

Análisis de series temporales

6. Metodología Box-Jenkins

5i hcf. 'dfcZ"Ernest Pons : UbUg

Grado en Estadística

6.0. Introducción

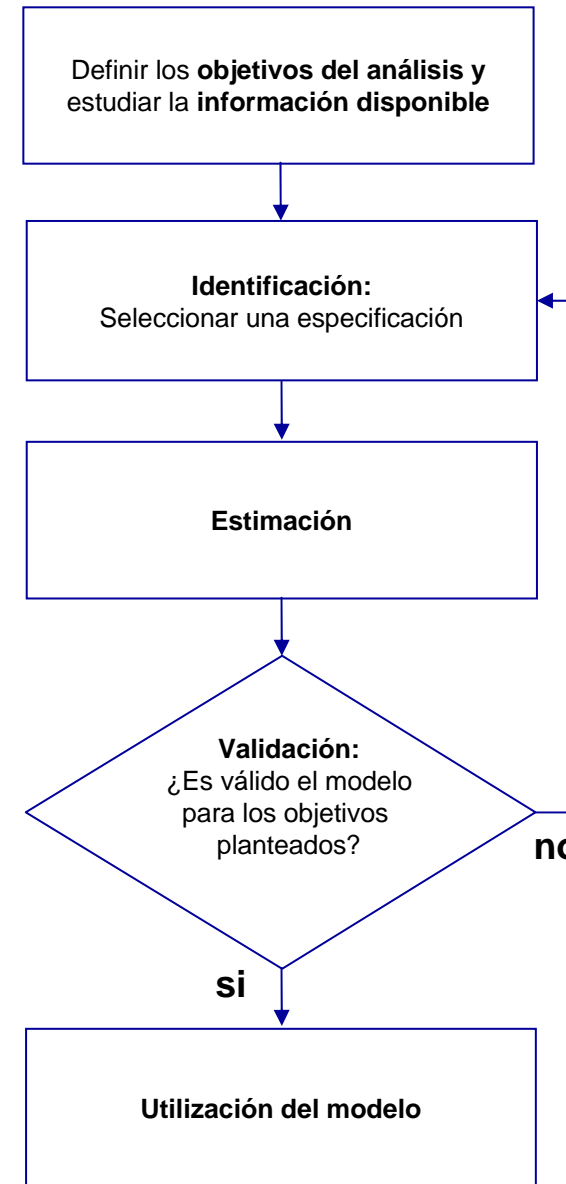
En las lecciones anteriores se ha seguido un enfoque donde, conociendo previamente el proceso estocástico (su especificación y los valores de los parámetros) hemos analizado sus características:

- Estacionariedad / Causalidad / Invertibilidad
- Esperanza
- Varianza y autocovarianzas
- FAS y FAP

A partir de ahora seguiremos un enfoque diferente: a partir de una serie temporal nos propondremos identificar el modelo **$SARIMA(p,d,q)(P,D,Q)_s$** que ha podido generar esta serie, para obtener predicciones de los valores futuros.

6.0. Introducción

La metodología planteada per Box y Jenkins (1970) propone combinar diferentes instrumentos en una pauta sistemática para construir modelos.



Recordemos los modelos básicos (ruido blanco)

Un proceso ruido blanco representa la evolución de una variable que

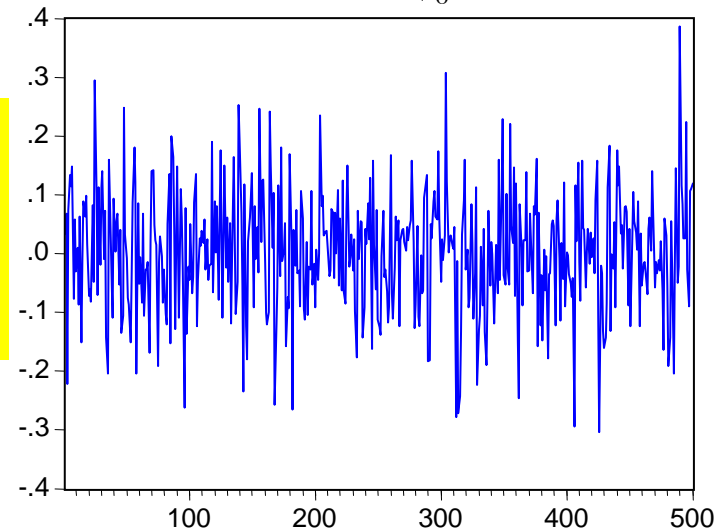
- oscila alrededor de una media constante,
- con una volatilidad constante,
- con una “historia” que no aporta información útil para predecir el futuro.

Podemos representar esta variable como: $y_t = \mu + \varepsilon_t$

$$E(y_t) = \mu \quad [E(\varepsilon_t) = 0] \quad E(\tilde{y}_t^2) = \sigma_y^2 = \gamma_0 = \sigma_\varepsilon^2 \quad \rho_k = \frac{\gamma_k}{\gamma_0} = 0 \quad ; \quad k \geq 1$$

Gráfico: 500 observaciones simuladas de un proceso ruido blanco:

$$y_t = \varepsilon_t ; \varepsilon_t \sim \text{iid } N(0, .01)$$

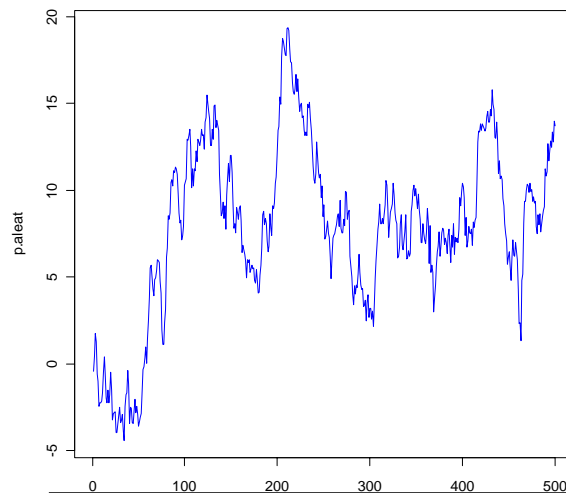


Recordemos los modelos básicos (camino aleatorio)

