

Taller #8, datos no i.i.d. (series) estacionarias y estacionarias y estacionarias

1. Sea $\rho(k) = \text{Corr}(Z_t, Z_{t+k})$

Para Datos 1

$$Q_{kk} = \text{Corr}(Z_t, Z_{t+k} | Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1})$$

$$\rho(k) = \text{Corr}(W_t, W_{t+k})$$

Para Datos 2

$$Q_{kk} = \text{Corr}(W_t, W_{t+k} | W_{t+1}, W_{t+2}, \dots, W_{t+k-1})$$

$$\rho(k) = \text{Corr}(X_t, X_{t+k})$$

Para Datos 3

$$Q_{kk} = \text{Corr}(X_t, X_{t+k} | X_{t+1}, X_{t+2}, \dots, X_{t+k-1})$$

Se prueba individualmente para $k=1, 2, 3, \dots, 25$

ACF

PACF

$$H_0: \rho(k) = 0$$

$$H_0: Q_{kk} = 0$$

$$H_1: \rho(k) \neq 0$$

$$H_1: Q_{kk} \neq 0$$

$$EP: \hat{\rho}(k) \overset{\text{prox}}{\sim} N(0, \frac{1}{100})$$

$$EP: \hat{Q}_{kk} \overset{\text{prox}}{\sim} N(0, \frac{1}{100})$$

$$CR: |\hat{\rho}(k)| \geq \frac{2}{\sqrt{100}}$$

$$CR: |\hat{Q}_{kk}| \geq \frac{2}{\sqrt{100}}$$

Conclusión: para datos 1 y datos 2 se observan cortes tanto en ACF como en PACF, por lo tanto se rechaza la hipótesis nula en ambos test y se concluye que estos procesos no provienen de un RB. En el caso de datos 3 no hay evidencia en contra de RB en el test ACF pero si hay un rechazo en PACF, por lo tanto tampoco proviene de un RB.

A pesar de que existe correlación en estos procesos, se observa que todos tienen media y varianzas constantes:

$$E(Z_t) = 0$$

$$\text{Var}(Z_t) = \text{Constante } \forall t$$

$$E(W_t) = 0$$

$$\text{Var}(W_t) = \text{Constante } \forall t$$

$$E(X_t) = 0$$

$$\text{Var}(X_t) = \text{Constante } \forall t$$

Por ende, si se comprueba que estos procesos son ergódicos entonces serán también estacionarios en covarianza.

gráficamente, en las ACF se comprueba que:

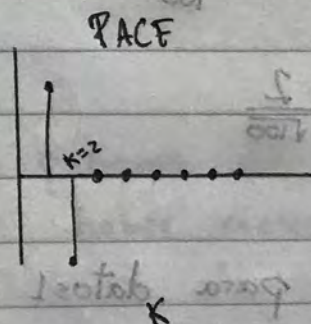
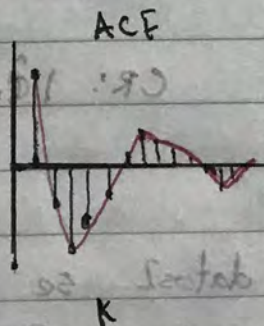
$$\lim_{k \rightarrow \infty} \rho(k) = 0 \text{ rápidamente para cada caso.}$$

Por lo tanto, estos procesos son ergódicos y se concluye que son estacionarios en covarianza.

Identificación de modelos ARMA(p,q)

• ACF-PACF:

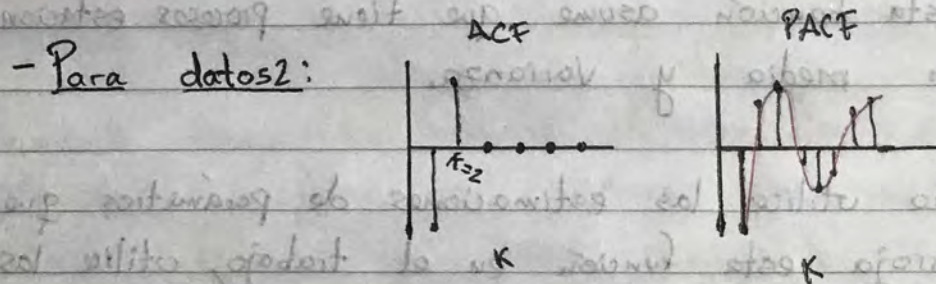
-Para datos1:



