

# Índice

|  |           |
|--|-----------|
| <b>1. Identificación y diagnosis</b>                                     | <b>2</b>  |
| 1.1. Instrumentos de identificación . . . . .                            | 3         |
| 1.2. Instrumentos de diagnosis . . . . .                                 | 4         |
| <b>2. Raíces unitarias</b>   | <b>4</b>  |
| 2.1. Notación: operadores retardo y diferencia y modelos ARIMA . . . . . | 4         |
| 2.1.1. Notación: ARIMA . . . . .   | 4         |
| 2.2. Raíces unitarias en los polinomios AR y MA . . . . .                | 5         |
| 2.3. Paseos aleatorios . . . . .   | 5         |
| <b>3. Modelos ARIMA estacionales (SARIMA)</b>                            | <b>6</b>  |
| 3.1. MA(1) estacional con raíz positiva . . . . .                        | 6         |
| 3.2. AR(1) estacional con raíz positiva . . . . .                        | 7         |
| 3.3. ARIMA $(0,0,1)x(0,0,1)$ . . . . .                                   | 9         |
| 3.4. ARIMA $(1,0,0)x(0,0,1)$ . . . . .                                   | 9         |
| 3.5. ARIMA $(1,0,0)x(1,0,0)$ . . . . .                                   | 10        |
| 3.6. ARIMA $(0,0,1)x(1,0,0)$ . . . . .                                   | 10        |
| <b>4. Resumen del análisis univariante de series temporales</b>          | <b>11</b> |
| 4.1. Ideas principales respecto a la modelización univariante . . . . .  | 11        |
| 4.2. Metodología . . . . .   | 12        |

# Lección 7. Modelos ARIMA y SARIMA. Identificación y diagnosis

Marcos Bujosa

1 de octubre de 2025

## Resumen

Repasaremos los instrumentos de identificación y diagnosis del análisis univariante. Extenderemos la notación para incorporar modelos con raíces unitarias y modelos estacionales. Finalmente resumiremos las ideas principales del análisis univariante.

- ([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.rcParams['text', usetex=False]
import matplotlib.pyplot as plt # data visualization
import datafram_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. Identificación y diagnosis

- Combinando herramientas gráficas y estadísticas, se puede inferir un modelo para los datos.
- Este proceso de especificación empírica de un modelo es conocido como "*identificación*".

El proceso de identificación puede estructurarse como una secuencia de preguntas:

1. ¿Es la serie estacionaria?
  2. ¿Tiene una media significativa?
  3. ¿Es persistente la ACF? ¿sigue alguna pauta reconocible?
  4. ¿Es persistente la PACF? ¿sigue alguna pauta reconocible?
- La identificación se apoya en estadísticos muestrales (media, autocorrelaciones, etc.) cuya representatividad respecto del proceso estocástico subyacente depende de la estacionariedad y la ergodicidad.
  - Tras inducir la estacionariedad, especificamos un modelo tentativo decidiendo cuál de las funciones ACF o PACF es finita y cuál es persistente

|                         | <b>ACF finita</b>                                      | <b>ACF persistente</b>                 |
|-------------------------|--|--|
| <b>PACF finita</b>      | Ruido blanco: retardos conjuntamente NO significativos | <u>AR</u> : orden indicado por la PACF |
| <b>PACF persistente</b> | <u>MA</u> : orden indicado por la ACF                  | <u>ARMA</u>                            |

La parametrización de mayor orden en modelos ARMA con series económicas suele ser ARMA(2, 1)

### 1.1. Instrumentos de identificación

|                                  | Instrumento  | Objetivo y observaciones   |
|----------------------------------|--|--|
| Transf.<br>logarítmica           | Gráficos rango-media y serie temporal.   | Conseguir independizar la variabilidad de los datos de su nivel. Las series económicas necesitan esta transformación frecuentemente. |
| $d$ , orden de<br>diferenciación | Gráfico de la serie temporal.<br>ACF (caída lenta y lineal).<br>Contrastes de raíz unitaria (DF o ADF y KPSS). | Conseguir que los datos fluctúen en torno a una media estable. En series económicas, $d$ suele ser 0, 1 ó 2.                         |
| Constante                        | Media de la serie transformada.<br>Desviación típica de la media.  | Si la media de la serie transformada es significativa, el modelo debe incluir un término constante.                                  |
| $p$ , orden AR.                  | Si PACF cae abruptamente en el retardo $p$ y la ACF decae lentamente.  | En series económicas $p$ suele ser $\leq 2$ .  |
| $q$ , orden MA.                  | Si ACF cae abruptamente en el retardo $q$ y PACF decae lentamente.   | En series económicas $q$ suele ser $\leq 1$ .  |