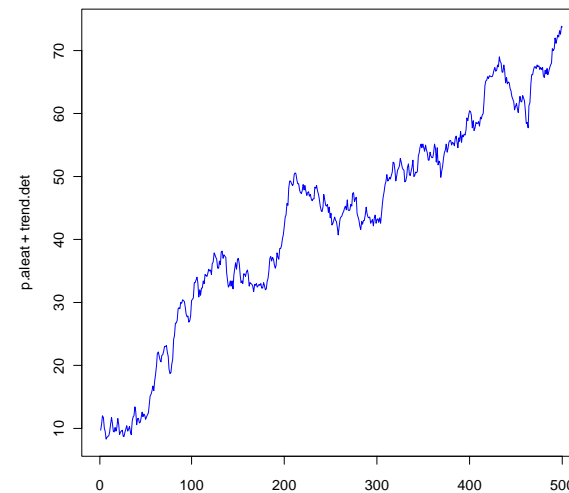
Un camino aleatorio representa una variable donde los incrementos son ruido blanco y, por tanto, imprevisibles. Es decir:

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

Gráfico: Simulación de dos caminos aleatorios, con y sin “deriva”. La característica fundamental de este tipo de proceso es que no presenta una media estable.

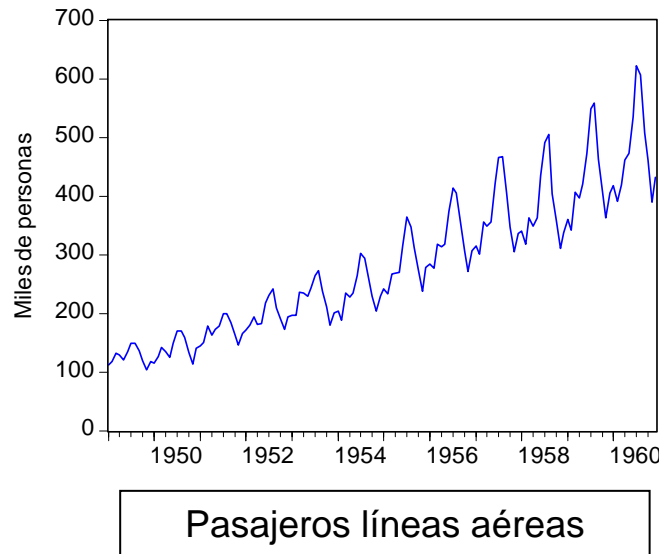


Camino aleatorio sin deriva



Camino aleatorio con deriva

Atención



Muchas series temporales “reales” presentan:

- tendencia,
- oscilaciones estacionales,
- una variabilidad creciente,
- efectos deterministas.

Pero los modelos ARMA describen procesos puramente estocásticos, no estacionales, con media y varianza constante. Por tanto, para la aplicación con series reales necesitamos más herramientas:

- **Transformaciones de datos** diseñadas para estabilizar la media y la varianza.
- **Modelos SARIMA.**

6.1. Identificación

Objetivo:

Identificar un modelo **$SARIMA(p,d,q)(P,D,Q)_s$** que, de forma verosímil, haya podido generar los valores observados.

Etapas:

1) Aplicar las transformaciones que sean necesarias para garantizar la estacionariedad de segundo orden:

- Estacionariedad en media.
- Estacionariedad en varianza.
- Estacionariedad autocovarianzas en la parte regular.
- Estacionariedad autocovarianzas en la parte estacional.

2) Identificar el modelo ARMA parte regular (p,q)

3) Identificar el modelo ARMA parte estacional (P,Q)

6.1. Identificación

Para modelizar cualquier serie temporal observada (Y_t), podemos proponer un modelo $ARIMA(p,d,q) \times (P,D,Q)_S$ una vez eliminada la componente determinista (Z_t):

$$\phi(B)\Phi_s(B)(1-B)^d(1-B^s)^D(y_t - Z_t)^{\lambda,m} = \delta + \theta(B)\Theta_s(B)\varepsilon_t$$

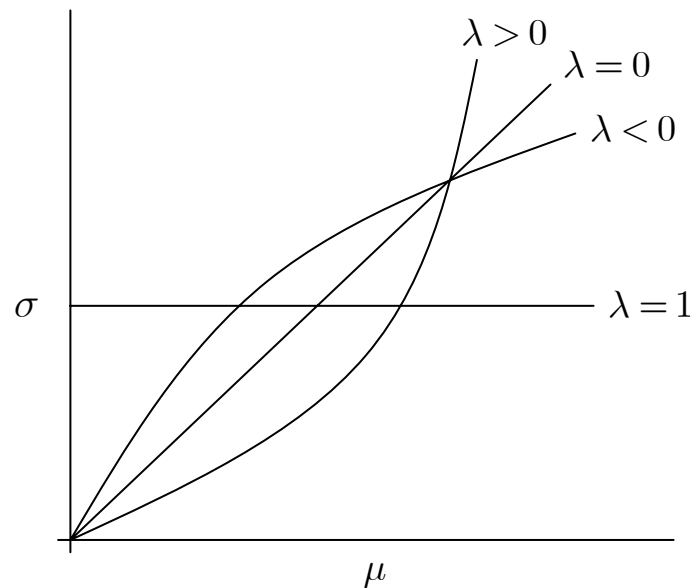
donde:

- λ , m definen la transformación de Box-Cox
- S es el orden de estacionalidad
- d y D definen el orden de diferenciación regular y estacional
- p y q son los órdenes autoregresivos y media móvil regular
- P y Q son los órdenes autoregresivos y media móvil estacional

Transformaciones (1): Box-Cox

Para estabilizar la varianza una opción es la transformación de Box-Cox:

$$y_t^{\lambda, m} = \frac{(y_t + m)^\lambda - 1}{\lambda}$$

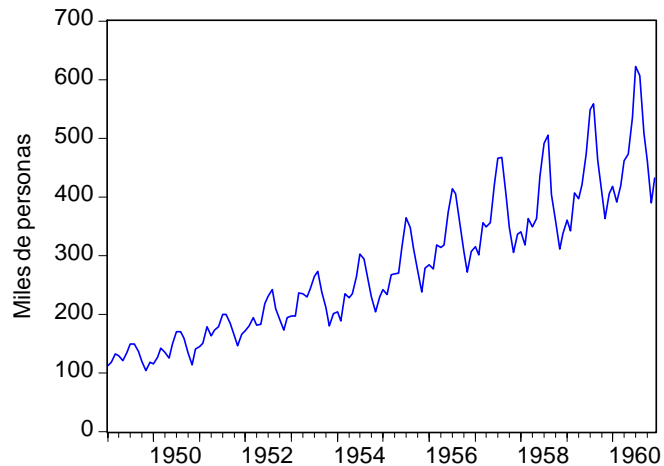


Cada transformación está caracterizada por un parámetro λ . Si es necesario se puede aplicar también un cambio de origen (parámetro m) cuando la transformación requiere que todos los valores sean positivos.

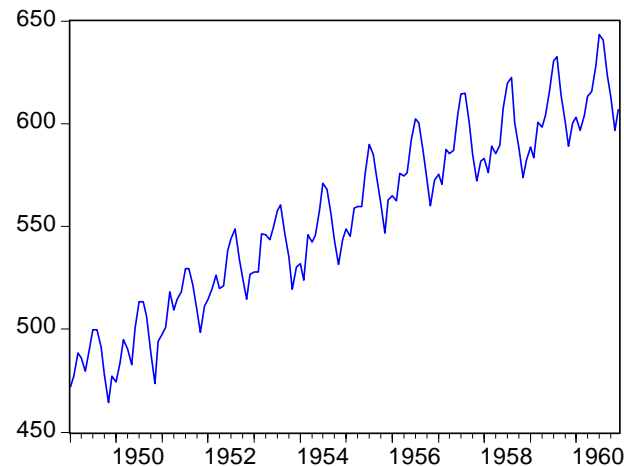
Para escoger la transformación se puede usar el gráfico media - desviación típica para varias submuestras.

Lo más habitual es encontrar series cuya variabilidad crece de forma aproximadamente lineal de manera que a menudo se puede utilizar la transformación logarítmica ($\lambda=0$)

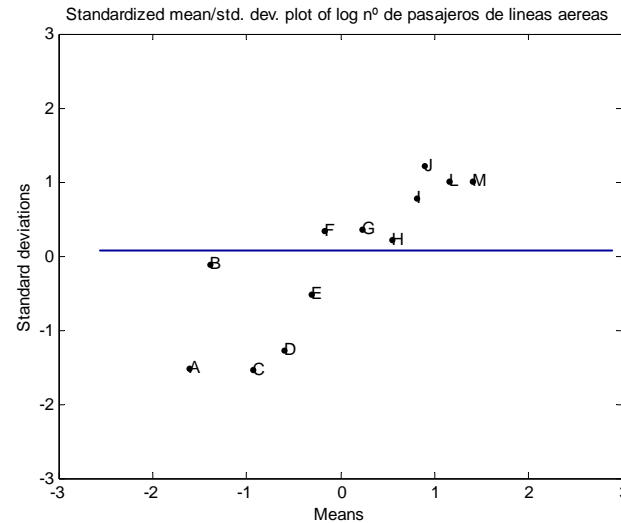
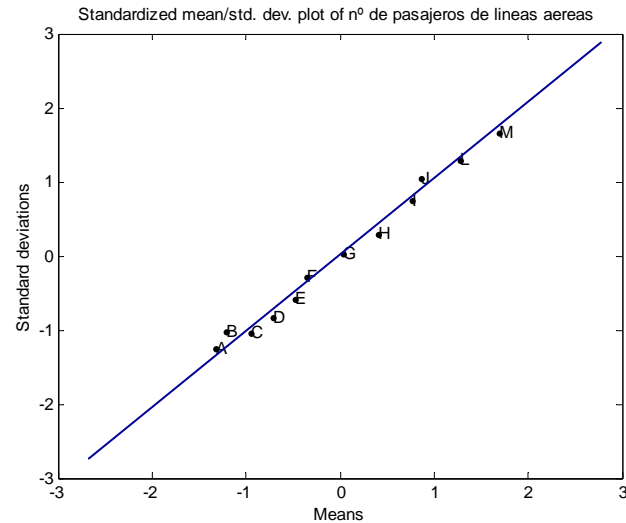
Transformaciones (1): Box-Cox



Pasajeros líneas aéreas



Log (Pasajeros líneas aéreas)



Los gráficos muestran que:

- la volatilidad crece de forma lineal con el nivel de la serie.
- tras la transformación la volatilidad es aproximadamente constante.

