\ell}^p \varphi_k Y_{t+\ell-k} + \sum_{k=\ell}^q -\vartheta_k U_{t+\ell-k} : \quad (5)$$

donde lo que era ortogonal a \mathcal{H}_{Y_t} se anula, y lo que pertenece a \mathcal{H}_{Y_t} no cambia.

A partir de esta expresión podemos obtener fórmulas recursivas tanto para los predictores como para sus actualizaciones cuando se disponen de nuevas observaciones.

3.1. Ejemplos

3.1.1. ARMA(1,1)

$$Y_{t+\ell} = c + \phi_1 Y_{t+\ell-1} + U_{t+\ell} - \theta_1 U_{t+\ell-1},$$

la ecuación (5) da lugar a dos casos en función de ℓ

$$\begin{cases} \widehat{Y_{t+1|t}} = c + \phi_1 Y_t - \theta_1 U_t & \text{cuando } \ell = 1 \\ \widehat{Y_{t+\ell|t}} = c + \phi_1 \widehat{Y_{t+\ell-1|t}} & \text{cuando } \ell \geq 2. \end{cases} \quad (6)$$

donde la segunda ecuación es recursiva (y la primera nos indica el paso inicial). Así:

Para $\ell = 1$ $\widehat{Y_{t+1|t}} = c + \phi_1 Y_t - \theta_1 U_t.$

Para $\ell = 2$

$$\begin{aligned} Y_{t+2} &= c + \phi_1 \widehat{Y_{t+1|t}} \\ &= c + \phi_1 (c + \phi_1 Y_t - \theta_1 U_t) \\ &= c(1 + \phi_1) + \phi_1^2 Y_t - \phi_1 \theta_1 U_t \end{aligned}$$

Para $\ell = 3$

$$\begin{aligned} Y_{t+3} &= c + \phi_1 \widehat{Y_{t+2|t}} \\ &= c + \phi_1 (c(1 + \phi_1) + \phi_1^2 Y_t - \phi_1 \theta_1 U_t) \\ &= c(1 + \phi_1 + \phi_1^2) + \phi_1^3 Y_t - \phi_1^2 \theta_1 U_t. \end{aligned}$$

Para $\ell = k$ Repitiendo el procedimiento llegamos a que:

$$\widehat{Y_{t+\ell|t}} = c(1 + \phi_1 + \dots + \phi_1^{\ell-1}) + \phi_1^\ell Y_t - \phi_1^{\ell-1} \theta_1 U_t.$$

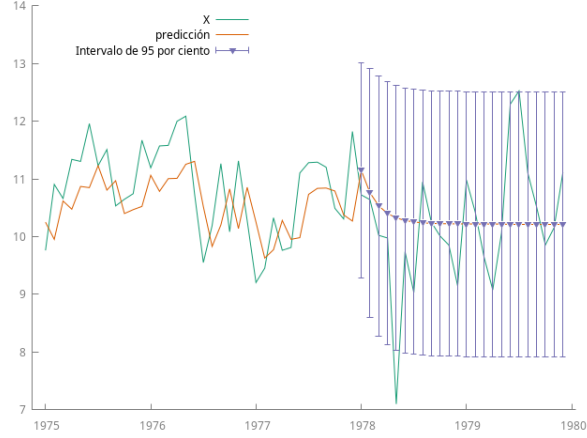
Como $|\phi_1| < 1$ y $|\theta| < 1$, la predicción cuando $\ell \rightarrow \infty$ tiende al valor esperado del proceso

$$\lim_{\ell \rightarrow \infty} \widehat{Y_{t+\ell|t}} = \frac{c}{1 - \phi_1} = \mu.$$

3.1.2. AR(1)

Basta sustituir $\theta = 0$ en (6) para particularizar el resultado un modelo AR(1),

