Modelos AutoRegresivos de Media Móvil (ARMA)

Randall Romero Aguilar, PhD

randall.romero@ucr.ac.cr

EC4301 - Macroeconometría

I Semestre 2020

Última actualización: 19 de abril de 2020

UCR
UNIVERSIDAD DE COSTA RICA

ESCUELA de ECONOMÍA
UNIVERSIDAD DE COSTA RICA

Tabla de contenidos

1. Introducción

2. Proceso de media móvil MA(q)

3. Proceso autorregresivo AR(p)

4. Proceso autorregresivo de media móvil ARMA(p,q)

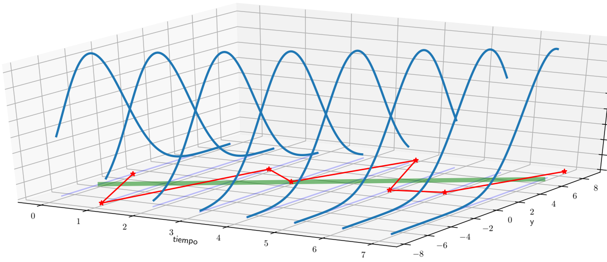
5. Estimación de modelos ARMA

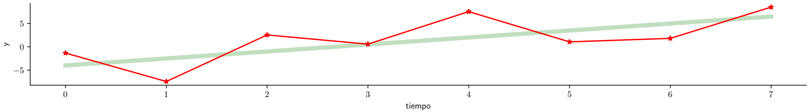
6. Pronósticos con modelos ARMA

1. Introducción

¿Qué es una serie de tiempo?

Una serie de tiempo $\{y_t\}_{t=1}^{T}$ es una realización de un proceso estocástico.





¿Qué nos gustaría hacer con esta serie de tiempo?

- Imaginemos a la serie de tiempo como la parte de un proceso estocástico para la cual ya tenemos las realizaciones del proceso (un valor por período)

$$\underbrace{y_1, y_2, \dots, y_T}_{\text{datos observados}}, \underbrace{y_{T+1}, y_{T+2}, \dots}_{\text{futuro}}$$

- Nos gustaría saber la distribución condicional

$$\mathbb{P}\left[y_{T+j} \leq x \mid y_1, \dots, y_T\right]$$

- Esto nos permitiría utilizar nuestra serie de tiempo para pronosticar valores futuros de la serie, así como precisar su variabilidad:

$$\mathbb{E}_T\left[y_{T+j}\right] \equiv \mathbb{E}\left[y_{T+j} \mid y_1 \dots, y_T\right]$$
$$\operatorname{Var}_T\left[y_{T+j}\right] \equiv \operatorname{Var}\left[y_{T+j} \mid y_1, \dots, y_T\right]$$

Nuestra tarea

- En la práctica, puede ser que nunca conozcamos el verdadero proceso generador de datos (PGD) (el proceso estocástico del cual fueron obtenidos los valores de nuestra serie de tiempo).
- Nuestra tarea es desarrollar modelos que capturen la esencia del verdadero PGD.
- Las ecuaciones en diferencia estocásticas son una manera muy conveniente de modelar procesos económicos dinámicos.

Ejemplo 1:

Controlando la oferta de dinero

- Suponga que la meta de oferta monetaria M^* del banco central crece 100g% por año:

$$M_t^* = (1+g)M_{t-1}^*$$

o en términos logarítmicos, con $m^* \equiv \log(M^*)$

$$m_t^* = \log(1+g) + m_{t-1}^*$$
$$m_t^* \approx g + m_{t-1}^*$$

- Para una condición inicial m_0^* dada, la solución es:

$$m_t^* = gt + m_0^*$$

- La cantidad efectiva de dinero m_t puede diferir de la meta m_t^*.
- El banco central intenta cerrar una proporción ρ de la brecha entre la meta y la cantidad efectiva del período anterior. El cambio en la oferta de dinero es:

$$\Delta m_t = \underbrace{\rho\left(m_{t-1}^* - m_{t-1}\right)}_{\text{política monetaria}} + \underbrace{\epsilon_t}_{\text{perturbación}}$$

- por lo que la oferta de dinero es

$$\begin{aligned} m_t &= \rho m_{t-1}^* + (1-\rho)m_{t-1} + \epsilon_t \\ &= \rho g(t-1) + \rho m_0^* + (1-\rho)m_{t-1} + \epsilon_t \\ &= \underbrace{\rho(m_0^* - g)}_{\text{intercepto}} + \underbrace{(\rho g)t}_{\text{tendencia}} + \underbrace{(1-\rho)m_{t-1}}_{\text{autorregresivo}} + \underbrace{\epsilon_t}_{\text{shock}} \end{aligned}$$

$$m_t = \underbrace{\rho(m_0^* - g)}_{\text{intercepto}} + \underbrace{(\rho g)t}_{\text{tendencia}} + \underbrace{(1-\rho)m_{t-1}}_{\text{autorregresivo}} + \underbrace{\epsilon_t}_{\text{shock}}$$

- Aunque la oferta monetaria es una variable continua, nuestro modelo es una ecuación en diferencia (discreta).
- Como las perturbaciones $\{\epsilon_t\}$ son aleatorias, la oferta de dinero es estocástica.
- Si supiéramos la distribución de $\{\epsilon_t\}$, podríamos calcular la distribución de cada elemento de $\{m_t\}$, porque está determinada completamente por los parámetros de la ecuación y por la secuencia $\{\epsilon_t\}$.
- Habiendo observado las primeras T observaciones de la serie $\{m_t\}$, podríamos pronosticar futuros valores. Por ejemplo:

$$\begin{aligned} m_{T+1} &= \rho\left(gT + m_0^*\right) + (1-\rho)m_T + \epsilon_{T+1} \\ \Rightarrow \mathbb{E}_T\left[m_{T+1}\right] &= \rho\left(gT + m_0^*\right) + (1-\rho)m_T \end{aligned}$$

Ruido blanco y el modelo clásico de regresión lineal

- En el MCRL se asume que

$$y_t = x_t'\beta + \epsilon_t$$

donde para todas las observaciones $t = 1, 2, \ldots, T$ el término de error cumple que:

$$\begin{aligned} \mathbb{E}[\epsilon_t] &= 0 && \text{(media cero)} \\ \mathrm{Var}[\epsilon_t] &= \sigma^2 && \text{(no hay heteroscedasticidad)} \\ \mathbb{E}[\epsilon_t \epsilon_\tau] &= 0 \quad \text{si } t \neq \tau && \text{(no hay autocorrelación)} \end{aligned}$$

- Es decir, el MCRL aplicado a series de tiempo asume que el error es un proceso de ruido blanco.
- Sin embargo, en la práctica rara vez se satisface ese supuesto cuando se ajusta un modelo de regresión lineal a datos de series de tiempo.

Ejemplo 2:

Estimando la demanda de dinero

- bccr
- demandadinero.ipynb

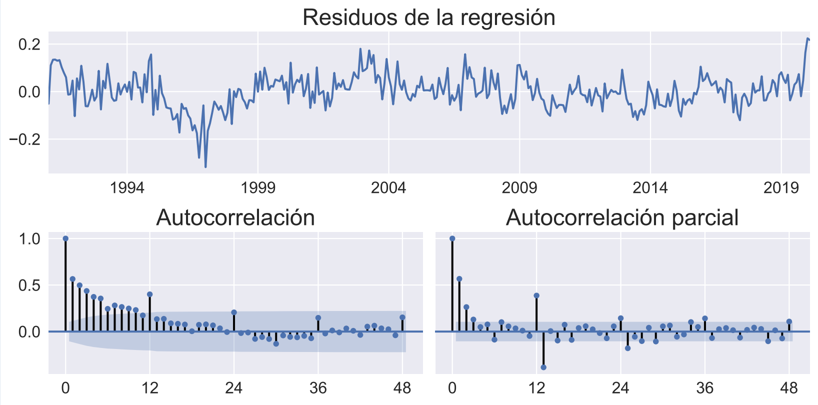
Consideremos la demanda de dinero en Costa Rica

$$\log(M_t) = \beta_0 + \beta_1 \log(q_t) + \beta_2 \log(p_t) + \beta_3 \log(i_t) + \epsilon_t$$

la cual estimamos con datos mensuales (1991m01 a 2020m01) del medio circulante M_t, el IMAE q_t, el IPC p_t, y la tasa básica pasiva i_t.

	coef	**std err**	**t**	**P$>$ |t|**	**[0.025**	**0.975]**
Intercept	5.5080	0.227	24.251	0.000	5.061	5.955
IMAE	1.1300	0.059	19.143	0.000	1.014	1.246
IPC	0.9463	0.021	45.664	0.000	0.905	0.987
Tbasica	-0.2409	0.018	-13.641	0.000	-0.276	-0.206

No obstante, los residuos de la regresión no parecen ruido blanco.



- Conocer el valor de un residuo puede ayudar a pronosticar el siguiente.
- Aún así, notemos que para pronosticar el valor del M1 en 2020m02, necesitaríamos pronosticar los valores de las demás variables.