Transformaciones (2): Diferencia

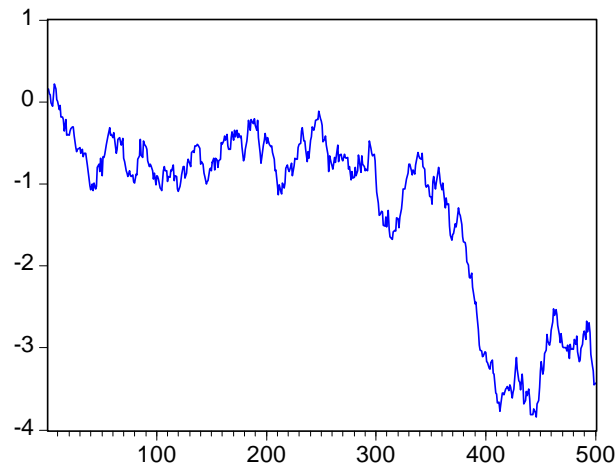
A menudo la tendencia de una serie temporal puede eliminarse tomando diferencias de los datos. Diremos en este caso que la serie es “integrada de orden 1” si la primera diferencia:

$$z_t = y_t - y_{t-1}$$

es estacionaria en media y varianza.

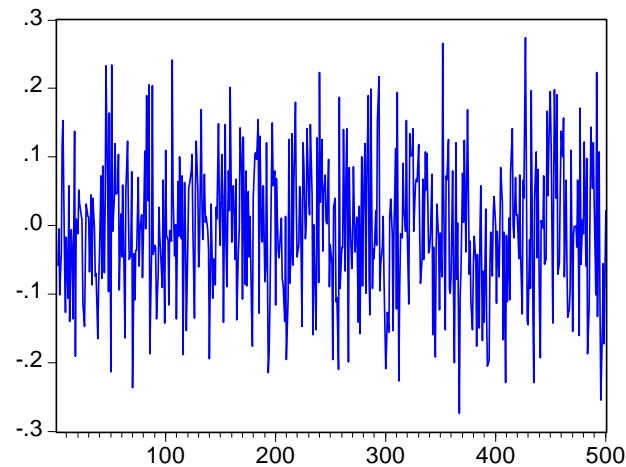
El gráfico muestra una serie generada según un camino aleatorio y su diferencia

$$y_t = y_{t-1} + \varepsilon_t ; \varepsilon_t \sim \text{iid } N(0, .01)$$



Camino aleatorio

$$z_t = y_t - y_{t-1}$$



Primera diferencia

Algunas series necesitan una diferencia adicional para conseguir una media estable.

Transformaciones: Interpretación

La primera diferencia del logaritmo de una serie es una aproximación a la tasa de crecimiento porcentual ya que si $y_t = (1 + \alpha_t)y_{t-1}$

podemos comprobar que: $\ln y_t - \ln y_{t-1} = \ln(1 + \alpha_t) \simeq \alpha_t$

A diferencia de la tasa de crecimiento “estándar”, la tasa logarítmica tiene la ventaja de ser aditiva, es decir:

$$\ln y_T - \ln y_0 = \sum_{t=1}^T (\ln y_t - \ln y_{t-1}) = \sum_{t=1}^T \ln(1 + \alpha_t)$$

En ocasiones también puede calcular-se la tasa de variación acumulada en “s” periodos:

$$\ln y_t - \ln y_{t-s}$$

Transformaciones: Interpretación

Transformación	Interpretación
$z_t = y_t - y_{t-1}$	Variaciones de la variable y_t
$z_t = \ln y_t - \ln y_{t-1}$	Tasa logarítmica (en tanto por u) de variación entre un periodo y el siguiente (indicador de crecimiento).
$w_t = z_t - z_{t-1}; z_t = y_t - y_{t-1}$	Cambio en la tasa logarítmica de variación entre un periodo y el siguiente (indicador de aceleración en el crecimiento).
$z_t = \ln y_t - \ln y_{t-s}$	Tasa logarítmica de variación acumulada en s periodos. Si s es el periodo estacional, se trata de un indicador de crecimiento acumulado en un ciclo estacional completo.

Resumen fase identificación

Parámetro	Instrumentos para identificación	Observaciones
m, λ	<ul style="list-style-type: none"> • Gráfico de la serie temporal • Gráfico media – desviación típica 	El objetivo es conseguir que la variabilidad de los datos sea independiente del nivel. La transformación más habitual es la logarítmica
d , orden de diferenciación	<ul style="list-style-type: none"> • Gráfico de la serie temporal • ACF (detectar decrecimiento lento y lineal) 	El objetivo es conseguir un perfil estacionario en media y varianza.
Constante	<ul style="list-style-type: none"> • Media de la serie analizada (niveles o diferenciada) • Desviación típica de la media 	Para decidir si el modelo debe incluir un término constante.
p , orden del polinomio AR	<ul style="list-style-type: none"> • FAP de orden p • FAS infinita 	La FAP tiene p valores diferentes de cero Un proceso AR finito equivale a un $MA(\infty)$
q , orden del polinomio MA	<ul style="list-style-type: none"> • FAS de orden q • FAP infinita 	La FAS tiene q valores diferentes de cero Un proceso MA finito e invertible equivale a un $AR(\infty)$