## 1.2. Instrumentos de diagnosis

|                               | Instrumento  | Possible diagnóstico  |
|-------------------------------|--|---|
| $d$ , orden de diferenciación | Proximidad a 1 de alguna raíz de los polinomios AR o MA. | Conviene diferenciar si la raíz es AR; o quitar una diferencia si es MA ( <i>salvo si hay tendencia determinista</i> ). |
| $d$ , orden de diferenciación | Gráfico de los residuos.                                 | Si muestra rachas largas de residuos positivos o negativos, puede ser necesaria una diferencia adicional.               |
| Constante                     | Media de los residuos.                                   | Si es significativa: añadir una constante.  |
| Constante                     | Constante estimada.                                      | Si NO es significativa: el modelo mejorará quitando el término constante.   |
| $p$ y $q$ ,                   | Contrastes de significación de los parámetros estimados. | Pueden sugerir eliminar parámetros irrelevantes.  |
| $p$ y $q$ ,                   | ACF/PACF residuos. Test Q de Ljung-Box.                  | Indican posibles pautas de autocorrelación no modelizadas.  |
| $p$ y $q$ ,                   | Correlaciones elevadas entre los parámetros estimados.   | Puede ser síntoma de sobreparametrización.  |

Una vez superadas las pruebas de diagnostico, aún se puede aplicar un análisis exploratorio; consistente en añadir parámetros AR y/o MA para comprobar si resultan significativos y mejoran el modelo.

## 2. Raíces unitarias

### 2.1. Notación: operadores retardo y diferencia y modelos ARIMA

El operador diferencia  $\nabla$  se define a partir del operador retardo como  $\nabla = (1 - \mathbf{B})$ :

$$\nabla Y_t = (1 - \mathbf{B})Y_t = Y_t - Y_{t-1}.$$

El operador diferencia estacional es  $\nabla_s = (1 - \mathbf{B}^S)$ :

$$\nabla_s Y_t = (1 - \mathbf{B}^S)Y_t = Y_t - Y_{t-S}.$$

#### 2.1.1. Notación: ARIMA

Extendemos la notación a procesos con raíces autorregresivas unitarias con “ARIMA( $p, d, q$ )”; donde  $d$  indica el número de diferencias que la serie necesita para ser  $I(0)$ ,

$$\phi_p * \nabla^d * \mathbf{Y} = \boldsymbol{\theta}_q * \mathbf{U};$$

es decir

$$\phi_p(\mathbf{B})\nabla^d Y_t = \boldsymbol{\theta}_q(\mathbf{B})U_t; \quad t \in \mathbb{Z}.$$

## 2.2. Raíces unitarias en los polinomios AR y MA

Cuando un polinomio tiene alguna raíz igual a uno se dice que tiene “raíces unitarias”.

Si el polinomio AR estimado tiene alguna raíz “próxima a uno” es síntoma de infradiferenciación.

Si el polinomio MA estimado tiene alguna raíz “próxima a uno” es síntoma de sobrediferenciación.

Ejemplos:

|   |   |
|---|---|
| Modelo expresado con raíces unitarias en<br>$\phi$ o $\theta$ | Modelo equivalente sin raíces unitarias en<br>$\phi$ o $\theta$ |
|---|---|

---

|   |  |
|---|--|
| $(1 - 1,5B + ,5B^2)Y_t = U_t$ $(1 - ,5B + 0,7B^2)\nabla^2 Y_t = (1 - B)U_t$ $\nabla Y_t = \beta + (1 - B)U_t$ | $(1 - 0,5B)\nabla Y_t = U_t$ $(1 - ,5B + 0,7B^2)\nabla Y_t = U_t$ $Y_t = \beta t + U_t \quad (\text{¡no estacionario!})$ |
|---|--|

## 2.3. Paseos aleatorios

Un paseo aleatorio representa una variable cuyos incrementos son ruido blanco:

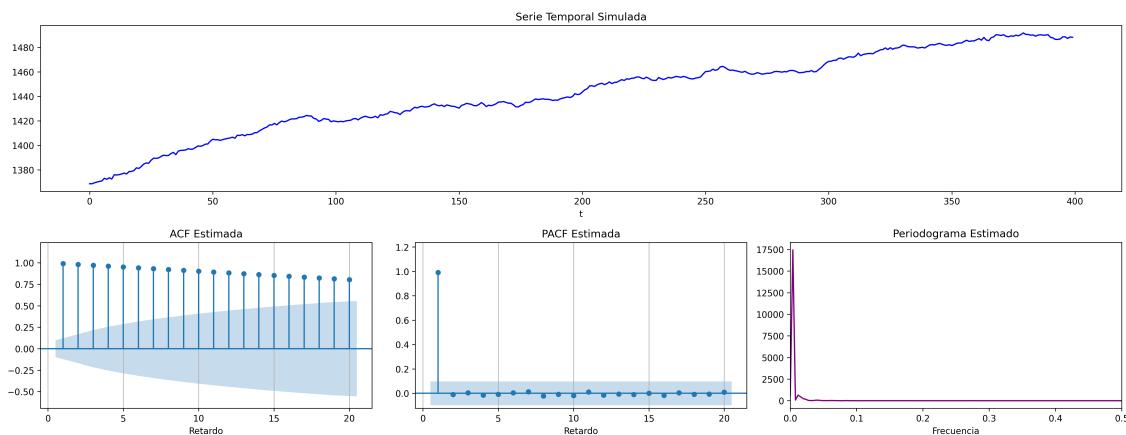
$$Y_t = \mu + Y_{t-1} + U_t.$$

Cuando  $\mu \neq 0$  se denomina *paseo aleatorio con deriva*:  $\nabla Y_t = \mu + U_t$ .

---

```
fig = plot_paseo_aleatorio_analysis(trend="t", pendiente=0.25, n=400, semilla=2025)
fig.savefig('./img/lecc07/ACF-RWcd.png', dpi=300, bbox_inches='tight')
```

---



El proceso tiene mayor inercia cuanto mayor es  $|\mu|$ . El signo de  $\mu$  determina el signo de la pendiente global.

Cuando  $\mu = 0$  se denomina sencillamente *paseo aleatorio*:  $\nabla Y_t = U_t$

---

```
fig = plot_paseo_aleatorio_analysis(n=400, semilla=2026)
fig.savefig('./img/lecc07/ACF-RW.png', dpi=300, bbox_inches='tight')
```

---



### 3. Modelos ARIMA estacionales (SARIMA)

El período estacional  $S$  es el número mínimo de observaciones necesarias para recorrer un ciclo estacional completo. Por ejemplo,  $S = 12$  para datos mensuales,  $S = 4$  para datos trimestrales,  $S = 24$  para datos horarios, etc.

Describiremos comportamientos estacionales con modelos ARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_S$

$$\phi_p(\mathbf{B})\Phi_P(\mathbf{B}^S)\nabla^d\nabla_S^D Y_t = \theta_q(\mathbf{B})\Theta_Q(\mathbf{B}^S)U_t; \quad t \in \mathbb{Z}$$

donde

$$\begin{aligned}\Phi_P(\mathbf{B}^S) &= 1 - \Phi_1\mathbf{B}^{1 \cdot S} - \Phi_2\mathbf{B}^{2 \cdot S} - \dots - \Phi_P\mathbf{B}^{P \cdot S} \\ \Theta_Q(\mathbf{B}^S) &= 1 - \Theta_1\mathbf{B}^{1 \cdot S} - \Theta_2\mathbf{B}^{2 \cdot S} - \dots - \Theta_Q\mathbf{B}^{Q \cdot S} \\ \nabla_S^D &= (1 - \mathbf{B}^S)^D\end{aligned}$$

Es decir, el modelo consta de polinomios autorregresivos y de media móvil tanto regulares (en minúsculas) como estacionales (en mayúsculas).

Veamos un ejemplo de un modelo MA(1) estacional y otro de un modelo AR(1) estacional...

#### 3.1. MA(1) estacional con raíz positiva

$$\text{MA}(1) \text{ estacional: } \Theta = 1 - 0.9z^{12} \Rightarrow X_t = (1 - 0.9\mathbf{B}^{12})U_t$$

---

```
SMA1 = [1, -0.9]
fig = plot_arma_seasonal_parametric_diagnostics(seasonal_ma_params=SMA1, s=12)
fig.savefig('img/lecc07/ACF-SMA1p.png', dpi=300, bbox_inches='tight')
```