En esta clase

- En esta clase aprenderemos a modelar series de tiempo en función de:
  - sus valores rezagados (procesos autorregresivos)
  - valores rezagados de un ruido blanco (procesos de media móvil)
- Primero estudiamos las propiedades teóricas de procesos estocásticos.
- Luego tratamos de identificar el PGD de nuestra serie a partir de sus estadísticos muestrales, comparándolos con los estadísticos de los procesos del punto anterior.
- Finalmente, utilizamos nuestro modelo estimado para
  - análisis de escenarios: ¿qué pasaría con la serie de tiempo si recibe una perturbación estocástica de cierta magnitud (función impulso respuesta)
  - pronósticos: ¿qué valores esperamos ver en el futuro para esta serie de tiempo?

Modelos que estudiaremos

Ruido blanco

Es una secuencia $\{\epsilon_t\}$ cuyos elementos satisfacen,

$$\mathbb{E}\left(\epsilon_t\right) = 0$$
$$\mathbb{E}\left(\epsilon_t^2\right) = \sigma^2$$
$$\mathbb{E}\left(\epsilon_t \epsilon_\tau\right) = 0 \quad \text{for } t \neq \tau$$

Proceso media móvil

Sea $\{\epsilon_t\}$ ruido blanco; el proceso estocástico

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

con $\theta_q \neq 0$ es llamado un proceso MA(q).

Proceso autorregresivo

Sea $\{\epsilon_t\}$ ruido blanco; el proceso estocástico

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

con $\phi_p \neq 0$ es llamado un proceso AR(p).

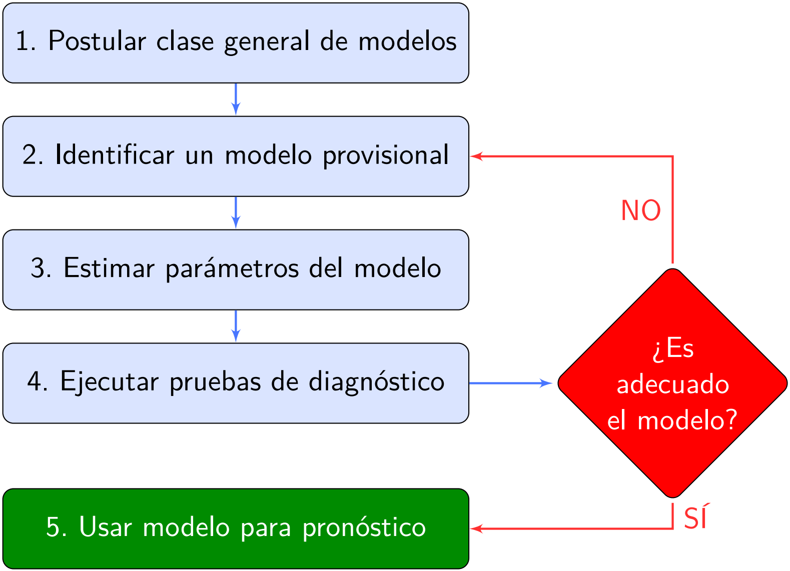
Autorregresivo media móvil

Sea $\{\epsilon_t\}$ ruido blanco; el proceso estocástico

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \dots$$
$$+ \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

es llamado proceso ARMA(p,q).

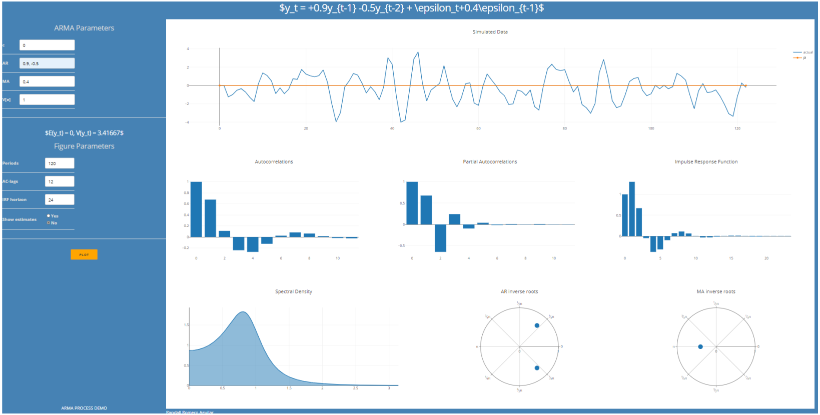
La metodología Box-Jenkins



Acerca de los ejemplos

- En esta clase veremos ilustraciones de distintos proceso AR, MA, y ARMA.
- Usted puede reproducirlas (y estudiar más casos específicos de estos procesos) con el paquete macrodemos que escribí en Python para este tema.
- Para instalarlo: `pip install macrodemos`
- Para ejecutarlo:

```
from macrodemos import ARMA_demo
ARMA_demo()
```



$y_t = +0.9y_{t-1} -0.5y_{t-2} + \epsilon_t+0.4\epsilon_{t-1}$
ARMA Parameters
c
0
AR
0.9, -0.5
MA
0.4
V[e]
1
$E(y_t) = 0, V(y_t) = 3.41667$
Figure Parameters
Periods
120
AC-lags
12
IRF horizon
24
Show estimates
Yes
No
PLOT
ARMA PROCESS DEMO
Simulated Data
actual
Autocorrelations
Partial Autocorrelations
Impulse Response Function
Spectral Density
AR inverse roots
MA inverse roots
Randall Romero Aguilar

2. Proceso de media móvil MA(q)

Proceso media móvil de primer orden: MA(1)

- Sea $\{\epsilon_t\}_{t=-\infty}^{\infty}$ un proceso ruido blanco.
- Se define el proceso MA(1) como:

$$\begin{aligned} y_t &= \mu + \epsilon_t + \theta\epsilon_{t-1} \\ &= \mu + (1 + \theta\,\mathrm{L})\epsilon_t \end{aligned}$$

- Su valor esperado es:

$$\begin{aligned} \mathbb{E}[y_t] &= \mu + \mathbb{E}[\epsilon_t] + \theta\,\mathbb{E}[\epsilon_{t-1}] \\ &= \mu + 0 + 0\theta \\ &= \mu \end{aligned}$$

- Su varianza es

$$
\begin{aligned}
\operatorname{Var}[y_t] &\equiv \mathbb{E}\left[(y_t - \mathbb{E}[y_t])^2\right] \\
&= \mathbb{E}\left[(y_t - \mu)^2\right] \\
&= \mathbb{E}\left[(\epsilon_t + \theta\epsilon_{t-1})^2\right] \\
&= \mathbb{E}\left[\epsilon_t^2 + 2\theta\epsilon_t\epsilon_{t-1} + \theta^2\epsilon_{t-1}^2\right] \\
&= \mathbb{E}\left[\epsilon_t^2\right] + 2\theta\,\mathbb{E}\left[\epsilon_t\epsilon_{t-1}\right] + \theta^2\,\mathbb{E}\left[\epsilon_{t-1}^2\right] \\
&= \sigma^2 + 2\theta \times 0\theta^2\sigma^2 \\
&= (1+\theta^2)\sigma^2
\end{aligned}
$$

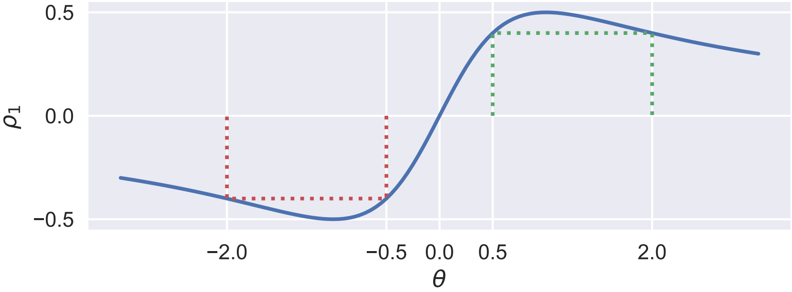
- Su autocovarianza j, para $j \geq 1$, es

$$\begin{aligned}
\operatorname{Cov}[y_t, y_{t-j}] &= \mathbb{E}\left[(y_t - \mu)(y_{t-j} - \mu)\right] \\
&= \mathbb{E}\left[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-j} + \theta\epsilon_{t-j-1})\right] \\
&= \mathbb{E}\left[\epsilon_t\epsilon_{t-j} + \theta\epsilon_t\epsilon_{t-j-1} + \theta\epsilon_{t-1}\epsilon_{t-j} + \theta^2\epsilon_{t-1}\epsilon_{t-j-1}\right] \\
&= 0 + \theta \times 0 + \theta\,\mathbb{E}\left[\epsilon_{t-1}\epsilon_{t-j}\right] + \theta^2 \times 0 \\
&= \theta\,\mathbb{E}\left[\epsilon_{t-1}\epsilon_{t-j}\right] = \begin{cases} \theta\sigma^2 & , \text{ si } j = 1 \\ 0 & , \text{ si } j > 1 \end{cases}
\end{aligned}$$

- Por lo tanto, su función de autocorrelación es

$$\rho_j = \begin{cases} \frac{\theta}{1+\theta^2} & \text{, si } j = 1 \\ 0 & \text{, si } j > 1 \end{cases}$$

- Notemos que la función de autocorrelación de $y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$ es la misma que para el proceso $z_t = \mu + \epsilon_t + \frac{1}{\theta}\epsilon_{t-1}$



- Resumiendo los resultados que hemos obtenido

$$\begin{aligned}\mathbb{E}[y_t] &= \mu \\ \mathrm{Var}[y_t] &= (1+\theta^2)\sigma^2 \\ \mathrm{Cov}[y_t, y_{t-j}] &= \begin{cases}\theta\sigma^2 & , \text{ si } j = 1 \\ 0 & , \text{ si } j > 1\end{cases}\end{aligned}$$

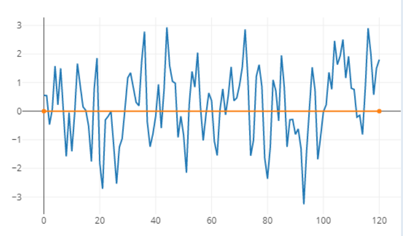
vemos que ninguno de estos momentos depende del tiempo t, por lo que el proceso MA(1) siempre es covarianza-estacionario.