Resumen fase identificación

Parámetro	Instrumentos para identificación	Observaciones
d, orden de diferenciación	<ul style="list-style-type: none"> • Raíces de los polinomios AR y MA 	<ul style="list-style-type: none"> • Una raíz próxima a la unidad en la parte AR significa que debemos diferenciar • Una raíz próxima a la unidad en la parte MA significa que debemos eliminar una diferencia
	<ul style="list-style-type: none"> • Gráfico de la serie de residuos 	Comprobar la no existencia de rachas de valores positivos o negativos
Constante	<ul style="list-style-type: none"> • Media muestral de los residuos • Desviación típica de la media 	Si la media es diferente de cero debemos añadir un término constante
p y q	<ul style="list-style-type: none"> • Test de significación de los parámetros estimados 	Permite eliminar parámetros no significativos
	<ul style="list-style-type: none"> • FAS y FAP de los residuos (blanquear) 	Permite detectar estructuras de autocorrelación no modelizadas
	<ul style="list-style-type: none"> • Test Q (Box-Pierce, Ljung-Box) 	Permite contrastar la hipótesis de que todos los coeficientes de autocorrelación son nulos
	<ul style="list-style-type: none"> • Correlaciones elevadas entre los parámetros estimados 	Puede ser un síntoma de sobreparametrización
	<ul style="list-style-type: none"> • "Sobreajuste" 	Debemos analizar si los parámetros AR y MA realmente mejoran la "calidad" del modelo

6.2. Estimación

Podemos estimar los coeficientes de la **función de autocorrelación simple** (FAS) usando el siguiente estimador:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad \text{donde} \quad \hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}) , \quad k = 1, 2, \dots$$

Para realizar inferencia sobre la función de autocorrelación simple (FAS) podemos utilizar los siguientes resultados:

- Para muestras grandes se cumple: $\text{s.e.}(\hat{\rho}_k) \simeq T^{-1/2}$
- Veremos como podemos contrastar: $H_0 : \rho_1 = \rho_2 = \rho_3 = \dots = \rho_K = 0$

6.2. Estimación

Podemos estimar la **función de autocorrelación parcial** (FAP) a partir de las siguientes ecuaciones:

$$(y_t - \bar{y}) = \hat{\phi}_{k1}(y_{t-1} - \bar{y}) + \hat{\phi}_{k2}(y_{t-2} - \bar{y}) + \dots + \hat{\phi}_{kk}(y_{t-k} - \bar{y}) + \hat{\varepsilon}_{kt} ; k = 1, 2, \dots$$

Se puede comprobar que:

$$\hat{\phi}_{kk} = \frac{|A_k|}{|B_k|}$$

donde:

$$A_k = \begin{bmatrix} 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{k-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \dots & \hat{\rho}_{k-1} & \hat{\rho}_2 \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \dots & \hat{\rho}_1 & \hat{\rho}_k \end{bmatrix} \quad B_k = \begin{bmatrix} 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{k-2} & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \dots & \hat{\rho}_{k-3} & \hat{\rho}_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \dots & \hat{\rho}_1 & 1 \end{bmatrix}$$

6.2. Estimación

Vamos a considerar tres métodos de estimación de modelos:

- 1) Método de estimación de Yule-Walker (YW) para modelos $AR(p)$.
- 2) Mínimos cuadrados no lineales (MCNL) para modelos $ARMA(p,q)$.
- 3) Máxima verosimilitud (MV) para modelos $ARMA(p,q)$.

6.2. Estimación

1) Método de estimación de Yule-Walker (YW)

Sólo permite estimar modelos AR(p) y se basa en las denominadas ecuaciones de Yule-Walker:

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots \phi_k \rho_{k-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots \phi_k \rho_{k-2} \\ &\dots\dots\dots \\ \rho_k &= \phi_{k-1} + \phi_2 \rho_{k-2} + \dots \phi_k\end{aligned}$$

En forma matricial:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_k \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_k \end{bmatrix}$$

Que podemos resolver utilizando, por ejemplo, el método de Cramer

6.2. Estimación

2) Mínimos cuadrados no lineales (MCNL)

El método se puede ilustra con un ejemplo (modelo MA(1)). En este caso tenemos:

$$\varepsilon_t = x_t + \theta \varepsilon_{t-1}$$

que, a través de sucesivas substituciones:

$$\varepsilon_t = x_t + \theta x_{t-1} + \theta^2 x_{t-2} + \dots + \theta^{t-1} x_1 + \theta \varepsilon_0$$

Utilizando la aproximación de Taylor podemos “linealizar” el modelo:

$$\varepsilon_t \approx \varepsilon_t^0 - (\theta - \theta^0) \left(- \frac{\partial \varepsilon_t}{\partial \theta} \right)_{\theta=\theta^0}$$

Para la estimación debemos evaluar las derivadas para lo cual tenemos dos opciones:

- De forma analítica:
$$\frac{\partial \varepsilon_t}{\partial \theta} = \theta \frac{\partial \varepsilon_{t-1}}{\partial \theta} + \varepsilon_{t-1}$$
- De forma numérica: métodos recursivos (Newton-Raphson, etc..)

6.2. Estimación

3) Máxima verosimilitud

Suponiendo normalidad y estacionariedad, la verosimilitud asociada a una muestra “finita” es:

$$\begin{aligned}\log L = \log f(x_1, \dots, x_T) &= \sum_{t=2}^T \log f(x_t | x_1, \dots, x_{t-1}) + f(x_1) = \\ &= -\frac{T}{2} \log(2\pi) - \frac{(T-1)}{2} \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T (x_t - E(x_t | x_1, \dots, x_{t-1}))^2 - \frac{1}{2} \log \sigma^2 - \frac{(x_1 - \mu)^2}{2\sigma^2}\end{aligned}$$

Por ejemplo, en el caso de un AR(1) tenemos:

$$\log L = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \sigma_\varepsilon^2 + \frac{1}{2} (1 - \phi^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T (x_t - c - \phi x_{t-1})^2 - \frac{\left(x_1 - \frac{c}{1-\phi}\right)^2}{2\left(\frac{\sigma_\varepsilon^2}{1-\phi^2}\right)}$$

6.3. Validación

Condiciones a comprobar:

- 1) Comprobar que los datos se distribuyen según una **normal**.
- 2) Todos los **coeficientes** del modelo deben resultar estadísticamente significativos.
- 3) El modelo debe cumplir con las condiciones de **estacionariedad** e **invertibilidad**.
- 4) Los **residuos** de la estimación del modelo deben comportarse según un ruido blanco.
- 5) El modelo propuesto debe ser **estable**.