---



|           | P. real   | P. imag.  | Módulo   | Frecuencia | Periodo | Multip. |
|-----------|-----------|-----------|----------|------------|---------|---------|
| <b>11</b> | 1.008819  | 0.000000  | 1.008819 | 0.000000   | Inf     | 1       |
| <b>9</b>  | 0.873663  | -0.504409 | 1.008819 | 0.083333   | 12.0    | 1       |
| <b>10</b> | 0.873663  | 0.504409  | 1.008819 | 0.083333   | 12.0    | 1       |
| <b>7</b>  | 0.504409  | -0.873663 | 1.008819 | 0.166667   | 6.0     | 1       |
| <b>8</b>  | 0.504409  | 0.873663  | 1.008819 | 0.166667   | 6.0     | 1       |
| <b>5</b>  | 0.000000  | -1.008819 | 1.008819 | 0.250000   | 4.0     | 1       |
| <b>6</b>  | 0.000000  | 1.008819  | 1.008819 | 0.250000   | 4.0     | 1       |
| <b>3</b>  | -0.504409 | -0.873663 | 1.008819 | 0.333333   | 3.0     | 1       |
| <b>4</b>  | -0.504409 | 0.873663  | 1.008819 | 0.333333   | 3.0     | 1       |
| <b>1</b>  | -0.873663 | -0.504409 | 1.008819 | 0.416667   | 2.4     | 1       |
| <b>2</b>  | -0.873663 | 0.504409  | 1.008819 | 0.416667   | 2.4     | 1       |
| <b>0</b>  | -1.008819 | 0.000000  | 1.008819 | 0.500000   | 2.0     | 1       |

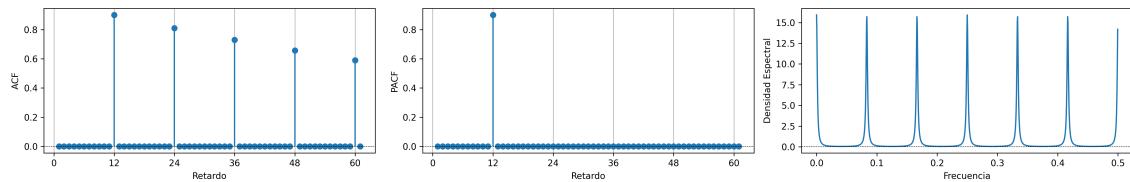
```
fig = plot_sarima_analysis(seasonal_ma_params=SMA1, s=12, seed=2025)
fig.savefig('./img/lecc07/Sim-SMA1p.png', dpi=300, bbox_inches='tight')
```



### 3.2. AR(1) estacional con raíz positiva

$$\text{AR}(1) \text{ estacional: } \Phi = 1 - 0.9z^{12} \Rightarrow (1 - 0.9B^{12})X_t = U_t$$

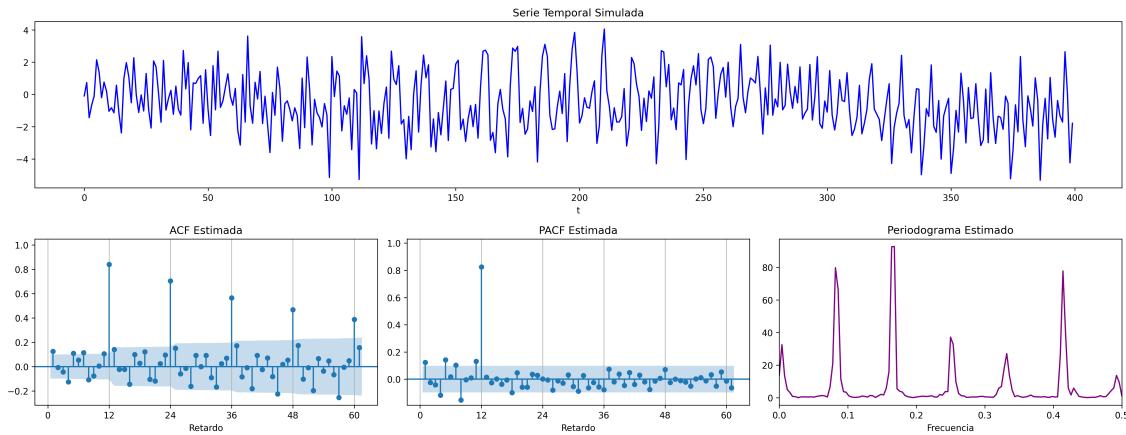
```
SAR1 = [1, -0.9]
fig = plot_arma_seasonal_parametric_diagnostics(seasonal_ar_params=SAR1, s=12)
fig.savefig('./img/lecc07/ACF-SAR1p.png', dpi=300, bbox_inches='tight')
```