Ejemplo 3:

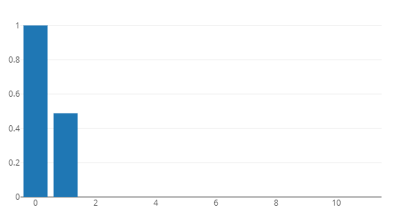
MA(1)

$$y_t = \epsilon_t + 0.8\epsilon_{t-1}$$

Realización

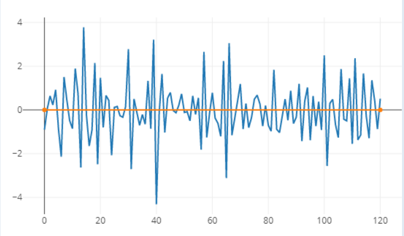


Autocorrelación

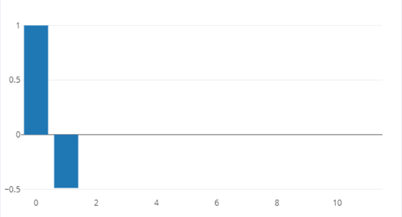


$$y_t = \epsilon_t - 0.8\epsilon_{t-1}$$

Realización



Autocorrelación



Invertibilidad de un proceso MA(1)

- Supongamos que $\mu = 0$, con lo que el proceso es

$$y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta\,\mathrm{L})\epsilon_t$$

- Siempre que $|\theta| < 1$ podemos invertir el polinomio $(1 + \theta\,\mathrm{L})$:

$$(1 + \theta\,\mathrm{L})^{-1} = (1 - \theta\,\mathrm{L} + \theta^2\,\mathrm{L}^2 - \theta^3\,\mathrm{L}^3 + \dots)$$

- con lo que

$$\begin{aligned}
\epsilon_t &= (1 + \theta\,\mathrm{L})^{-1} y_t \\
&= (1 - \theta\,\mathrm{L} + \theta^2\,\mathrm{L}^2 - \theta^3\,\mathrm{L}^3 + \dots) y_t \\
&= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 y_{t-3} + \dots \\
\Rightarrow y_t &= \epsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - \dots
\end{aligned}$$

- Es decir, podemos representar el proceso MA(1) como un proceso AR(∞).

- Recordemos que si bien un proceso MA(1) con parámetro θ tiene exactamente la misma función de autocorrelación que un proceso con parámetro $\frac{1}{\theta}$, solo uno de ellos puede ser invertible, porque si $|\theta| < 1$, entonces $|\frac{1}{\theta}| > 1$.
- Para ciertos métodos de estimación, sólo será posible estimar el modelo MA(1) si es invertible.
- Por ello, para modelos no invertibles se suele cambiar el parámetro por su recíproco.

El proceso MA(q)

- Es fácil extender el proceso MA(1) para incluir más rezagos.
- El proceso MA(q) es

$$
\begin{aligned}
y_t &= \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\
&= \mu + \left(1 + \theta_1 \, \mathrm{L} + \cdots + \theta_q \, \mathrm{L}^q\right) \epsilon_t \\
&= \mu + \Theta(\mathrm{L}) \epsilon_t
\end{aligned}
$$

  con ϵ_t ruido blanco.
- Su valor esperado es

$$
\begin{aligned}
\mathbb{E}\left(y_t\right) &= \mathbb{E}\left(\mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}\right) \\
&= \mu + \mathbb{E}\left(\epsilon_t\right) + \theta_1 \, \mathbb{E}\left(\epsilon_{t-1}\right) + \cdots + \theta_q \, \mathbb{E}\left(\epsilon_{t-q}\right) \\
&= \mu
\end{aligned}
$$

- Su varianza es

$$\begin{aligned}\mathrm{Var}[y_t] &= \mathbb{E}\left[(y_t-\mu)^2\right] \\ &= \mathbb{E}\left[(\epsilon_t+\theta_1\epsilon_{t-1}+\cdots+\theta_q\epsilon_{t-q})^2\right] \\ &= \sigma^2+\theta_1^2\sigma^2+\cdots+\theta_q^2\sigma^2 \\ &= (1+\theta_1^2+\cdots+\theta_q^2)\sigma^2\end{aligned}$$

- mientras que su autocovarianza es

$$\gamma_j = \begin{cases}\left(\theta_j+\theta_{j+1}\theta_1+\theta_{j+2}\theta_2+\cdots+\theta_q\theta_{q-j}\right)\sigma^2 & \text{si } j=1,2,\ldots,q \\ 0 & \text{para } j>q.\end{cases}$$

- es decir, una característica distintiva de un proceso MA(q) es que todas sus autocorrelaciones para rezagos mayores a q son cero.

Invertibilidad de un proceso MA(q)

- El proceso MA(q)
$$y_t = \mu + \Theta(\mathrm{L})\epsilon_t$$
será invertible si y solo si las raíces del polinomio $\Theta(z)$ están todas fuera del círculo unitario.
- En ese caso, el proceso se puede representar por
$$\epsilon_t = \Theta^{-1}(\mathrm{L})(y_t - \mu)$$
lo cual corresponde a un proceso AR(∞).

Función impulso respuesta de un proceso MA(q)

- La función de impulso respuesta está definida por

$$\Psi(j) = \frac{\partial y_{t+j}}{\partial \epsilon_t}$$

es decir, nos dice cuánto cambia y luego de j períodos ante una perturbación.

- Para series estacionarias, podemos escribir

$$\Psi(j) = \frac{\partial y_t}{\partial \epsilon_{t-j}}$$

- Pero como $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$ es fácil ver que

$$\Psi(j) = \theta_j$$

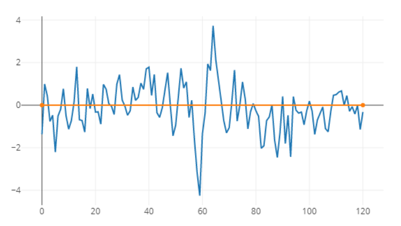
es decir, la función de impulso respuesta es idéntica a los coeficientes del proceso MA(q).

Ejemplo 4:

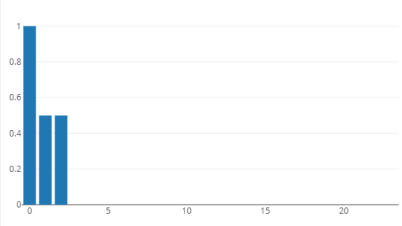
MA(2) y MA(4)

$$y_t = \epsilon_t + 0.5\epsilon_{t-1} + 0.5\epsilon_{t-2}$$

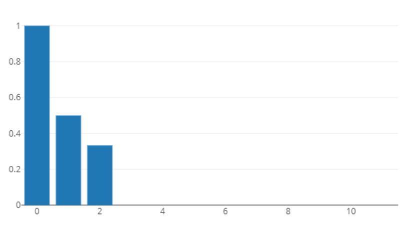
Realización



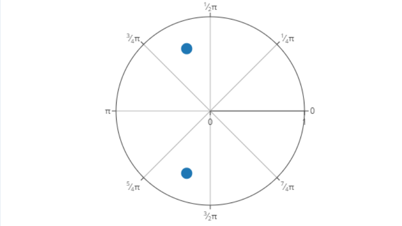
Función impulso respuesta



Autocorrelación

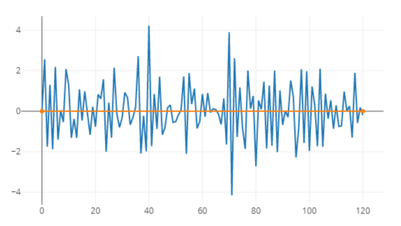


Raíces de $z^2 + 0.5z + 0.5$

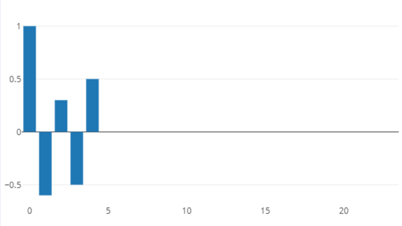


$$y_t = \epsilon_t - 0.6\epsilon_{t-1} + 0.3\epsilon_{t-2} - 0.5\epsilon_{t-3} + 0.5\epsilon_{t-4}$$

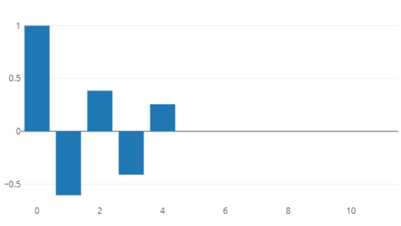
Realización



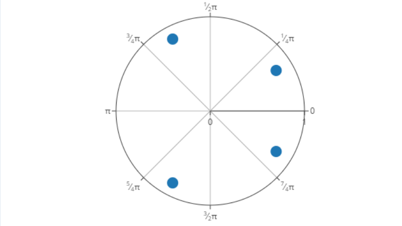
Función impulso respuesta



Autocorrelación



Raíces de $z^4 - 0.6z^3 + 0.3z^2 - 0.5z + 0.5$



Proceso media móvil de orden infinito: MA(∞)