6.3. Validación

1) Normalidad

La hipótesis de normalidad se puede contrastar a través del test de Jarque-Bera:

$$JB = n \left(\frac{CA^2}{6} + \frac{(CK - 3)^2}{24} \right) \underset{H_0}{\sim} \chi^2_2$$

basado en los coeficientes de asimetría y apuntamiento:

$$CA = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}_X} \right)^3$$

$$CK = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}_X} \right)^4$$

6.3. Validación

4) Residuos

Para comprobar si los **residuos** son ruido blanco podemos utilizar los siguientes estadísticos:

Box-Pierce (1970):

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_m = 0$$

$$BP = T \sum_{k=1}^m \gamma_k^2 \stackrel{H_0}{\sim} \chi_m^2$$

Ljung-Box (1978):

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_m = 0$$

$$LB = T(T+2) \sum_{k=1}^m \frac{\gamma_k^2}{(T-k)} \stackrel{H_0}{\sim} \chi_m^2$$

6.4. Predicción

$$\Phi(B)X_t = \Theta(B)\varepsilon_t$$

$$X_t = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t = \Psi(B)\varepsilon_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_1\varepsilon_{t-2} + \dots$$

Objetivo: con información hasta el momento T , predecir hasta $T+k$

Supongamos que tenemos datos de una serie: $X_1, X_2, X_3, \dots, X_T$

$$X_{T+k} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{T+k-j} = \sum_{j=0}^{k-1} \psi_j \varepsilon_{T+k-j} + \sum_{j=k}^{\infty} \psi_j \varepsilon_{T+k-j}$$

Por lo tanto, usaremos predicciones que sean lineales:

$$\hat{X}_T(k) = \psi_k^* \varepsilon_T + \psi_{k+1}^* \varepsilon_{T-1} + \psi_{k+2}^* \varepsilon_{T-2} + \dots$$

6.4. Predicción

Criterio: Minimizar el error cuadrático medio

$$\min_{\psi^*} E[X_{T+k} - \hat{X}_T(k)]^2$$

Sabemos que $E[X_{T+k} - \hat{X}_T(k)]^2 = \sigma_\varepsilon^2 \sum_{j=0}^{k-1} \psi_j^2 + \sigma_\varepsilon^2 \sum_{j=0}^{\infty} (\psi_{T+j} - \psi_{T+j}^*)^2$

$$\frac{\partial E[X_{T+k} - \hat{X}_T(k)]^2}{\partial \psi^*} = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} 2(\psi_{T+j} - \psi_{T+j}^*)(-1) = 0 \Rightarrow \psi_{T+j}^* = \psi_{T+j}$$

Por tanto: $\hat{X}_T(k) = \psi_k \varepsilon_T + \psi_{k+1} \varepsilon_{T-1} + \psi_{k+2} \varepsilon_{T-2} + \dots$

Una interpretación alternativa

$$X_{T+k} = \sum_{j=0}^{k-1} \psi_j \varepsilon_{T+k-j} + \sum_{j=0}^{\infty} \psi_{k+j} \varepsilon_{T-j}$$

Per condicionamiento:

$$E(X_{T+k} | X_T, X_{T-1}, \dots) = \sum_{j=0}^{\infty} \psi_{k+j} \varepsilon_{T-j} = \psi_k \varepsilon_T + \psi_{k+1} \varepsilon_{T-1} + \psi_{k+2} \varepsilon_{T-2} + \dots$$

Por lo tanto:

$$\hat{X}_T(k) = E(X_{T+k} | X_T, X_{T-1}, \dots)$$

En síntesis, para la función de pérdida cuadrática, el predictor óptimo es la esperanza condicionada, donde el conjunto de condicionamiento esta formado por la información pasada sobre la propia serie temporal.

Propiedades del error de predicción

$$X_{T+k} = \sum_{j=0}^{k-1} \psi_j \varepsilon_{T+k-j} + \sum_{j=0}^{\infty} \psi_{k+j} \varepsilon_{T-j}$$

Como: $X_{T+k} = \hat{X}_T(k) + e_T(k)$

El error de predicción es: $e_T(k) = X_{T+k} - \hat{X}_T(k) = \sum_{j=0}^{k-1} \psi_j \varepsilon_{T+k-j}$

Propiedades:

- 1) La predicción $\hat{X}_T(k)$ y el error de predicción $e_T(k)$ están incorrelacionados
- 2) No tiene sesgo: $E(e_T(k)) = 0$
- 3) La varianza del error de predicción es $Var(e_T(k)) = \sigma_\varepsilon^2 \sum_{j=0}^{k-1} \psi_j^2$
- 4) Los errores referidos a diversos periodos $e_T(k)$ $k > 1$ están correlacionados entre sí

Propiedades del error de predicción

Si nos fijamos ahora en los sucesivos errores sobre la predicción a un período

$$e_T(1), e_{T+1}(1), e_{T+2}(1), \dots$$

son incorrelacionados porque $e_T(1) = X_{T+1} - \hat{X}_T(1) = \varepsilon_{T+1}$