Se observa un patrón de cola (exponencial sinusoidal) en la ACF y un patrón de corte en la PACF con el último corte en $k=2$, Así que se identifica un modelo AR(2)

$$\text{AR}(2): Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t \text{ con } a_t \sim \text{RB} \sim N(0, \sigma^2)$$

PD: El patrón de cola exponencial sinusoidal en la ACF implica que los modelos identificados deben tener parte $AR(p)$ con $p \geq 2$ y que el polinomio autorregresivo $\Phi_p(B)$ tiene raíces complejas



Se identifica un modelo $MA(z)$, ya que hay patrón de corte en la ACF con el último corte en $k=2$ y patrón de cola (exponencial sinusoidal) en la PACF.

$$MA(z): W_t = \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t \text{ con } a_t \text{ un RB} \sim N(0, \sigma^2)$$

PD: El patrón de cola exponencial sinusoidal en la PACF implica que los modelos identificados deben tener parte $MA(q)$ con $q \geq 2$ y que el polinomio $\Theta_q(B)$ tiene raíces complejas.

- Para datos3: En este caso los patrones no son muy claros, pero al no haber rechazos en ACF no debería hablarse de un patrón de corte, Así que se asumirá un patrón de cola negativa tanto en ACF como en PACF.

Como en este caso ambos patrones son de cola, se identifica un modelo $ARMA(p,q)$ pero no es posible identificar los

órdenes p y q .

• Metodos automáticos de p y q no son (p)RA

- auto.arima() → Esta función asume que tiene procesos estacionarios en media y Varianza.

→ No utilice las estimaciones de parámetros que arroja esta función en el trabajo, utilice las de la función Arima().

Datos 1:

• ARMA(2,2) de media cero:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t \quad \text{con } a_t \text{ un } RB \sim N(0, \sigma^2)$$

• ARMA(2,0) = AR(2) → el mismo identificado por ACF-PACF.

Datos 2:

• Sólo se identifican ARMA(0,2) = MA(2) → el mismo identificado por ACF-PACF

Datos 3:

• ARMA(1,1) de media cero

$$X_t = \phi_1 X_{t-1} + \theta_1 a_{t-1} + a_t \quad \text{con } a_t \text{ un } RB \sim N(0, \sigma^2)$$

• ARMA(0,0) → ¡ojo! el método está asumiendo de forma errónea que el proceso es un RB.

• EACF

Recuerde tener en cuenta al identificar modelos con los diferentes métodos que datos1 debe ser modelado con parte $AR(p)$ con $p \geq 2$ y datos2 con parte $MA(q) \geq 2$, en este caso no se da, pero si llegara a identificar un modelo que no cumpla estas condiciones no debe ser tenido en cuenta.

- datos1 : Se ubica el triángulo de "o" de mayor área lo más arriba y a la izquierda posible, en este caso el triángulo trazado tiene vértice en: (2,2), por lo que el modelo identificado es un ARMA(2,2)

ARMA(2,2):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t \text{ con } a_t \text{ un RB} \sim N(0, \sigma^2)$$

- datos2 : Se pueden ubicar 2 triángulos de área grande en los vértices (3,2) y (0,2).

ARMA(3,2):

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \phi_3 W_{t-3} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t \text{ con } a_t \text{ un RB} \sim N(0, \sigma^2)$$

MA(2): el mismo que se había identificado anteriormente.

- datos3: El triángulo de mayor área estaría con vértice en (0,0) así que no se toma porque asumiría RB, por lo tanto se toma el triángulo con vértice en (1,1)

ARMA(1,1): el mismo identificado anteriormente.