- Definimos el proceso MA(∞) como

$$y_t = \mu + \psi_0\epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \dots$$

- Su media es

$$\mathbb{E}[y_t] = \mu$$

- Su varianza es

$$\gamma_0 = \left(\psi_0^2 + \psi_1^2 + \psi_2^2 + \dots\right)\sigma^2$$

la cual es finita siempre y cuando

$$\sum_{j=0}^{\infty}\psi_j^2 < \infty$$

3. Proceso autorregresivo AR(p)

Proceso autorregresivo de primer orden: AR(1)

- Sea $\{\epsilon_t\}_{t=-\infty}^{\infty}$ un proceso ruido blanco.
- Se define el proceso AR(1) como:

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \epsilon_t \\ y_t - \phi y_{t-1} &= c + \epsilon_t \\ (1 - \phi\,\mathrm{L})y_t &= c + \epsilon_t \end{aligned}$$

- Siempre que $|\phi| < 1$ podemos invertir el término $(1 - \phi\,\mathrm{L})$

$$\begin{aligned} y_t &= (1 - \phi\,\mathrm{L})^{-1}\,(c + \epsilon_t) \\ &= (1 - \phi\,\mathrm{L})^{-1}c + (1 - \phi\,\mathrm{L})^{-1}\epsilon_t \\ &= \frac{c}{1-\phi} + \left(1 + \phi\,\mathrm{L} + \phi^2\,\mathrm{L}^2 + \dots\right)\epsilon_t \\ &= \frac{c}{1-\phi} + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots \end{aligned}$$

- Vemos por tanto que si $|\phi| < 1$ el proceso AR(1) puede escribirse como el proceso MA(∞)

$$y_t = \frac{c}{1-\phi} + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots$$

- Por lo que su valor esperado es:

$$\mathbb{E}[y_t] \equiv \mu = \frac{c}{1-\phi}$$

- y su varianza es

$$\begin{aligned}\operatorname{Var}[y_t] \equiv \gamma_0 &= \left(1 + \phi^2 + \phi^4 + \phi^6 + \dots\right)\sigma^2 \\ &= \frac{\sigma^2}{1-\phi^2}\end{aligned}$$

- La representación MA(∞) nos muestra que y_t depende de la perturbación presente ϵ_t y de todas las pasadas $\epsilon_{t-1}, \epsilon_{t-2}, \ldots,$ pero no depende de ninguna de las perturbaciones futuras $\epsilon_{t+1}, \epsilon_{t+2}, \ldots$.
- Por ello, la covarianza de y_t con cualquier perturbación posterior ϵ_{t+k} (con $k > 0$) es cero.

- Sabiendo que un proceso AR(1) es estacionario si $|\phi| < 1$, es fácil obtener sus momentos sin necesidad de obtener su representación MA(∞).
- En el caso de la media:

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \epsilon_t \\ \mathbb{E}[y_t] &= c + \phi\, \mathbb{E}[y_{t-1}] + \mathbb{E}[\epsilon_t] \\ \mu &= c + \phi\mu + 0 \\ \Rightarrow \mu &= \frac{c}{1-\phi} \end{aligned}$$

- Restando la tercera línea de la primera, vemos que el proceso puede escribirse como un AR(1) para la desviación respecto a la media sin la constante:

$$\underbrace{y_t - \mu}_{\tilde{y}_t} = \phi(\underbrace{y_{t-1} - \mu}_{\tilde{y}_{t-1}}) + \epsilon_t$$

- Para obtener su varianza elevamos al cuadrado la expresión anterior y tomamos su valor esperado

$$
\begin{aligned}
\mathbb{E}\left(\tilde{y}_t^2\right) &= \mathbb{E}\left[\left(\phi \tilde{y}_{t-1} + \epsilon_t\right)^2\right] \\
\gamma_0 &= \mathbb{E}\left[\phi^2 \tilde{y}_{t-1}^2 + 2\phi \tilde{y}_{t-1}\epsilon_t + \epsilon_t^2\right] \\
&= \phi^2 \gamma_0 + 2 \times 0 + \sigma^2 \\
\Rightarrow \gamma_0 &= \frac{\sigma^2}{1-\phi^2}
\end{aligned}
$$

- Para obtener su autocovarianza γ_j, para $j \geq 1$, escribimos

$$\tilde{y}_t = \phi\tilde{y}_{t-1} + \epsilon_t$$

multiplicamos ambos lados por $\tilde{y}_{t-j}$ y tomamos esperanza

$$\begin{aligned}\mathbb{E}\left[\tilde{y}_t\tilde{y}_{t-j}\right] &= \mathbb{E}\left[\phi\tilde{y}_{t-1}\tilde{y}_{t-j} + \tilde{y}_{t-j}\epsilon_t\right] \\ \gamma_j &= \phi\gamma_{j-1}\end{aligned}$$

- Vemos que la función de autocorrelación es la solución de una ecuación en diferencia homogénea de primer orden, cuya solución es

$$\gamma_j = \phi^j\gamma_0$$

- Resumiendo los resultados que hemos obtenido

$$
\begin{aligned}
\mathbb{E}[y_t] = \mu &= \frac{c}{1-\phi} \\
\operatorname{Var}[y_t] = \gamma_0 &= \frac{\sigma^2}{1-\phi^2} \\
\operatorname{Cov}[y_t, y_{t-j}] = \gamma_j &= \phi^j \gamma_0 \\
\rho_j &= \phi^j
\end{aligned}
$$

vemos que ninguno de estos momentos depende del tiempo t, por lo que el proceso AR(1) es covarianza-estacionario siempre y cuando $|\phi| < 1$.

- A partir de la representación MA(∞) del proceso AR(1) con $|\phi| < 1$

$$y_t = \mu + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots$$

es fácil observar que su función de impulso respuesta

$$\Psi(j) = \frac{\partial y_{t+j}}{\partial \epsilon_t} = \frac{\partial y_t}{\partial \epsilon_{t-j}} = \phi^j$$

es idéntica a su función de autocorrelaciones.

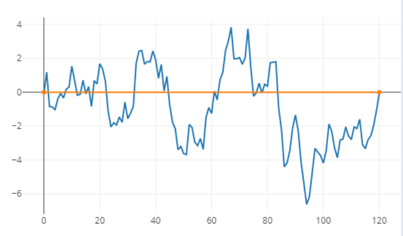
Ejemplo 5:

AR(1)

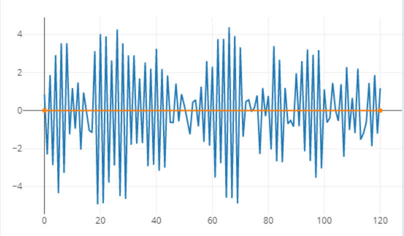
$$y_t = 0.9y_{t-1} + \epsilon_t$$

$$y_t = -0.9y_{t-1} + \epsilon_t$$

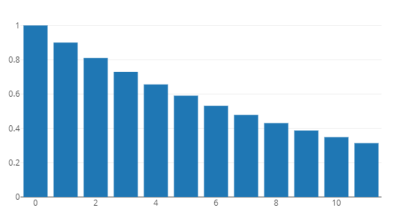
Realización



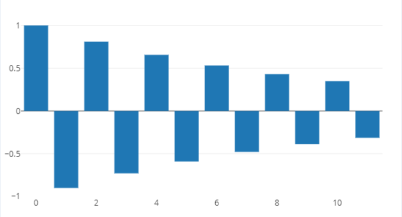
Realización



Autocorrelación



Autocorrelación



El proceso AR(p)

- Es fácil extender el proceso AR(1) para incluir más rezagos.
- El proceso AR(p) es

$$
\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \\
y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} &= c + \epsilon_t \\
\left(1 - \phi_1 \mathrm{L}^1 - \cdots - \phi_p \mathrm{L}^p\right) y_t &= c + \epsilon_t \\
\Phi(\mathrm{L}) y_t &= c + \epsilon_t
\end{aligned}
$$

- El proceso AR(p) es una ecuación en diferencia de orden p.
- Esta ecuación es estable si y solo si las raíces del polinomio $1 - \phi_1 z^1 - \cdots - \phi_p z^p$ están todas fuera del círculo unitario.

- Si el proceso es estable, resolvemos para y_t

$$\begin{aligned} y_t &= \Phi^{-1}(\mathrm{L})\left(c + \epsilon_t\right) \\ &= \Phi^{-1}(1)c + \Phi^{-1}(\mathrm{L})\epsilon_t \\ &= \frac{c}{1 - \phi_1 - \cdots - \phi_p} + \Phi^{-1}(\mathrm{L})\epsilon_t \end{aligned}$$

- Su valor esperado es

$$\mathbb{E}\left(y_t\right) = \frac{c}{1 - \phi_1 - \cdots - \phi_p} = \mu$$

- Similar a lo que obtuvimos para el proceso AR(1), podemos escribir el proceso AR(p) en términos de desviación de la media $\tilde{y} \equiv y - \mu$.