Pero, en general, cuando calculamos predicciones para un mismo período $k > 1$ son correlacionados:

$$e_T(k) = X_{T+k} - \hat{X}_T(k) = \varepsilon_{T+k} + \psi_1 \varepsilon_{T+k-1} + \dots + \psi_{k-1} \varepsilon_{T+1}$$

$$e_{T-j}(k) = X_{T-j+k} - \hat{X}_{T-j}(k) = \varepsilon_{T-j+k} + \psi_1 \varepsilon_{T-j+k-1} + \dots + \psi_{k-1} \varepsilon_{T-j+1}$$

$$\text{cov}(e_T(k), e_{T-j}(k)) = \sum_{i=T+1}^{T+k-j} \psi_{i-T} \psi_{i-T-j} \sigma^2 \quad j < k$$

Predicción con un modelo AR(1)

$$X_t = \phi X_{t-1} + \varepsilon_t$$

Para k=1: $X_{T+1} = \phi X_T + \varepsilon_{T+1} \Rightarrow E(X_{T+1} | I_T) = \phi X_T$

Para k=2: $X_{T+2} = \phi X_{T+1} + \varepsilon_{T+2} \Rightarrow E(X_{T+2} | I_T) = \phi^2 X_T$

Para cualquier k: $\hat{X}_T(k) = \phi^k X_T$

Podemos observar que los coeficientes decrecen de forma geométrica

Predicción con un modelo AR(p)

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

Para k=1: $X_{T+1} = \phi_1 X_T + \phi_2 X_{T-1} + \dots + \phi_p X_{T-p+1} + \varepsilon_{T+1}$

$$\Rightarrow \hat{X}_T(1) = E(X_{T+1} | I_T) = \phi_1 X_T + \phi_2 X_{T-1} + \dots + \phi_p X_{T-p+1}$$

Para k=2: $X_{T+2} = \phi_1 X_{T+1} + \phi_2 X_T + \dots + \phi_p X_{T-p+2} + \varepsilon_{T+2}$

$$\Rightarrow \hat{X}_T(2) = E(X_{T+2} | I_T) = \phi_1 \hat{X}_T(1) + \phi_2 X_T + \dots + \phi_p X_{T-p+2}$$

Para cualquier k: $\hat{X}_T(k) = \phi_1 \hat{X}_T(k-1) + \phi_2 \hat{X}_T(k-2) + \dots + \phi_p \hat{X}_T(k-p)$

Necesitamos calcular las predicciones previas sobre k-1, k-2, etc..

Predicción con un modelo MA(1)

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

Para $k=1$: $X_{T+1} = \varepsilon_{T+1} - \theta \varepsilon_n \Rightarrow \hat{X}_T(1) = E(X_{T+1} | I_T) = E(X_{T+1}) = -\theta \varepsilon_n$

Para $k=2$: $X_{T+2} = \varepsilon_{T+2} - \theta \varepsilon_{T+1} \Rightarrow \hat{X}_T(2) = 0$

Para cualquier $k > 2$: $\hat{X}_T(k) = 0$

Es decir, la predicción es la media del proceso

Predicción con un modelo MA(q)

$$X_t = 1 - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

$$\hat{X}_T(k) = E(X_{T+k} \mid I_T) = \begin{cases} -\theta_k \varepsilon_T - \theta_{k+1} \varepsilon_{T-1} - \theta_{k+2} \varepsilon_{T-2} - \dots - \theta_q \varepsilon_{T-q+k} & k \leq q \\ 0 & k > q \end{cases}$$

$$\text{on } \varepsilon_T = \frac{1}{1 - \theta_1 B - \dots - \theta_q B^q} X_T$$

Predicción con un modelo ARMA(1,1)

$$(1 - \phi B)X_t = (1 - \theta B)\varepsilon_t$$

$$X_{T+k} = \phi X_{T+k-1} + \varepsilon_{T+k} - \theta \varepsilon_{T+k-1}$$

Para $k=1$: $\hat{X}_T(1) = \phi X_T - \theta \varepsilon_T$ on $\varepsilon_T = \frac{1 - \phi B}{1 - \theta B} X_T$

Para $k=2$: $\hat{X}_T(2) = \phi \hat{X}_T(1) = \phi(\phi X_T - \theta \varepsilon_T)$

Para cualquier $k > 2$: $\hat{X}_T(k) = \phi \hat{X}_T(k-1) = \phi^2 \hat{X}_T(k-2) = \dots \phi^{k-1} \hat{X}_T(1)$

Predicción con un modelo ARMA(p,q) general

$$\Phi(B)X_t = \Theta(B)\varepsilon_t$$

$$X_{T+k} = \phi_1 X_{T+k-1} + \dots + \phi_p X_{T+k-p} + \varepsilon_{T+k} - \theta_1 \varepsilon_{T+k-1} - \dots - \theta_q \varepsilon_{T+k-q}$$

$$\hat{X}_T(k) = \phi_1 \hat{X}_T(k-1) + \dots + \phi_p \hat{X}_T(k-p) + \hat{\varepsilon}_T(k) + \theta_1 \hat{\varepsilon}_T(k-1) + \dots + \theta_q \hat{\varepsilon}_T(k-q)$$

Donde:

$$\hat{X}_T(j) = E(X_{T+j} \mid X_T, X_{T-1}, \dots) \quad j \geq 1$$

$$\hat{X}_T(j) = X_{T+j} \quad j \leq 0$$

$$\hat{\varepsilon}_T(j) = 0 \quad j \geq 1$$

$$\hat{\varepsilon}_T(j) = \varepsilon_{T+j} = X_{T+j} - \hat{X}_{T+j-1}(1) \quad j \leq 0$$