Ajustes

- datos1: AR(2): $\hat{z}_t = 0,691882 z_{t-1} - 0,479471 z_{t-2}$

ARMA(2,2): $\hat{z}_t = 1,03527 z_{t-1} - 0,51592 z_{t-2} - 0,42592 \hat{a}_{t-1}$

$-0,22599 \hat{a}_{t-2}$

- datos2: MA(2): $\hat{w}_t = -0,394121 \hat{a}_{t-1} + 0,565021 \hat{a}_{t-2}$

ARMA(3,2): $\hat{w}_t = 0,086058 w_{t-1} - 0,147569 w_{t-2} - 0,02241 w_{t-3}$
 $- 0,4463 \hat{a}_{t-1} + 0,675202 \hat{a}_{t-2}$

- datos3: ARMA(1,1): $\hat{x}_t = 0,489672 x_{t-1} - 0,791686 \hat{a}_{t-1}$

ARMA(2,2): $\hat{x}_t = 1,611587 x_{t-1} - 0,685478 x_{t-2} - 1,948239 \hat{a}_{t-1}$
 $+ 0,999906 \hat{a}_{t-2}$

Se prueba significancia de parámetros: cuando un modelo tenga muchos parámetros interesa probarlos todos para analizar si se puede reducir el modelo; se debe correr el modelo con y sin los parámetros y comparar en calidad de supuestos, ajustes y pronósticos para concluir si es conveniente reducirlo.

Pruebas a realizar:

$K = \#$ parámetros

$$H_0: \theta_j = 0 \quad H_1: \theta_j \neq 0 \quad EP: T_0 = \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)} \sim t_{100-K}$$

$$H_0: \theta_i = 0 \quad H_1: \theta_i \neq 0 \quad EP: T_0 = \frac{\hat{\theta}_i}{SE(\hat{\theta}_i)} \sim t_{100-K}$$

$$CR: \text{Val } P = P(|t_{100-K}| > |T_0|)$$

Validación de supuestos sobre errores de ajuste

Se desea probar sobre a_t : 1) Ruido blanco
2) Normalidad

1) Para probar RB: $\rho(k) = \text{Corr}(a_t, a_{t+k})$; $\rho_{kk} = \text{Corr}(a_t, a_{t+k} | a_{t+1}, a_{t+2}, \dots, a_{t+k-1})$

Test ACF - PACE

Para $K = 1, 2, \dots, 25$ se prueba individualmente:

ACF

PACF

$$H_0: \rho(k) = 0$$

$$H_0: \rho_{kk} = 0$$

$$H_1: \rho(k) \neq 0$$

$$H_1: \rho_{kk} \neq 0$$

$$EP: \hat{\rho}(k) \overset{aprox}{\sim} N(0, \frac{1}{100})$$

$$EP: \hat{\rho}_{kk} \overset{aprox}{\sim} N(0, \frac{1}{100})$$

$$CR: |\hat{\rho}(k)| > \frac{z}{\sqrt{100}}$$

$$CR: |\hat{\rho}_{kk}| > \frac{z}{\sqrt{100}}$$

$$\text{con } \hat{\rho}(k) = \frac{\sum_{t=1}^{100-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^{100} \hat{a}_t^2}$$

Test Ljung-Box

Con $m = 6, 12, 18, 24$

$$H_0: \rho(1) = \rho(2) = \rho(3) = \dots = \rho(m) = 0$$

H_1 : al menos un $\rho(k) \neq 0$ para $k = 1, 2, 3, \dots, m$.

$$EP: Q_{LB} = 100 * 98 \sum_{k=1}^m \frac{\hat{\rho}^2(k)}{100-k} \approx \chi^2_m$$

$$CR: \text{val } P = P(\chi^2_m \geq Q_{LB})$$

- datos 1: Analizando las gráficas de residuales se observa media cero y aunque podría haber problemas de varianza no es tan grave como para rechazar el supuesto de varianza constante.

• En los test ACF, PACE y Ljung-Box no se rechaza en ninguna prueba y por lo tanto no se rechaza RB.

• Como no se rechaza RB se evalúa normalidad. En la gráfica de normalidad se ven en los extremos los valores atípicos que se observan en las gráficas de residuales. Hay algo de desviación en la normalidad aunque no es suficiente para rechazar el supuesto. Entre los dos modelos no se identifica cuál se acerca más a cumplir el supuesto.

- datos2: • Se observa media cero y varianza constante, se evidencian problemas de carencia de ajuste.