|           | P. real   | P. imag.  | Módulo   | Frecuencia | Periodo | Multip. |
|-----------|-----------|-----------|----------|------------|---------|---------|
| <b>11</b> | 1.008819  | 0.000000  | 1.008819 | 0.000000   | Inf     | 1       |
| <b>9</b>  | 0.873663  | -0.504409 | 1.008819 | 0.083333   | 12.0    | 1       |
| <b>10</b> | 0.873663  | 0.504409  | 1.008819 | 0.083333   | 12.0    | 1       |
| <b>7</b>  | 0.504409  | -0.873663 | 1.008819 | 0.166667   | 6.0     | 1       |
| <b>8</b>  | 0.504409  | 0.873663  | 1.008819 | 0.166667   | 6.0     | 1       |
| <b>5</b>  | 0.000000  | -1.008819 | 1.008819 | 0.250000   | 4.0     | 1       |
| <b>6</b>  | 0.000000  | 1.008819  | 1.008819 | 0.250000   | 4.0     | 1       |
| <b>3</b>  | -0.504409 | -0.873663 | 1.008819 | 0.333333   | 3.0     | 1       |
| <b>4</b>  | -0.504409 | 0.873663  | 1.008819 | 0.333333   | 3.0     | 1       |
| <b>1</b>  | -0.873663 | -0.504409 | 1.008819 | 0.416667   | 2.4     | 1       |
| <b>2</b>  | -0.873663 | 0.504409  | 1.008819 | 0.416667   | 2.4     | 1       |
| <b>0</b>  | -1.008819 | 0.000000  | 1.008819 | 0.500000   | 2.0     | 1       |

Evidentemente las raíces son iguales a las del caso anterior (aunque ahora corresponden al polinomio autorregresivo).

```
fig = plot_sarima_analysis(seasonal_ar_params=SAR1, s=12, seed=2025)
fig.savefig('./img/lecc07/Sim-SAR1p.png', dpi=300, bbox_inches='tight')
```



Con estos dos ejemplos hemos podido apreciar que:

- las pautas de autocorrelación son análogas a las de los MA(1) y AR(2), pero ahora los retardos significativos corresponden a los retardos estacionales, es decir, a múltiplos del período estacional  $S$ .
- En estos ejemplos, en los que  $S = 12$ , los retardos estacionales son: 12, 24, 36, 48, 60,...
- las correlaciones correspondientes a los “retardos regulares” (es decir, todos menos los estacionales) son no significativas en general.

Veamos ahora un par de ejemplos de modelos estacionales multiplicativos (i.e., con parte regular y parte estacional).

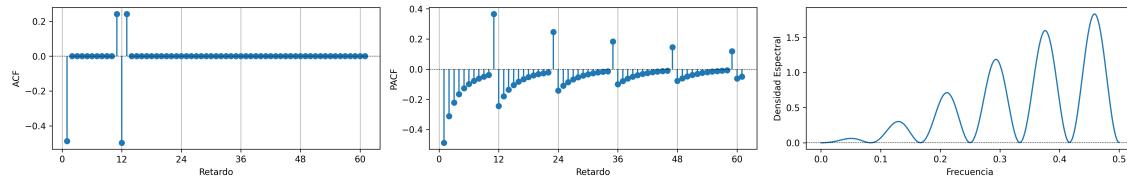
### 3.3. ARIMA (0,0,1)x(0,0,1)

$$\text{ARIMA}(0, 0, 1) \times (0, 0, 1)_{12}: \quad X_t = (1 - 0.9B)(1 - 0.9B^{12})U_t$$

---

```
MA1 = [1, -0.8]
SMA1 = [1, -0.9]
fig = plot_arma_seasonal_parametric_diagnostics(ma_params=MA1, seasonal_ma_params=SMA1, s=12)
fig.savefig('./img/lecc07/ACF-MA1SMA1.png', dpi=300, bbox_inches='tight')
```