Ejemplo: ARMA(2,2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

$$k=1 \quad X_{T+1} = \phi_1 X_T + \phi_2 X_{T-1} + \varepsilon_{T+1} - \theta_1 \varepsilon_T - \theta_2 \varepsilon_{T-1}$$

$$\hat{X}_T(1) = E(X_{T+1} | I_T) = \phi_1 X_T + \phi_2 X_{T-1} - \theta_1 \hat{\varepsilon}_T - \theta_2 \hat{\varepsilon}_{T-1}$$

Donde:

$$\hat{X}_T(0) = X_T$$

$$\hat{X}_T(-1) = X_{T-1}$$

$$\hat{\varepsilon}_T = \frac{(1 - \phi_1 B - \phi_2 B^2)}{(1 - \theta_1 B - \theta_2 B^2)} X_T$$

$$\hat{\varepsilon}_{T-1} = X_{T-1} - \hat{X}_{T-2}(1)$$

Actualización de predicciones

Supongamos que tenemos información hasta T y deseamos calcular predicciones hasta $T+k$.

$$\hat{X}_T(1), \hat{X}_T(2), \dots, \hat{X}_T(k)$$

Si ahora tenemos un dato más, podemos revisar la predicción referida a $T+k$?

Paso 1:
$$e_T(k) = X_{T+k} - \hat{X}_T(k) = \sum_{j=0}^{k-1} \psi_j \varepsilon_{T+k-j}$$

Paso 2:
$$e_{T-1}(k+1) = \sum_{j=0}^{k+1-1} \psi_j \varepsilon_{T-1+k+1-j} = \sum_{j=0}^k \psi_j \varepsilon_{T+k-j}$$

$$e_{T-1}(k+1) = \sum_{j=0}^{k-1} \psi_j \varepsilon_{T+k-j} + \psi_k \varepsilon_T = e_T(k) + \psi_k \varepsilon_T$$

Paso 3:

$$X_{T+k} - \hat{X}_{T-1}(k+1) = X_{T+k} - \hat{X}_T(k) + \psi_k \varepsilon_T$$

$$\hat{X}_T(k) = \hat{X}_{T-1}(k+1) + \psi_k \varepsilon_T$$

$$\hat{X}_{T+1}(k) = \hat{X}_T(k+1) + \psi_k \varepsilon_{T+1}$$

Fuentes de error en las predicciones

Las predicciones siempre están sujetas a error. Como mínimo debemos considerar tres fuentes de error:

- **Error en la especificación** del modelo ARMA. Cualquier modelo no deja de ser una aproximación.
- **Error en la innovación** del modelo ARMA. Las innovaciones futuras son desconocidas en el momento de calcular la predicción.
- **Error en los parámetros** del modelo ARMA. Los coeficientes del modelo ARMA son el resultado de un proceso de estimación y, por tanto, están sujetos a un error de estimación.

Evaluación de predicciones

Las medidas más habituales para evaluar la precisión de una predicción son:

El error cuadrático medio: $E Q M = \frac{1}{k} \sum_{j=1}^k e_T^2(j) = \frac{1}{k} \sum_{j=1}^k \left(\hat{X}_T(j) - X_{T+j} \right)^2$

La raíz cuadrada del error cuadrático medio: $R E Q M = \sqrt{\frac{1}{k} \sum_{j=1}^k e_T^2(j)}$

El error absoluto medio: $E A M = \frac{1}{k} \sum_{j=1}^k |e_T(j)| = \frac{1}{k} \sum_{j=1}^k \left| \hat{X}_T(j) - X_{T+j} \right|$

El error porcentual absoluto medio: $E P A M = \frac{1}{k} \sum_{j=1}^k \frac{\left| \hat{X}_T(j) - X_{T+j} \right|}{X_{T+j}}$

Anexo. Combinación de predicciones

Sean $f_t^{(1)}$ y $f_t^{(2)}$ dos predicciones de X_t con errores:

$$e_t^{(j)} = X_t - f_t^{(j)} \text{ tq}$$

$$E(e_t^{(j)}) = 0, \quad E(e_t^{(j)})^2 = \sigma_j^2, \quad i$$

$$E(e_t^{(1)} e_t^{(2)}) = \rho \sigma_1 \sigma_2$$

Podemos considerar ahora una combinación de predicciones que sea una media ponderada de las dos predicciones individuales.

$$C_t = p f_t^{(1)} + (1 - p) f_t^{(2)}$$

El error de predicción es:

$$e_t^{(c)} = X_t - C_t = k e_t^{(1)} + (1 - k) e_t^{(2)}$$

Anexo. Combinación de predicciones

Por lo tanto, la varianza de este error (combinación) es:

$$\sigma_c^2 = p^2 \sigma_1^2 + (1-p)^2 \sigma_2^2 + 2p(1-p)\rho\sigma_1\sigma_2$$

Esta expresión la podemos minimizar respecto al valor p :

$$p^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

y substituyendo en la expresión anterior, concluimos que la varianza mínima que se puede alcanzar es igual a:

$$\sigma_{c^*}^2 = \frac{\sigma_1^2 \sigma_2^2 (1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

Observemos que $\sigma_{c^*}^2 < \min(\sigma_1^2, \sigma_2^2)$, excepto si $\rho = \frac{\sigma_1}{\sigma_2}$ o $\rho = \frac{\sigma_2}{\sigma_1}$

Si se cumple alguna de estas dos igualdades, la varianza de la combinación de predicciones será igual a la menor de las varianzas de los dos errores de predicción “simples”.