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \cdots + \phi_p \tilde{y}_{t-p} + \epsilon_t$$

- Para obtener su varianza y autocovarianzas multiplicamos la expresión anterior por $\tilde{y}_{t-j}$, con $j \geq 0$, y calculamos el valor esperado

$$\mathbb{E}\left[\tilde{y}_t\tilde{y}_{t-j}\right] = \mathbb{E}\left[\phi_1\tilde{y}_{t-1}\tilde{y}_{t-j} + \cdots + \phi_p\tilde{y}_{t-p}\tilde{y}_{t-j} + \epsilon_t\tilde{y}_{t-j}\right]$$

$$\gamma_j = \begin{cases} \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma^2 & \text{si } j = 0 \\ \phi_1\gamma_{j-1} + \cdots + \phi_p\gamma_{j-p} & \text{si } j > 0 \end{cases}$$

- Si dividimos la última línea entre la varianza γ_0 obtenemos las

Ecuaciones Yule-Walker

$$\rho_j = \phi_1\rho_{j-1} + \cdots + \phi_p\rho_{j-p}, \text{ con } j > 0$$

- Esta es la misma ecuación en diferencia que el proceso original, por lo que en principio se puede resolver con los métodos convencionales, una vez que tengamos p valores iniciales $\rho_0, \rho_1, \ldots, \rho_{p-1}$.
- En general, no es tan sencillo obtener despejar los valores de las autocorrelaciones ρ_j.

- Para obtener la función de impulso respuesta, podríamos partir de la forma

$$y_t = \mu + \Phi^{-1}(\mathrm{L})\epsilon$$

pero esto requiere que obtengamos la forma explícita del polinomio de rezagos

$$\Phi^{-1}(\mathrm{L}) = 1 + \psi_1 \mathrm{L} + \psi_2 \mathrm{L}^2 + \dots$$

- Ahora bien, reconociendo que un proceso AR(p) es una ecuación en diferencia de orden p, encontramos que

$$\frac{\partial y_{t+j}}{\partial \epsilon_t} = c_1\lambda_1^j + c_2\lambda_2^j + \cdots + c_p\lambda_p^j$$

donde λ_j son las raíces del polinomio característico $\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p$ (ver tema 2, p.19-20).

- Para un proceso AR(p) estacionario, sabemos que todas las raíces λ_j están dentro del círculo unitario.
- Por ello, a partir de

$$\frac{\partial y_{t+j}}{\partial \epsilon_t} = c_1\lambda_1^j + c_2\lambda_2^j + \cdots + c_p\lambda_p^j$$

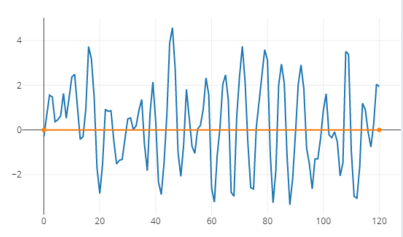
  sabemos que la función de impulso respuesta converge a cero en el largo plazo.
- La forma de la función depende especialmente del valor λ_j más grande:
  - Si es positivo, decae geométricamente,
  - Si es negativo, decae alternando de signo,
  - Si es complejo, decae oscilando periódicamente (ver tema 2, p.25).

Ejemplo 6:

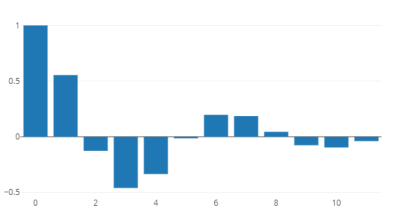
AR(2): $y_t = 0.9y_{t-1} - 0.625y_{t-2} + \epsilon_t$

$$y_t = 0.9y_{t-1} - 0.625y_{t-2} + \epsilon_t$$

Realización



Autocorrelación



Con las ecuaciones Yule-Walker

$$\phi_1 + \phi_2\rho_1 = \rho_1$$
$$\phi_1\rho_1 + \phi_2 = \rho_2$$

encontramos

$$\rho_1 = \frac{\phi_1}{1-\phi_2} \approx 0.5538$$
$$\rho_2 = \phi_1\rho_1 + \phi_2 \approx -0.1265$$

Las demás autocorrelaciones las encontramos con

$$\rho_j = 0.9\rho_{j-1} - 0.625\rho_{j-2}$$

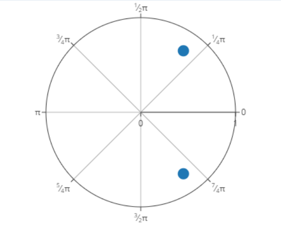
La ecuación característica es

$$\lambda^2 - 0.9\lambda + 0.625 = 0$$

con raíces

$$\lambda = 0.45 \pm 0.65i$$
$$\approx 0.79\left(\cos 0.97 \pm i \sin 0.97\right)$$

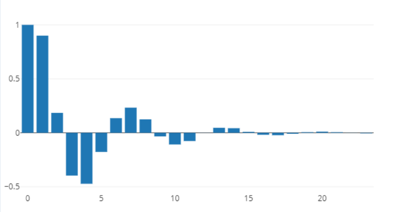
por lo que el proceso es estacionario.



Usando las fórmulas del tema 2 (pp.19-25), encontramos su función impulso respuesta

$$0.79^j\left[\cos(0.97j) - \tfrac{9}{13}\sin(0.97j)\right]$$

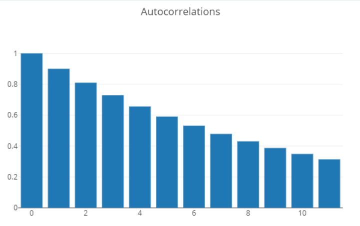
la cual converge a cero, oscilando periódicamente conforme $j \to \infty$.



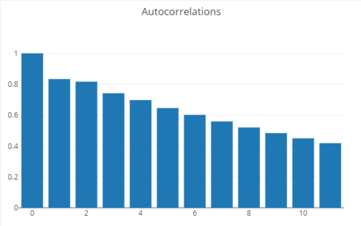
Ejemplo 7:
AR(1) vs AR(2)

- Consideremos estos dos procesos autorregresivos:

$$y_t = 0.9y_{t-1} + \epsilon_t$$



$$y_t = 0.5y_{t-1} + 0.4y_{t-2} + \epsilon_t$$



- Este caso ilustra que es sumamente difícil identificar el orden p de un proceso AR(p) a partir de su autocorrelograma.
- Para resolver este problema, a continuación estudiaremos la autocorrelación parcial.

Autocorrelación parcial

- La autocorrelación parcial mide la correlación restante entre y_t y y_{t-k} una vez que se han eliminado la influencia de $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$.

$$y_t = \underbrace{a_1^{(k)} y_{t-1} + a_2^{(k)} y_{t-2} + \cdots + a_{k-1}^{(k)} y_{t-k+1}}_{\text{“eliminamos” este efecto}} + a_k^{(k)} y_{t-k}$$

- Es decir, las primeros m autocorrelaciones parciales vienen de

$$\begin{aligned}
y_t &= a_1^{(1)} y_{t-1} \\
y_t &= a_1^{(2)} y_{t-1} + a_2^{(2)} y_{t-2} \\
&\vdots \\
y_t &= a_1^{(m-1)} y_{t-1} + a_2^{(m-1)} y_{t-2} + \cdots + a_{m-1}^{(m-1)} y_{t-m+1} \\
y_t &= a_1^{(m)} y_{t-1} + a_2^{(m)} y_{t-2} + \cdots + a_{m-1}^{(m)} y_{t-m+1} + a_m^{(m)} y_{t-m}
\end{aligned}$$

- Para encontrar el valor de $a_k^{(k)}$ basta con resolver

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1^{(k)} \\ a_2^{(k)} \\ a_3^{(k)} \\ \vdots \\ a_k^{(k)} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix}$$

- En adelante, denotaremos la k-ésima correlación parcial por $\varphi(k) \equiv a_k^{(k)}$

- Comparando las ecuaciones del proceso AR(p) y de la autocorrelación parcial k:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$
$$y_t = a_1^{(k)} y_{t-1} + a_2^{(k)} y_{t-2} + \cdots + a_{k-1}^{(k)} y_{t-k+1} + a_k^{(k)} y_{t-k}$$

  vemos que

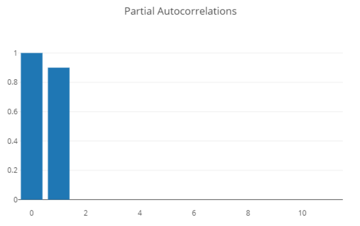
  - si $k = p$, entonces $\varphi_k = \phi_p$
  - si $k > p$, entonces $\varphi_k = 0$
  - si $k = 1$, entonces $\varphi_1 = \rho_1$

Ejemplo 8:

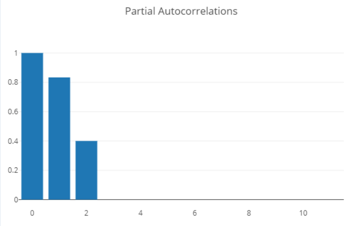
AR(1) vs AR(2), con autocorrelación parcial

- Consideremos de nuevo estos procesos autorregresivos:

$y_t = 0.9y_{t-1} + \epsilon_t$



$y_t = 0.5y_{t-1} + 0.4y_{t-2} + \epsilon_t$



- Ahora es muy sencillo identificar el orden p de los proceso AR.

Para encontrar la segunda autocorrelación parcial, resolvemos

$$\begin{bmatrix} a_1 \\ \varphi_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \frac{1}{1-\rho_1^2} \begin{bmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

de donde $\varphi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$.