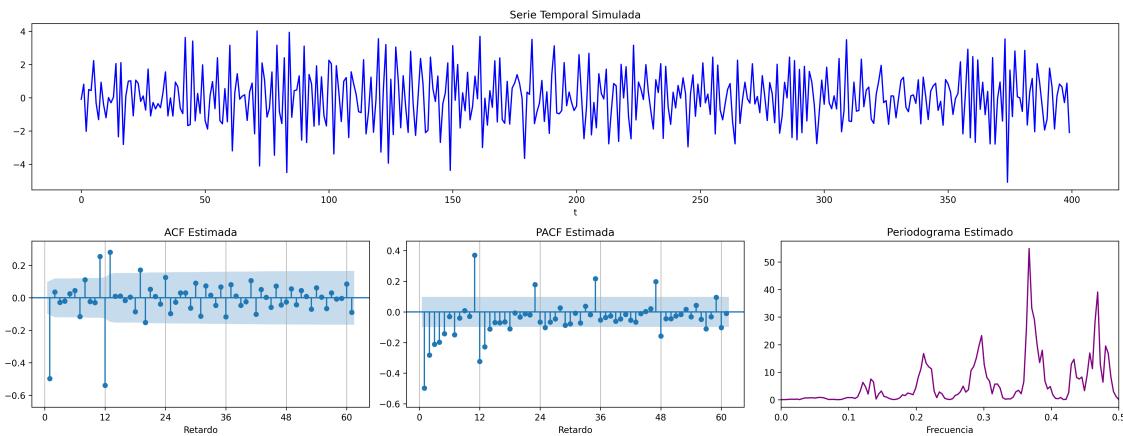
---




---

```
fig = plot_sarima_analysis(ma_params=MA1, seasonal_ma_params=SMA1, s=12, seed=2025)
fig.savefig('./img/lecc07/Sim-MA1SMA1.png', dpi=300, bbox_inches='tight')
```

---



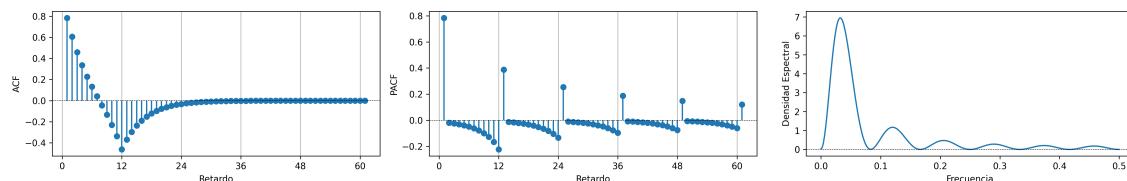
### 3.4. ARIMA (1,0,0)x(0,0,1)

$$\text{ARIMA}(1, 0, 0) \times (0, 0, 1)_{12}: \quad (1 - 0.9B)X_t = (1 - 0.9B^{12})U_t$$

---

```
AR1 = [1, -0.8]
SMA1 = [1, -0.9]
fig = plot_arma_seasonal_parametric_diagnostics(ar_params=AR1, seasonal_ma_params=SMA1, s=12)
fig.savefig('./img/lecc07/ACF-AR1SMA1.png', dpi=300, bbox_inches='tight')
```

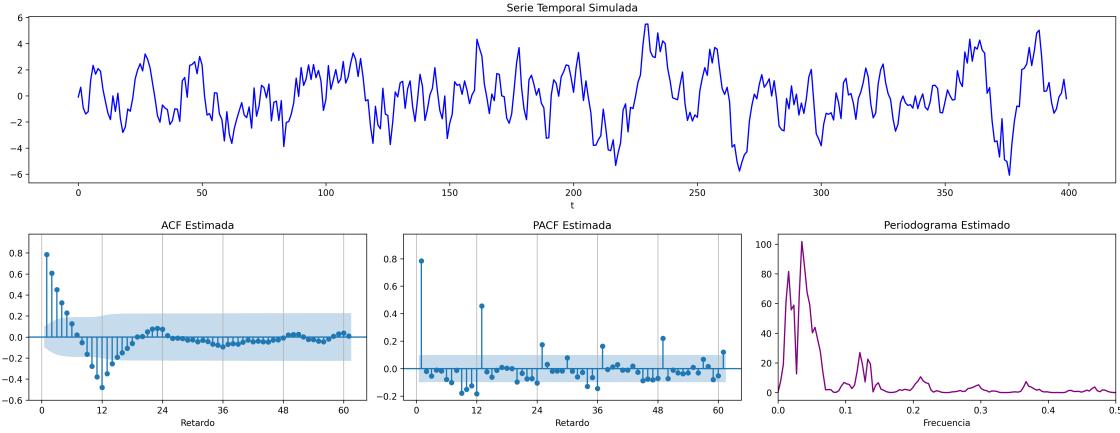
---




---

```
fig = plot_sarima_analysis(ar_params=AR1, seasonal_ma_params=SMA1, s=12, seed=2025)
fig.savefig('./img/lecc07/Sim-AR1SMA1.png', dpi=300, bbox_inches='tight')
```