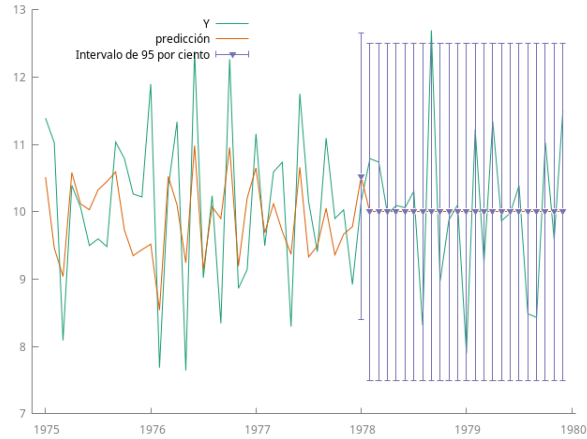
$$\widehat{Y_{t+\ell|t}} = \frac{c}{1-\phi_1} + \phi_1^\ell Y_t, \quad \ell = 1, 2, \dots,$$



3.1.3. MA(1)

De igual manera, sustituyendo $\phi = 0$ en (6) particularizamos para un modelo MA(1),

$$\begin{cases} \widehat{Y_{t+1|t}} &= \frac{c}{1-\phi_1} - \theta U_t, \\ \widehat{Y_{t+\ell|t}} &= \frac{c}{1-\phi_1}, \quad \text{para } \ell \geq 2. \end{cases}$$

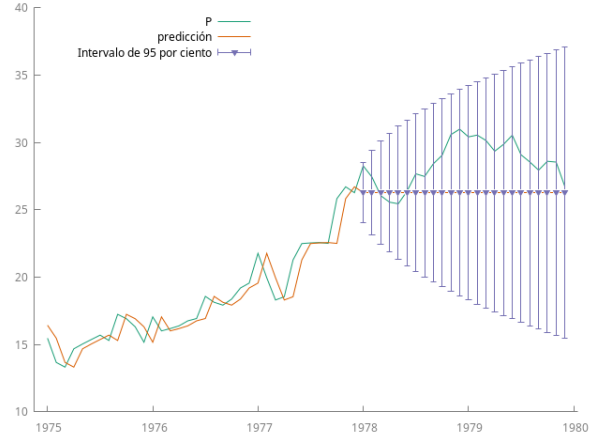


En ambos casos se observa que $\widehat{Y_{t+\ell|t}}$ converge a su valor esperado $\mu = E(Y_{t+\ell})$ cuando $\ell \rightarrow \infty$. Esta propiedad se verifica en todos los modelos ARMA estacionarios, causales e invertibles.

3.1.4. Paseo aleatorio sin deriva

Para un paseo aleatorio $\phi = 1$, $\theta = 0$ y $\mu = 0$ en (6):

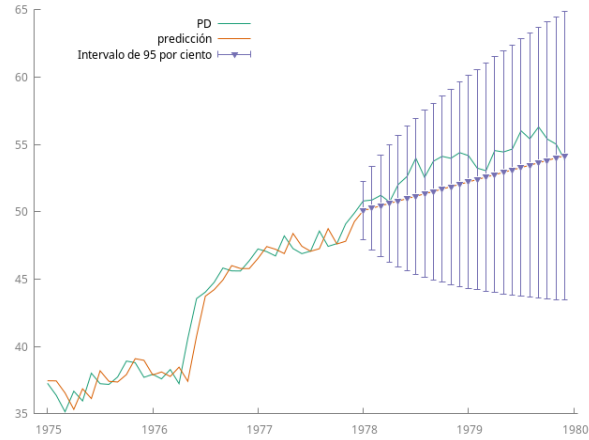
$$\widehat{Y_{t+\ell|t}} = Y_t, \quad \ell = 1, 2, \dots,$$



3.1.5. Paseo aleatorio con deriva

Si el paseo aleatorio tiene deriva, el modelo es $\phi = 1$, $\theta = 0$ y $\mu \neq 0$ en (6):

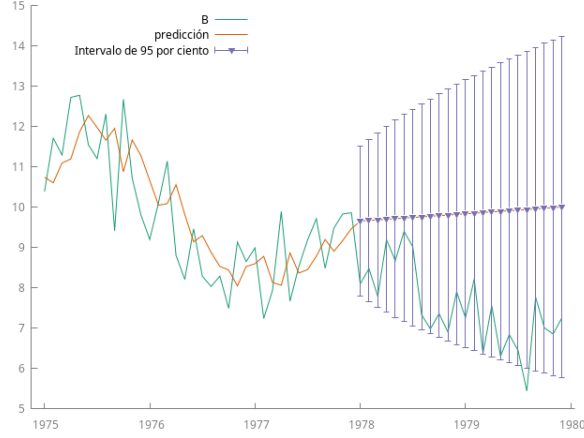
$$\widehat{Y_{t+\ell|t}} = c \cdot \ell + Y_t, \quad \ell = 1, 2, \dots,$$