- Para el proceso AR(1) sabemos que $\rho_k = \phi^k$, con lo que comprobamos que $\varphi_2 = \frac{\phi^2 - \phi^2}{1-\phi^2} = 0$.
- En un ejemplo anterior encontramos que para un proceso AR(2) se cumple

$$\rho_1 = \frac{\phi_1}{1-\phi_2} = \frac{5}{6} \qquad \rho_2 = \phi_1\rho_1 + \phi_2 = \frac{49}{60}$$

Al sustituir en la expresión para φ_2 comprobamos que

$$\varphi_2 = \frac{\frac{49}{60} - \frac{25}{36}}{1 - \frac{25}{36}} = 0.4 = \phi_2.$$

4. Proceso autorregresivo de media móvil ARMA(p,q)

Procesos ARMA

Definición: proceso ARMA

Sea $\{\epsilon_t\}$ un proceso ruido blanco; el proceso estocástico

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

con $\phi_p, \theta_q \neq 0$ es llamado un proceso ARMA(p,q).

- ARMA = AutoRegressive Moving Average, (autorregresivo media móvil)
- Los procesos ARMA son importantes porque todo proceso estacionario puede ser aproximado por un proceso ARMA.

- Similar a lo que encontramos con los procesos AR(p), si asumimos que es estacionario su media satisface la relación

$$\mu = c + \phi_1 \mu + \cdots + \phi_p \mu$$

- Por lo que el proceso puede escribirse sin el intercepto, si expresamos la variable y como desviación de su media

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \cdots + \phi_p \tilde{y}_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

- Las fórmulas para la varianza y la autocovarianza se obtienen aplicando las definiciones del caso, pero tienden a ser más complicadas.

El proceso $\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \cdots + \phi_p \tilde{y}_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$ puede expresarse en términos de polinomios de rezagos:

$$\begin{aligned}
\tilde{y}_t - \phi_1 \tilde{y}_{t-1} - \cdots - \phi_p \tilde{y}_{t-p} &= \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\
\left(1 - \phi_1 \mathrm{L} - \cdots - \phi_p \mathrm{L}^p\right) \tilde{y}_t &= \left(1 + \theta_1 \mathrm{L} + \cdots + \theta_q \mathrm{L}^q\right) \epsilon_t \\
\Phi(\mathrm{L}) \tilde{y}_t &= \Theta(\mathrm{L}) \epsilon_t
\end{aligned}$$

Estabilidad

Si las raíces del polinomio $\Phi(z)$ están todas fuera del círculo unitario, el proceso es estable.

$$\tilde{y}_t = \Phi(\mathrm{L})^{-1} \Theta(\mathrm{L}) \epsilon_t$$

Invertibilidad

Si las raíces del polinomio $\Theta(z)$ están todas fuera del círculo unitario, el proceso es invertible.

$$\epsilon_t = \Theta(\mathrm{L})^{-1} \Phi(\mathrm{L}) \tilde{y}_t$$

Sobreparametrización de un proceso ARMA

- Supongamos que tenemos un proceso ARMA(p, q) $\Phi(\mathrm{L})y_t = \Theta(\mathrm{L})\epsilon_t$, en el cual los polinomios $\Theta(\mathrm{L})$ tiene una raíz en común. En este caso podemos

$$
\begin{aligned}
\Phi(\mathrm{L})y_t &= \Theta(\mathrm{L})\epsilon_t \\
(1 - r\,\mathrm{L})\Phi^*(\mathrm{L})y_t &= (1 - r\,\mathrm{L})\Theta^*(\mathrm{L})\epsilon_t \\
\Phi^*(\mathrm{L})y_t &= \Theta^*(\mathrm{L})\epsilon_t
\end{aligned}
$$

- Es decir, podemos representar el mismo proceso con un modelo ARMA(p-1, q-1).
- Decimos que:
  - el modelo ARMA(p,q) está sobreparametrizado.
  - el modelo ARMA(p-1, q-1) es una representación más parsimoniosa del proceso generador de datos.

Ejemplo 9:

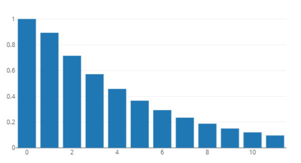
Sobreparametrización

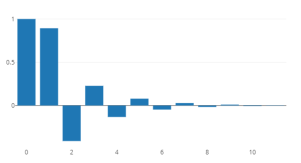
Consideremos estos dos procesos ARMA

Autocorrelación | Autocorrelación parcial

$$y_t = 1.3y_{t-1} - 0.4y_{t-2} + \epsilon_t + 0.1\epsilon_{t-1} - 0.3\epsilon_{t-2}$$

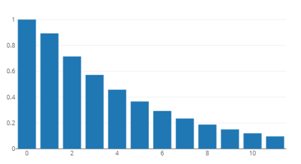
ARMA(2,2)

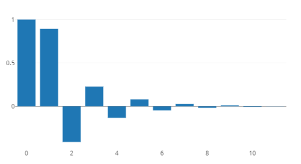




$$y_t = 0.8y_{t-1} + \epsilon_t + 0.6\epsilon_{t-1}$$

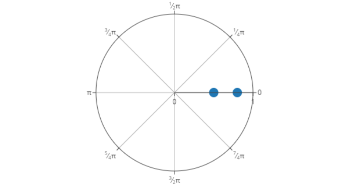
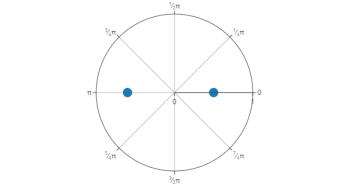
ARMA(1,1)





¡Sus funciones ACF y PACF son idénticas!

- Las raíces de los polinomios de rezagos del proceso ARMA(2,2)

Raíces AR	Raíces MA
	

- Vemos que ambas tienen a $z = 0.5$ como una raíz.
- Al "eliminarla" de ambos polinomios, terminamos con el mismo proceso pero con menos parámetros.
- Esta representación ARMA$(1, 1)$ es una versión más parsimoniosa del mismo proceso.

5. Estimación de modelos ARMA

Introducción (muy breve) al estimador de máxima verosimilitud

- Sea $f(y|\theta)$ la función de densidad conjunta de la variable $Y = [Y_1, \ldots, Y_n]$. Entonces, para una muestra observada $\mathbf{y}$ de esta distribución, la función del vector de parámetros θ definida por
$$L(\theta|\mathbf{y}) = f(\mathbf{y}|\theta)$$
se conoce como la función de verosimilitud.
- El estimador de máxima verosimilitud es el valor del vector de parámetros θ que maximiza la función de verosimilitud
$$\hat{\theta}_{\text{ML}} \equiv \underset{\theta}{\operatorname{argmax}} \underbrace{\mathcal{L}(\theta|\mathbf{y})}_{\ln L[(\theta|\mathbf{y})]}$$
- Es decir, $\hat{\theta}_{\text{ML}}$ es el parámetro que hace más plausible haber obtenido la muestra $\mathbf{y}$ si la verdadera distribución era $f(y|\theta)$.

Ejemplo 10:

Estimador máximo verosímil de una distribución exponencial

- Supongamos que tenemos una muestra $\{x_1, \dots, x_N\}$ de valores tomados de realizaciones independientes de una distribución exponencial con parámetro λ
- La función de densidad de una observación es $f(x|\lambda) = \lambda e^{-\lambda x}$
- La función de verosimilitud es la probabilidad conjunta de observar esta muestra:

$$L(\lambda|x_1, \dots, x_N) = \prod_{i=1}^{N} \lambda e^{-\lambda x_i}$$

o bien, tomando su logaritmo

$$\begin{aligned}\mathcal{L}(\lambda|x_1, \dots, x_N) &= \sum_{i=1}^{N} [\ln \lambda - \lambda x_i] \\ &= N \ln \lambda - \lambda \sum_{i=1}^{N} x_i\end{aligned}$$

- Para encontrar el máximo:

$$\frac{\partial \mathcal{L}(\lambda)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{i=1}^{N} x_i = 0$$

- Por lo tanto, el estimador de máxima verosimilitud es:

$$\hat{\lambda}_{\mathsf{ML}} = \frac{N}{\sum_{i=1}^{N} x_i} = \frac{1}{\bar{x}}$$

donde $\bar{x}$ es el promedio simple de los datos.