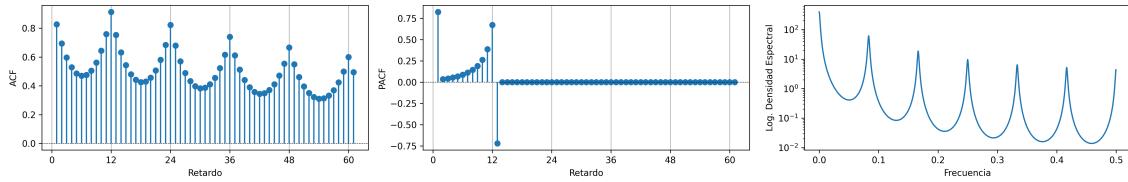
---



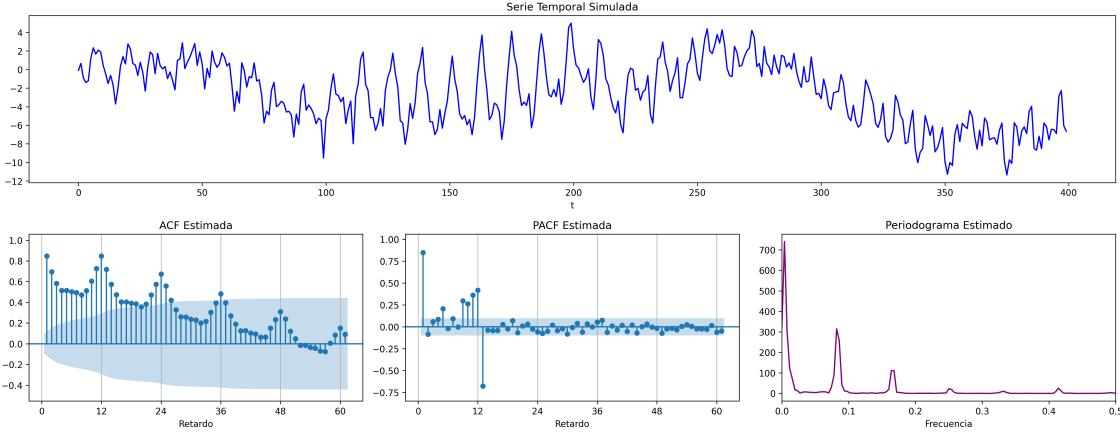
### 3.5. ARIMA (1,0,0)x(1,0,0)

$$\text{ARIMA}(1,0,0) \times (1,0,0)_{12}: \quad (1 - 0.9B)(1 - 0.9B^{12})X_t = U_t$$

```
AR1 = [1, -0.8]
SAR1 = [1, -0.9]
fig = plot_arima_seasonal_parametric_diagnostics(ar_params=AR1, seasonal_ar_params=SAR1, s=12, logs=True)
fig.savefig('./img/lecc07/ACF-AR1SAR1.png', dpi=300, bbox_inches='tight')
```



```
fig = plot_sarima_analysis(ar_params=AR1, seasonal_ar_params=SAR1, s=12, seed=2025)
fig.savefig('./img/lecc07/Sim-AR1SAR1.png', dpi=300, bbox_inches='tight')
```



### 3.6. ARIMA (0,0,1)x(1,0,0)

$$\text{ARIMA}(0,0,1) \times (1,0,0)_{12}: \quad (1 - 0.9B^{12})X_t = (1 - 0.9B)U_t$$

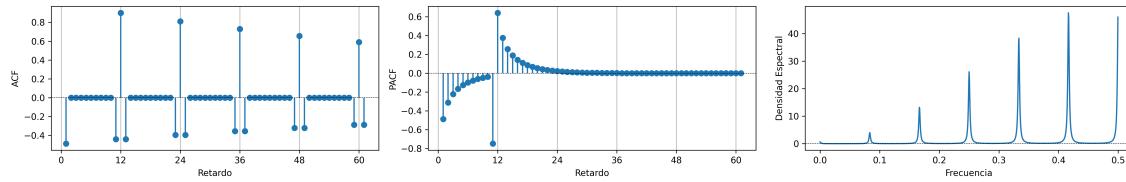
---

```

MA1 = [1, -0.8]
SAR1 = [1, -0.9]
fig = plot_arima_seasonal_parametric_diagnostics(ma_params=MA1, seasonal_ar_params=SAR1, s=12)
fig.savefig('./img/lecc07/ACF-MA1SAR1.png', dpi=300, bbox_inches='tight')