3.1.6. IMA(1)

Basta sustituir $\phi_1 = 1$ en (6) para particularizar el resultado a un modelo IMA(1),

$$\begin{cases} \widehat{Y_{t+1|t}} &= c + Y_t - \theta_1 U_t. \\ \widehat{Y_{t+\ell|t}} &= c \cdot \ell + Y_t - \theta_1 U_t \quad \text{cuando } \ell \geq 2. \end{cases}$$

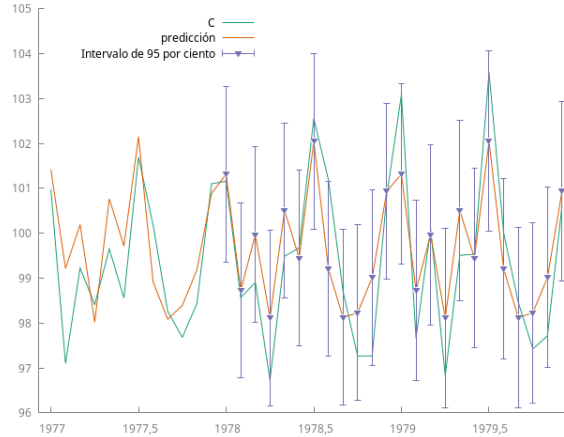


3.1.7. IMA(1) estacional

Para $\nabla_{12}Y_t = (1 - \Theta_1 B^{12})U_t$ (con $\mu = 0$ para simplificar); donde $j = 1, 2, \dots, 12$:

$$\text{predicción de cada mes } \begin{cases} \widehat{Y_{t+j}|t}} &= Y_{t+j-12} - \Theta_1 U_{t+j-12-1} \\ \widehat{Y_{t+j+(12\ell)|t}} &= \widehat{Y_{t+j}|t} \quad \text{cuando } \ell \geq 1. \end{cases}$$

Desde el segundo año, cada predicción es como la del mismo mes del año anterior.



4. Varianza de las predicciones lineales

Consideremos primero modelos lineales, causales y estacionarios.

Sea $Y_t = \psi(B)U_t$ la representación MA del proceso, donde $\psi = \phi^{-1} * \theta$. Entonces:

$$Y_{t+\ell} = \sum_{i=0}^{\infty} \psi_i U_{t+\ell-i}.$$

La predicción lineal óptima será su proyección ortogonal sobre \mathcal{H}_{Y_t} :

$$\widehat{Y_{t+\ell}|t}} = \sum_{j=0}^{\infty} \psi_{\ell+j} U_{t-j}. \quad (7)$$

Restando $\widehat{Y_{t+\ell|t}}$ de $Y_{t+\ell}$ obtenemos el error de predicción:

$$e_t(\ell) = Y_{t+\ell} - \widehat{Y_{t+\ell|t}} = U_{t+\ell} + \psi_1 U_{t+\ell-1} + \cdots + \psi_{\ell-1} U_{t+1},$$

cuya esperanza es cero.

Asumiendo que \mathbf{Y} tiene distribución gaussiana, $\widehat{Y_{t+\ell|t}} = E(Y_{t+\ell} | \mathbf{Y}_{[:t]})$; y por tanto

$$E[e_t(\ell)^2] = \text{Var}(e_t(\ell)) = \sigma^2(1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)$$

donde $1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2$ es la suma del cuadrado de los ℓ primeros coeficientes de la secuencia $\boldsymbol{\psi} = \boldsymbol{\phi}^{-1} * \boldsymbol{\theta}$, que es una secuencia de cuadrado sumable.

Y ahora consideremos modelos lineales, causales NO estacionarios.