- Pensemos en un proceso ARMA estacionario como una distribución conjunta de $Y_1, \dots, Y_T$, donde

$$\mathbb{E}(Y_t) = \mu \qquad\qquad \text{Cov}(Y_t, Y_{t-j}) = \gamma_j$$

- Entonces

$$\mathbb{E}\begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \mu$$

$$\text{Var}\begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix} = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \dots & 1 \end{bmatrix} = \Omega$$

- Supongamos que tenemos una serie de tiempo débilmente estacionaria, con datos generados por el proceso ARMA descrito anteriormente.

$$\mathbf{y} = [y_1, y_2, \ldots, y_T]'$$

- Si asumimos que el ruido blanco ϵ_t tiene una distribución normal, entonces el proceso ARMA tiene una distribución normal multivariada
- Su función de log-verosimilitud es

$$\mathcal{L}(\theta) = -\frac{T}{2}\ln(2\pi) + \frac{1}{2}\ln|\Omega^{-1}| - \frac{1}{2}(\mathbf{y}-\mu)'\Omega^{-1}(\mathbf{y}-\mu)$$

donde

$$\theta = [c, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]$$

- En general, los modelos ARMA se estiman por el método de máxima verosimilitud.
- Para ello, se igualan a cero las derivadas de la función $\mathcal{L}(\theta)$ con respecto a cada uno de los parámetros presentes en θ.
- El resultado es un sistema de ecuaciones no lineales que carece de solución cerrada.
- Por ello, es necesario recurrir a métodos numéricos para resolver estos sistemas de ecuaciones.
- Por razones de tiempo, no cubrimos estos métodos en este curso. Para más detalles, consulte Greene (2012, capítulo 5) y Miranda y Fackler (2002, capítulos 3 y 4)

Estimación de modelos autorregresivos

- Para el modelo AR(p)

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \qquad \text{Var}(\epsilon_t) = \sigma^2$$

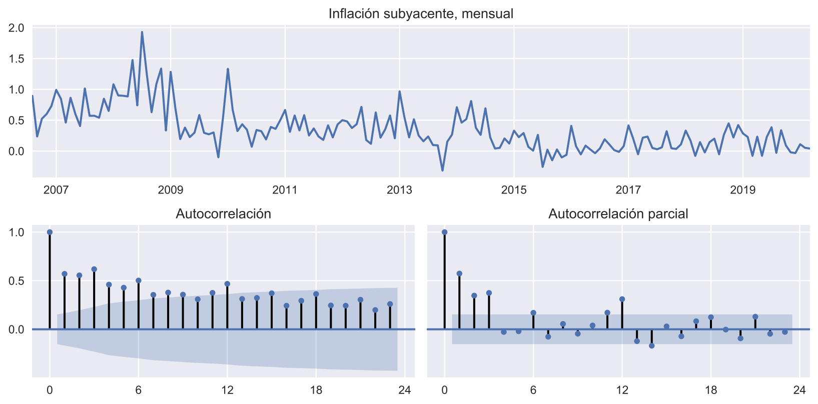
  es posible estimar los parámetros $c, \phi_1, \ldots, \phi_p$ por mínimos cuadrados ordinarios.
- El resultado es equivalente a un estimador de máxima verosimilitud condicional.
- Se llama así porque los primeros p valores de la serie se toman como dados.
- Si la muestra es grande, el resultado es equivalente al estimador de máxima verosimilitud exacto.
- El parámetro σ^2 se estima como el promedio de los cuadrados de los residuos de la regresión anterior.

Ejemplo 11:

Estimando un modelo AR(3) para la inflación mensual

- `bccr.ServicioWeb`
- `ISI-AR3.ipynb`

Veamos la serie histórica y los autocorrelogramas de la inflación mensual subyacente de Costa Rica



Consideremos este proceso AR(3) para modelar la inflación mensual subyacente

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \epsilon_t$$

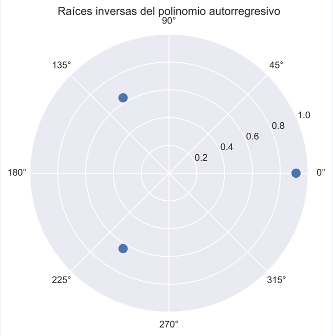
la cual estimamos con datos mensuales (2006m08 a 2020m03, indicador 25725).

	coef	**std err**	**z**	**P$>$ |z|**	**[0.025**	**0.975]**
const	0.3535	0.104	3.384	0.001	0.149	0.558
ar.L1.isi	0.2534	0.073	3.469	0.001	0.110	0.396
ar.L2.isi	0.1996	0.074	2.679	0.008	0.054	0.346
ar.L3.isi	0.3690	0.073	5.059	0.000	0.226	0.512

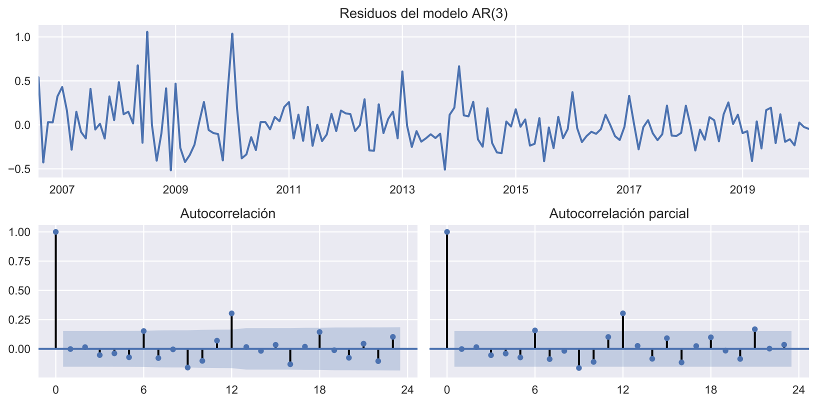
- El modelo ajustado es

$$\tilde{y}_t = 0.2534\tilde{y}_{t-1} + 0.1996\tilde{y}_{t-2} + 0.3690\tilde{y}_{t-3} + \epsilon_t$$
$$\tilde{y}_t = y_t - 0.3535$$
$$\epsilon_t \sim \mathsf{N}(0, 0.2522^2)$$

- Las raíces inversas del polinomio $1 - 0.253\,\mathrm{L} - 0.200\,\mathrm{L}^2 - 0.369\,\mathrm{L}^3$ están dentro del círculo unitario, por lo que el proceso es estacionario.



Los residuos de la regresión parecen ruido blanco, excepto por la presencia de un efecto estacional (rezago 12).



En Stata:

```
. arima isi, arima(3,0,0)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -7.3405352
Iteration 1:   log likelihood = -7.3208124
Iteration 2:   log likelihood = -7.3199769
Iteration 3:   log likelihood = -7.3197271
Iteration 4:   log likelihood = -7.3196206
(switching optimization to BFGS)
Iteration 5:   log likelihood = -7.3195774
Iteration 6:   log likelihood = -7.3195452
Iteration 7:   log likelihood = -7.3195446

ARIMA regression

Sample:  2006m8 - 2020m3                        Number of obs     =        164
                                                Wald chi2(3)      =     202.08
Log likelihood = -7.319545                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
         isi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
isi          |
       _cons |   .3534968   .1181878     2.99   0.003      .121853    .5851406
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .2533589   .0724494     3.50   0.000     .1113607    .3953571
         L2. |   .1995733   .0514003     3.88   0.000     .0988305    .3003161
         L3. |   .3689968   .0712193     5.18   0.000     .2294095    .5085841
-------------+----------------------------------------------------------------
      /sigma |   .2521808   .0103002    24.48   0.000     .2319929    .2723688
------------------------------------------------------------------------------
```

Iteraciones del método numérico

Selección de modelos

- En la estimación de máxima verosimilitud de un modelo ARMA(p,q) hay un supuesto implícito: que sabemos el orden del proceso, es decir, que sabemos el número "correcto" de rezagos p, q
- En la práctica, esto rara vez sucede.
- Tenemos una disyuntiva: Entre más rezagos incluyamos
  - "mejor" será el ajuste del modelo a los datos.
  - "peor" se vuelve la precisión de los parámetros que se estiman.
- La metodología de Box-Jenkins sugiere buscar modelos parsimoniosos.

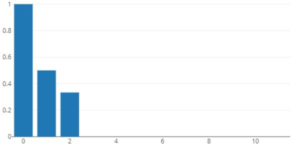
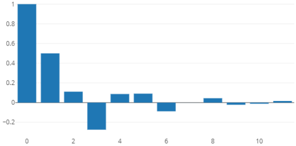
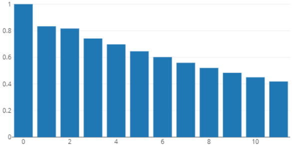
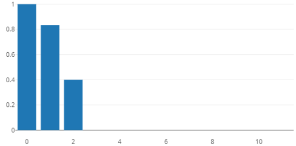
- La parsimonia (usar tan pocos parámetros como sea necesario) tiene sus beneficios a la hora de hacer pronósticos.
- Muchos modelos estructurales complejos tienen un ajuste muy alto a la muestra en que se estiman, pero hacen pronósticos muy pobres fuera de la muestra.
- Sorprendentemente, modelos ARMA univariados sencillos pueden hacer mejores pronósticos.
- La idea es que entre más parámetros haya que estimar, más posibilidad hay de hacerlo mal.

La filosofía de modelación de Box-Jenkins

El enfoque para pronosticar de Box-Jenkins consiste de cuatro etapas:

1. Transformar los datos (de ser necesario) para que el supuesto de estacionariedad sea razonable.
2. Adivinar valores pequeños de p y q para un modelo ARMA(p,q) que pueda describir la serie transformada.
3. Estimar los parámetros de $\phi(\mathrm{L})$ y $\theta(\mathrm{L})$.
4. Realizar análisis de diagnóstico para confirmar que el modelo es consistente con los datos observados.

Distinguiendo los procesos AR(p) de los MA(q)

	Autocorrelación	Autocorrelación parcial
MA(q)	se parte en q 	no se parte 
AR(p)	no se parte 	se parte en p 

- El autocorrelograma y el autocorrelograma parcial ayudan a reconocer series ARMA(p,0) y ARMA(0, q), pero no series ARMA(p,q) con $pq \neq 0$.
- Para esto, recurrimos a criterios de selección.
- Estos criterios tratan de resolver la disyuntiva de que a mayores valores de p, q, "mejor" será el ajuste pero "peor" la precisión de la estimación.
- Los criterios más usuales son el de Akaike (Akaike) y el de Bayes (BIC).