```

---



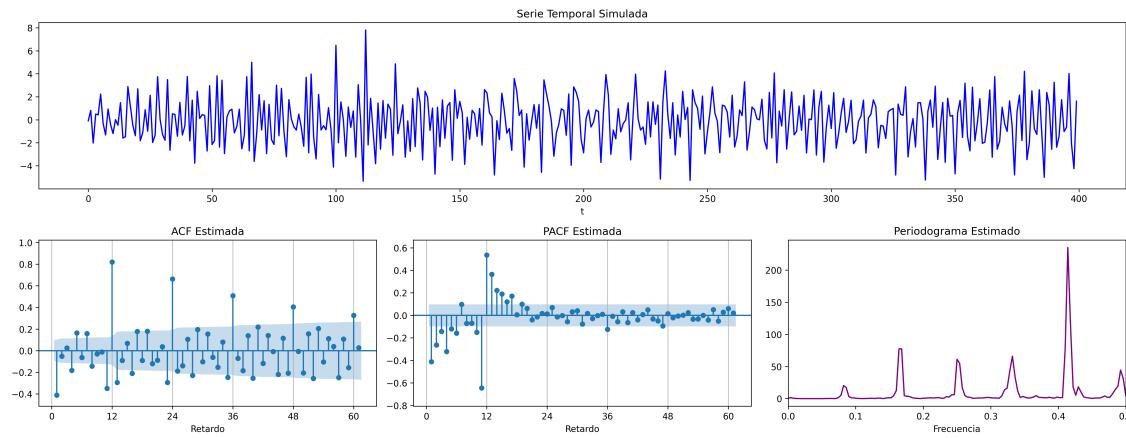

---

```

fig = plot_sarima_analysis(ma_params=MA1, seasonal_ar_params=SAR1, s=12, seed=2025)
fig.savefig('./img/lecc07/Sim-MA1SAR1.png', dpi=300, bbox_inches='tight')

```

---



En estos cuatro ejemplos hemos podido apreciar que

- en el entorno de los retardos estacionales surgen una serie de coeficientes significativos (“satélites”) que proceden de la interacción entre las estructuras regular y estacional
- Estos satélites son útiles para identificar en qué retardos estacionales hay autocorrelaciones no nulas, pero no requieren una parametrización especial.

## 4. Resumen del análisis univariante de series temporales

### 4.1. Ideas principales respecto a la modelización univariante

- Es una modelización sin variables exógenas
- Modelizan la interdependencia temporal con polinomios de órdenes reducidos.
- Está especialmente indicada para hacer predicción.
- Parte de dos supuestos sobre el proceso estocástico subyacente:
  1. es débilmente estacionario

2. tiene representación como proceso lineal:  $Y_t = \mu + \sum_{j=0}^{\infty} a_j U_{t-j}$ ; con  $\mu \in \mathbb{R}$ ,  $a \in \ell^2$  y  $U \sim WN(0, \sigma^2)$

- Utiliza múltiples instrumentos: (a) gráficos (b) función de autocorrelación (c) función de autocorrelación parcial, (d) estadístico Q de Ljung-Box, etc...
- Si la serie original no "parece" débilmente estacionaria, se induce esta propiedad mediante las transformaciones adecuadas

|                         | <b>ACF finita</b>                                      | <b>ACF persistente</b>                 |
|-------------------------|--|--|
| <b>PACF finita</b>      | Ruido blanco: retardos conjuntamente NO significativos | <u>AR</u> : orden indicado por la PACF |
| <b>PACF persistente</b> | <u>MA</u> : orden indicado por la ACF                  | <u>ARMA</u>                            |

## 4.2. Metodología

Tres fases:

**Identificación** Se elige una especificación provisional para el proceso estocástico subyacente en función de las características medibles de los datos ("dejar que los datos hablen")

**Estimación** suele requerir métodos iterativos (*Gretl se encarga de esto*)

**Diagnóstico** de la calidad estadística del modelo ajustado. Algunos controles estándar son:

- Significatividad de los parámetros estimados
- Estacionariedad y homocedasticidad de los residuos
- ¿Existe un patrón de autocorrelación residual que podría ser modelado? ¿O hemos logrado que los residuos sean **ruido blanco**?

Si la diagnosis no es satisfactoria, se vuelve a la primera fase.

Si la diagnosis es satisfactoria... ¡hemos logrado un modelo aceptable!... que podremos usar para realizar pronósticos.