- Suponga que \mathbf{Y} es SARIMA ($\boldsymbol{\phi}$ tiene raíces en el círculo unidad) pero que simultáneamente \mathbf{U} tiene un comienzo (que $U_t = 0$ para $t < 0$; por tanto $Y_t = \psi(B)U_n = 0$ para $t < 0$).
- El cálculo sigue siendo el mismo. Aunque ahora $\boldsymbol{\psi} = \boldsymbol{\phi}^{-\triangleright} * \boldsymbol{\theta}$ no es de cuadrado sumable, en las sumas solo hay una cantidad finita de términos no nulos.
- La incertidumbre en las predicciones difiere entre modelos estacionarios y no estacionarios. En los estacionarios, como $\psi_k \rightarrow 0$ cuando k aumenta, la varianza de la predicción tiene a un valor constante: la varianza marginal del proceso.

Por ejemplo, en un modelo AR(1), donde $\psi_k = \phi^k$, cuando ℓ tiende a infinito

$$\lim_{\ell \rightarrow \infty} \text{Var}(e_T(\ell)) = \sigma^2(1 + \psi_1^2 + \psi_2^2 + \cdots) = \sigma^2/(1 - \phi^2).$$

Si $|\phi|$ está próximo uno, la varianza marginal será mucho más grande que la varianza condicionada, σ^2 . Pero la incertidumbre es finita pues la varianza marginal también lo es.

- En modelos no estacionarios, como la serie $\sum \psi_i^2$ no converge, la incertidumbre en la predicción a largo plazo crece indefinidamente: **no se puede anticipar el comportamiento de un proceso no estacionario a largo plazo.**

4.1. Intervalos de confianza para las previsiones

Si asumimos que la distribución de las innovaciones es gaussiana, podemos calcular intervalos de confianza para diferentes horizontes de predicción.

Como $Y_{t+\ell}$ tendrá distribución normal con esperanza $\widehat{Y_{t+\ell|t}}$ y varianza

$$\sigma^2(1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2),$$

tendremos que

$$Y_{t+\ell} \in \left(\widehat{Y_{t+\ell|t}} \pm z_{1-\alpha} \sqrt{\text{Var}(e_t(\ell))} \right)$$

donde $z_{1-\alpha}$ son los valores críticos de la distribución normal.

En modelos MA(q) la varianza deja de crecer si $\ell > q$, pues $\psi_k = 0$ para todo $k > q$.

5. Adaptación de las predicciones a las nuevas observaciones

Contamos con predicciones para $t + 1$ hasta $t + j$ basadas en \mathcal{H}_{Y_t} . Queremos actualizarlas con $\mathcal{H}_{Y_{t+1}}$. Según (7), la predicción de $Y_{t+\ell}$ con información hasta t es:

$$\widehat{Y_{t+\ell|t}} = \psi_\ell U_t + \psi_{\ell+1} U_{t-1} + \dots$$

Comparando $\widehat{Y_{t+1|t}}$ con Y_{t+1} obtenemos el error de predicción $U_{t+1} = Y_{t+1} - \widehat{Y_{t+1|t}}$.

La nueva predicción para $Y_{t+\ell}$, incorporando Y_{t+1} como información adicional, será:

$$\widehat{Y_{t+\ell|t+1}} = \psi_{\ell-1} U_{t+1} + \psi_\ell U_t + \psi_{\ell+1} U_{t-1} + \dots$$

Restando las dos previsiones, tenemos que:

$$\widehat{Y_{t+\ell|t+1}} - \widehat{Y_{t+\ell|t}} = \psi_{\ell-1} U_{t+1}.$$

Al observar Y_{t+1} y obtener el error de previsión U_{t+1} , podremos revisar las previsiones:

$$\widehat{Y_{t+\ell|t+1}} = \widehat{Y_{t+\ell|t}} + \psi_{\ell-1} U_{t+1}$$

(las predicciones se revisan sumando un múltiplo del último error de predicción).

Si no hay error de previsión un periodo adelante ($U_{t+1} = 0$), no hay revisión del resto.

6. Medidas del error de previsión

Error Medio (ME) $= \frac{1}{T} \sum_{t=1}^T e_t$

Promedio de los errores de previsión.

Raíz del Error Cuadrático Medio (RMSE) $= \sqrt{\frac{1}{T} \sum_{t=1}^T e_t^2}$

Es la norma euclídea de vector de errores.