• Al igual que en datos1 no se rechaza RB en test ACF, PACF y Ljung-Box.

• Se rechaza normalidad, la nube de puntos tiene concavidad hacia abajo por lo que hay asimetría, una distribución de cola a la derecha.

- datos3: • Se observa media cero y varianza constante.

• Nuevamente no hay rechazos contra RB en test ACF, PACF y Ljung-Box

• Nuevamente se rechaza normalidad, los puntos no siguen apropiadamente la recta teórica de normalidad.

Resumen

	¿A _t RB?	¿A _t Normales?	¿modelo válido?
datos1: AR(2)	si	si	si
ARMA(2,2)	si	si	si
datos2: MA(2)	si	NO	NO
ARMA(3,2)	si	NO	NO
datos3: ARMA(1,1)	si	NO	NO
ARMA(2,2)	si	NO	NO

Como los modelos para cada caso son ambos válidos o inválidos entonces para elegir los mejores modelos se analizan las medidas de ajuste y pronóstico. En este caso como no

Se aplicó validación cruzada se elige sólo con AIC y BIC

- datos1: el mejor modelo es el AR(2) por parsimonia
- datos2: tanto por AIC como por BIC el mejor modelo es el MA(2)
- datos3: el mejor modelo es el ARMA(1,1) por parsimonia.

Verificación de estacionariedad e invertibilidad

1) Sobre el AR(2) ajustado a Z_t (datos1)

- ¿Es invertible? \rightarrow Sí, todo AR(p) es invertible por definición.
- ¿Es estacionario? \rightarrow Será estacionario si las raíces del polinomio $\Phi_z(B)$ tienen módulo mayor a 1.

Polinomio estimado $\hat{\Phi}_2(B) = 1 - \hat{\alpha}_1 B - \hat{\alpha}_2 B^2 = 1 - 0,691882B + 0,479471B^2$
y sus raíces son la solución de:

$$1 - 0,691882X + 0,479471X^2 = 0$$

$$X = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{0,691882 \pm \sqrt{(-0,691882)^2 - 4(0,479471)(1)}}{2(0,479471)}$$

$$X = \frac{0,691882 \pm \sqrt{-1,439183298}}{0,958942} = \frac{0,691882 \pm \sqrt{1,439183298}}{0,958942} \quad i; \quad i = \sqrt{-1}$$

$\underbrace{\hspace{1.5cm}}_a \quad \underbrace{\hspace{1.5cm}}_b$

Las raíces son complejas y tienen módulo: $|X| = \sqrt{a^2 + b^2}$

$$|X| = \sqrt{\left(\frac{0,691882}{0,958942}\right)^2 + \left(\frac{\sqrt{1,439183298}}{0,958942}\right)^2} = 1,4442 > 1$$

Es decir, ambas raíces tienen módulo $> 1 \therefore$ el $AR(z)$ ajustado a Z_t es un proceso estacionario.

2) Sobre el $MA(z)$ ajustado a W_t (datos 2)

- ¿Es estacionario? \rightarrow Sí, todo $MA(q)$ es estacionario por definición.
- ¿Es invertible? \rightarrow Será invertible si las raíces del polinomio $\Theta_z(B)$ tienen módulo mayor a 1.

Polinomio estimado $\hat{\Theta}_z(B) = 1 + \hat{\Theta}_1 B + \hat{\Theta}_2 B^2 = 1 - 0,394121B + 0,565021B^2$
y sus raíces son solución de:

$$1 - 0,394121x + 0,565021x^2 = 0$$

$$x = \frac{0,394121 \pm \sqrt{(-0,394121)^2 - 4(0,565021)(1)}}{2(0,565021)} = \frac{0,394121 \pm \sqrt{-2,104752637}}{1,130042}$$

$$x = \underbrace{\frac{0,394121}{1,130042}}_a \pm \underbrace{\frac{\sqrt{2,104752637}}{1,130042}}_b i$$

$$\text{módulo} = \sqrt{a^2 + b^2} = \sqrt{\left(\frac{0,394121}{1,130042}\right)^2 + \left(\frac{\sqrt{2,104752637}}{1,130042}\right)^2} = 1,33036 > 1$$

Como ambas raíces tienen módulo mayor a 1, el $MA(z)$ ajustado a W_t es un proceso invertible.