- Sean $\mathcal{L}$ el máximo de la función log-verosimilitud, T el número de observaciones, y $K = p + q + 2$ el número de parámetros estimado.
- Entonces

Criterio de información de Akaike

$$\text{AIC} = \underbrace{-2\mathcal{L}}_{\text{“desajuste”}} + \underbrace{2K}_{\text{penalización}}$$

Criterio de información de Bayes

$$\text{BIC} = \underbrace{-2\mathcal{L}}_{\text{“desajuste”}} + \underbrace{\ln(T)K}_{\text{penalización}}$$

- Se escoge la combinación p, q que minimiza el criterio de información.
- En la práctica, en ocasiones AIC y BIC escogen modelos distintos.

Ejemplo 12:

Seleccionando p, q para un modelo ARMA de inflación

- bccr.ServicioWeb
- ISI-AR3.ipynb

En Stata, para el modelo AR(3) que estimamos en el ejemplo anterior:

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	T= 164	$\mathcal{L}$=	-7.319545	K= 5	24.63909	40.13842

Calculamos los criterios de información:

$$\begin{aligned} \text{AIC} &= -2\mathcal{L} + 2K \\ &= -2 \times -7.3195 + 2 \times 5 = 24.63909 \end{aligned}$$

$$\begin{aligned} \text{BIC} &= -2\mathcal{L} + \ln(T)K \\ &= -2 \times -7.3195 + \ln(164) \times 5 = 40.13842 \end{aligned}$$

- Si calculamos los dos criterios para una combinación de valores p, q obtenemos

AIC

	q=0	q=1	q=2
p=0	128.02	89.62	83.77
p=1	64.31	30.74	31.92
p=2	46.27	32.21	34.44
p=3	24.64	26.51	28.32
p=4	26.50	28.43	25.10

BIC

	q=0	q=1	q=2
p=0	134.22	98.92	96.17
p=1	73.61	43.14	47.42
p=2	58.67	47.71	53.04
p=3	40.14	45.11	50.02
p=4	45.10	50.13	49.90

- Esto nos indica que la serie de tiempo debe modelarse como un proceso AR(3).

6. Pronósticos con modelos ARMA

Pronósticos con modelos ARMA

- Sea $y^*_{t+j|t}$ un pronóstico de y_{t+j} pasado en datos observados hasta t.
- Definimos el "mejor" pronóstico de este tipo como aquel que minimiza el error cuadrático medio

$$\text{MSE}\left(y^*_{t+j|t}\right) \equiv \mathbb{E}\left(y_{t+j} - y^*_{t+j|t}\right)^2$$

- Aunque no lo probamos aquí, resulta que el pronóstico con el menor error cuadrático medio es la esperanza de y_{t+j} condicional en los datos hasta t:

$$y^*_{t+j|t} = \mathbb{E}\left(y_{t+j} | y_t, y_{t-1}, \dots\right) \equiv \mathbb{E}_t\left(y_{t+j}\right)$$

Pronosticando valores puntuales de la serie

- Es muy sencillo obtener de manera recursiva estos pronósticos para los modelos ARMA, siguiendo estos 3 pasos:
  1. Se escribe la ecuación ARMA de manera que y_t quede a la izquierda y todos los otros términos a la derecha.
  2. Se sustituye el índice t por $T + h$.
  3. En el lado derecho de la ecuación, se sustituyen:
     - observaciones futuras con sus pronósticos,
     - errores futuros con cero,
     - errores pasados con sus respectivos residuos.
- Empezando con $h = 1$, se repiten los pasos 2 y 3 para valores $h = 2, 3, \dots$ hasta que todos los pronósticos hayan sido calculados.

Para ilustrar el procedimiento, consideremos este proceso ARMA(1,2)

$$(1-\phi\,\mathrm{L})\tilde{y}_t = (1+\theta_1\,\mathrm{L}+\theta_2\,\mathrm{L}^2)\epsilon_t$$

- Paso 1: $\tilde{y}_t = \phi\tilde{y}_{t-1} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$
- Para $h=1$:
$$\begin{aligned}\tilde{y}_{T+1} &= \phi\tilde{y}_T + \epsilon_{T+1} + \theta_1\epsilon_T + \theta_2\epsilon_{T-1}\\ \tilde{y}^*_{T+1|T} &= \phi\tilde{y}_T + \theta_1\hat{\epsilon}_T + \theta_2\hat{\epsilon}_{T-1}\end{aligned}$$
- Para $h=2$:
$$\begin{aligned}\tilde{y}_{T+2} &= \phi\tilde{y}_{T+1} + \epsilon_{T+2} + \theta_1\epsilon_{T+1} + \theta_2\epsilon_T\\ y^*_{T+2|T} &= \phi\tilde{y}^*_{T+1|T} + \theta_2\hat{\epsilon}_T\end{aligned}$$
- Para $h=3$:
$$\begin{aligned}\tilde{y}_{T+3} &= \phi\tilde{y}_{T+2} + \epsilon_{T+3} + \theta_1\epsilon_{T+2} + \theta_2\epsilon_{T+1}\\ y^*_{T+3|T} &= \phi\tilde{y}^*_{T+2|T}\end{aligned}$$

- Es fácil comprobar que, siguiendo este procedimiento, una vez que $h > p$, $h > q$ la ecuación de pronóstico será

$$\tilde{y}^*_{T+h} = \phi_1 \tilde{y}^*_{T+h-1} + \cdots + \phi_p \tilde{y}^*_{T+h-p}$$

- Esto es una ecuación en diferencia de orden p, de la cual sabemos que sus raíces están dentro del círculo unitario.
- Por ello

$$\lim_{h\to\infty} \tilde{y}^*_{T+h} = 0 \qquad \text{es decir} \qquad \lim_{h\to\infty} y^*_{T+h} = \mu$$

- Esto nos dice que, para valores grandes de h, el pronóstico es simplemente la media del proceso.

Cuantificando la incertidumbre de los pronósticos puntuales

- Para saber qué tan precisos esperamos que sean estos pronósticos, necesitamos cuantificar su error cuadrático medio

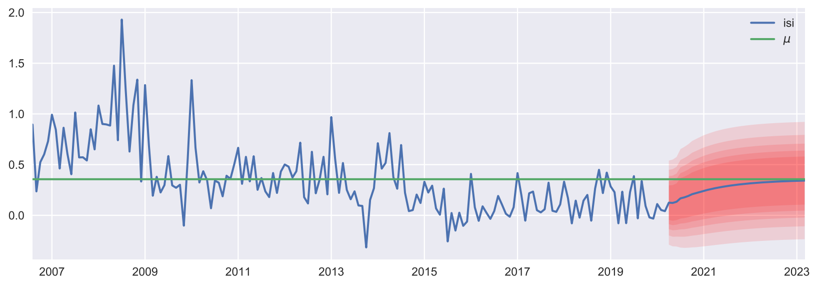
$$\begin{aligned}\text{MSE}\left(y^*_{t+j|t}\right) &\equiv \mathbb{E}\left(y_{t+j} - y^*_{t+j|t}\right)^2 \\ &= \mathbb{E}\left(y_{t+j} - \mathbb{E}_t\left(y_{t+j}\right)\right)^2 \\ &= \mathrm{Var}_t\left(y_{t+j}\right)\end{aligned}$$

- Es decir, necesitamos cuantificar la varianza de y_{t+j} condicional en los datos disponibles en t.
- Por razones de tiempo, no vamos a desarrollar estas fórmulas durante este curso.

Ejemplo 13:

Pronosticando la inflación

- bccr.ServicioWeb
- ISI-AR3.ipynb



isi
μ
2.0
1.5
1.0
0.5
0.0
2007
2009
2011
2013
2015
2017
2019
2021
2023

¿Qué podría salir mal?

- En todas las fórmulas que hemos elaborado para los pronósticos hasta ahora, está implícito que conocemos los verdaderos valores de los parámetros.
- En la práctica, esos parámetros son estimados a partir de los datos.
- Tomemos por ejemplo un proceso AR(1):

$$
\begin{aligned}
y_{t+1} &= \phi y_t + \epsilon_{t+1} && \text{(valor verdadero)} \\
y^*_{t+1|t} &= \hat{\phi} y_t && \text{(pronóstico)}
\end{aligned}
$$

$$
\underbrace{y_{t+1} - y^*_{t+1|t}}_{\text{error de pronóstico}} = \underbrace{\left(\phi - \hat{\phi}\right)}_{\text{“sesgo”}} y_t + \underbrace{\epsilon_{t+1}}_{\text{innovación}}
$$

Referencias I

Becketti, Sean (2013). *Introduction to Time Series Using Stata*. 1ª ed. Stata Press. ISBN: 978-1-59718-132-7.

Enders, Walter (2015). *Applied Econometric Time Series*. 4ª ed. Wiley. ISBN: 978-1-118-80856-6.

Greene, William H. (2012). *Econometric Analysis*. 7ª ed. Prentice Hall. ISBN: 978-0-13-139538-1.

Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.

Kirchgässner, Gebhard, Jürgen Wolters y Uwe Hassler (2013). *Introduction to Modern Time Series Analysis*. 2ª ed. Springer. ISBN: 978-3-642-33435-1.

Miranda, Mario J. y Paul L. Fackler (2002). *Applied Computational Economics and Finance*. MIT Press. ISBN: 0-262-13420-9.