Error Absoluto Medio (MAE) $= \frac{1}{T} \sum_{t=1}^T |e_t|$

Promedio de los errores de previsión en términos absolutos

Error Porcentual Medio (MPE) $= \frac{1}{T} \sum_{t=1}^T 100 \frac{e_t}{y_t}$

Promedio de los errores de previsión porcentuales

Error Porcentual Absoluto Medio (MAPE) $= \frac{1}{T} \sum_{t=1}^T 100 \left| \frac{e_t}{y_t} \right|$

Promedio del valor absoluto de los errores porcentuales

Gretl muestra otros estadísticos adicionales (véase el manual).

7. Modelos con logaritmos

Si se tomaron logaritmos para estabilizar la varianza, es necesario recuperar las previsiones de la serie original a partir de las de la serie transformada. Si

$$Y_{t+\ell} = \ln(X_{t+\ell})$$

entonces la previsión puntual de la serie original es

$$\widehat{X_{t+\ell|t}} = \exp \left(\widehat{Y_{t+\ell|t}} + \frac{1}{2} \text{Var}(e_t(\ell)) \right)$$

Y el Intervalo de confianza

$$I_{1-\alpha}(X_{t+\ell}) = (a, b)$$

donde

$$\begin{aligned} a &= \exp \left(\widehat{Y_{t+\ell|t}} - (z_{1-\frac{\alpha}{2}}) \sqrt{\text{Var}(e_t(\ell))} \right) \\ b &= \exp \left(\widehat{Y_{t+\ell|t}} + (z_{1-\frac{\alpha}{2}}) \sqrt{\text{Var}(e_t(\ell))} \right) \end{aligned}$$

que es no simétrico alrededor de la previsión.

8. Nota sobre subespacio sobre el que proyectamos al prever

Sea un proceso estocástico lineal causal y estacionario de segundo orden definido por

$$Y_t = \sum_{k=0}^{\infty} \psi_k U_{t-k},$$

donde:

- $\sum_{k=0}^{\infty} \psi_k^2 < \infty$,
- U es un proceso de ruido blanco con $E[U_t] = 0$ y $E[U_t^2] = \sigma^2 < \infty$,
- el proceso es **invertible**, es decir, existe una secuencia $\{\pi_j\}_{j \geq 0}$ con $\sum_{j=0}^{\infty} \pi_j^2 < \infty$ tal que

$$U_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}.$$

Bajo las hipótesis anteriores se cumple que los subespacios cerrados de L^2 generados por el pasado de Y_t y por el pasado de las innovaciones U_t son iguales:

$$\mathcal{H}_{Y_t} = \mathcal{H}_{U_t}$$

donde $\mathcal{H}_{Z_n} = \overline{\text{sp}}\{1, Z_n, Z_{n-1}, \dots\}$ denota la clausura en la norma de L^2 del subespacio engendrado por las variables aleatorias $\{1, Z_n, Z_{n-1}, Z_{n-2}, \dots\}$. He incluido la constante 1 para poder considerar procesos con esperanza no nula. La “H” de la notación \mathcal{H}_{Z_n} tiene como objetivo recordar que se trata de un espacio de Hilbert y que dicho espacio está engendrado por las variables aleatorias que constituyen la “Historia” del proceso desde Z_n hacia atrás.

Demostración:

1. **Primera inclusión \subseteq :**

Cada Y_t se puede expresar como combinación lineal de las innovaciones pasadas:

$$Y_t = \sum_{k=0}^{\infty} \psi_k U_{t-k}.$$

Por tanto, todo Y_t pertenece al subespacio cerrado generado por $\{1, U_t, U_{t-1}, U_{t-2}, \dots\}$, y al incluir la clausura en L^2 obtenemos

$$\mathcal{H}_{Y_t} \subseteq \mathcal{H}_{U_t}$$

2. **Segunda inclusión \supseteq :**

Por invertibilidad, cada U_t se puede escribir como combinación de las Y pasadas:

$$U_t = \sum_{m=0}^{\infty} \pi_m Y_{t-m}.$$

Así, cada U_t pertenece al subespacio cerrado generado por $\{1, Y_t, Y_{t-1}, Y_{t-2}, \dots\}$, y al incluir la clausura en L^2 obtenemos la inclusión opuesta.

$$\mathcal{H}_{Y_t} \supseteq \mathcal{H}_{U_t}$$