3) Sobre el ARMA(1,1) ajustado a X_t (datos3)

- Es estacionario ya que $|\hat{\phi}| = |0,489672| < 1$
- Es invertible ya que $|\hat{\theta}| = |1 - 0,791686| = 0,791686 < 1$

Pronósticos para $L=2$

$n=100$

1) modelo AR(2) ajustado a Z_t (datos1)

$$\hat{Z}_{100}(L) = 0,691882 \hat{Z}_{100}(L-1) - 0,479471 \hat{Z}_{100}(L-2)$$

$$\text{Con } \hat{Z}_{100}(L-j) = \begin{cases} \text{Valor observado } Z_{100+L-j} & \text{si } L-j \leq 0 \\ \text{Pronóstico para } L-j \text{ periodos después de } t=100 & \text{si } L-j > 0 \end{cases}$$

Con $L=2$

$$\hat{Z}_{100}(2) = 0,691882 \hat{Z}_{100}(1) - 0,479471 \hat{Z}_{100}(0) \quad (1)$$

$$\text{Con } \hat{Z}_{100}(0) \text{ valor observado en } t=100 \rightarrow -0,5485614 \quad (2)$$

y con $\hat{Z}_{100}(1)$ pronóstico un período después de $t=100$

$$\hat{Z}_{100}(1) = 0,691882 \hat{Z}_{100}(0) - 0,479471 \hat{Z}_{100}(-1)$$

$$\hat{Z}_{100}(1) = 0,691882(-0,5485614) - 0,479471(1,747155)$$

$$\hat{Z}_{100}(1) = -1,21725 \quad (3)$$

(2) y (3) en (1)

$$\hat{Z}_{100}(2) = 0,691882(-1,21725) - 0,479471(-0,5485614) = \boxed{-0,5791741}$$

2) Modelo MA(2) ajustado a W_t (datos 2):

$$\hat{W}_{100}(L) = -0,394121 \hat{a}_{100}(L-1) + 0,565021 \hat{a}_{100}(L-2)$$

$$\text{Con } \hat{a}_{100}(L-j) \begin{cases} \text{Residuo } \hat{a}_{100+L-j} & \text{si } L-j \leq 0 \\ 0 & \text{si } L-j > 0 \end{cases}$$

$$\text{Con } L=2$$

$$\hat{W}_{100}(2) = -0,394121 \hat{a}_{100}(1) + 0,565021 \hat{a}_{100}(0)$$

$$\hat{W}_{100}(2) = 0,565021 (0,7759357) = \boxed{0,43842}$$

3) Modelo ARMA(1,1) ajustado a X_t (datos 3):

$$\hat{X}_{100}(L) = 0,489672 \hat{X}_{100}(L-1) - 0,791686 \hat{a}_{100}(L-1)$$

$$\text{Con } \hat{X}_{100}(L-j) \begin{cases} \text{Valor observado } X_{100+L-j} & \text{si } L-j \leq 0 \\ \text{Pronóstico para } L-j \text{ periodos} & \text{si } L-j > 0 \\ \text{después de } L=100 & \end{cases}$$

$$\text{Con } \hat{a}_{100}(L-j) \begin{cases} \text{Residuo } \hat{a}_{100+L-j} & \text{si } L-j \leq 0 \\ 0 & \text{si } L-j > 0 \end{cases}$$

$$\text{Para } L=2$$

$$\hat{X}_{100}(2) = 0,489672 \hat{X}_{100}(1) - 0,791686 \hat{a}_{100}(1)$$

$$\hat{X}_{100}(z) = 0,489672 \hat{X}_{100}(1) \quad (1)$$

$$\hat{X}_{100}(1) = 0,489672 \hat{X}_{100}(0) - 0,791686 \hat{a}_{100}(0)$$

$$\hat{X}_{100}(1) = 0,489672(0,9384119) - 0,791686(0,7805878)$$

$$\hat{X}_{100}(1) = -0,1584664 \quad (2)$$

② en ①

$$\hat{X}_{100}(z) = 0,489672(-0,1584664) = \boxed{-0,07759656}$$