

Métodos Numéricos I:

Modelos de series temporales.

Contenidos

1. Series temporales.
 - Valores esperados.
 - Procesos estacionarios.
 - Ecuaciones de diferencias.
 - Predicción.
2. Modelos de series temporales
 - Medias móviles: Modelos $MA(q)$.
 - Procesos autorregresivos: $AR(p)$.
 - Series no estacionarias.
 - Heterocedasticidad
 - * Modelo $ARCH(q)$
 - * Modelo $GARCH(p,q)$.

Series temporales.

Una serie temporal es una secuencia ordenada de valores, correspondientes a la magnitud de una variable en un determinado instante en el tiempo

$$\{X_\tau\}_{\tau=1}^\infty \equiv X_1, X_2, \dots, X_t, \dots$$

- Ejemplo:

- Ruido blanco gaussiano: Secuencia de variables aleatorias con una distribución normal $\mathcal{N}(0, \sigma)$

$$\epsilon_1, \epsilon_2, \dots, \epsilon_t, \dots$$

- Caminante Browniano: Secuencia de variables aleatorias

$$0, \epsilon_1, \sum_{\tau=1}^2 \epsilon_\tau, \sum_{\tau=1}^3 \epsilon_\tau, \dots, \sum_{\tau=1}^t \epsilon_\tau, \dots$$

en la cual $\{\epsilon_\tau\}_{\tau=1}^\infty$ es ruido blanco gaussiano.

La hipótesis que realizamos es que la serie temporal de longitud T está generada mediante la extracción de muestras de una distribución de densidad de probabilidad

$$P\left(\{X_\tau\}_{\tau=1}^T\right)$$

- Ejemplo: En una secuencia determinista, en la cual la trayectoria es única, la distribución es un producto de distribuciones *delta*

$$P \left(\{X_\tau\}_{\tau=1}^T \right) = \prod_{\tau=1}^T \delta (X_\tau - G(X_{\tau-1}, \tau))$$

- Ejemplo: Para ruido blanco gaussiano la distribución de densidad de probabilidad es factorizable

$$P \left(\{\epsilon_\tau\}_{\tau=1}^T \right) = \prod_{\tau=1}^T \left[\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \epsilon_\tau^2 \right\} \right]$$

La **distribución marginal** de la variable X_t se obtiene mediante integración de la distribución completa respecto al resto de las variables

$$\begin{aligned} P(X_t) = & \int dX_1 \int dX_2 \dots \int dX_{t-1} \int dX_{t+1} \dots \\ & \dots \int dX_T P \left(\{X_\tau\}_{\tau=1}^T \right) . \end{aligned}$$

Valores esperados.

El valor esperado de una determinada función de los valores de una serie temporal $F(\{X_\tau\}_{\tau=1}^T)$ es

$$\mathbf{E} \left[F \left(\{X_\tau\}_{\tau=1}^T \right) \right] = \int dX_1 \int dX_2 \dots \int dX_T F(\{X_\tau\}_{\tau=1}^T) P \left(\{X_\tau\}_{\tau=1}^T \right)$$

En la práctica los valores esperados se obtienen mediante un promedio sobre realizaciones de la serie temporal

$$\left\{ X_\tau^{(i)} \right\}_{\tau=1}^T \equiv X_1^{(i)}, X_2^{(i)}, \dots, X_T^{(i)}, \quad i = 1, 2, \dots, I$$

La estimación empírica mediante el promedio sobre I realizaciones es

$$\langle F \rangle = \frac{1}{I} \sum_{i=1}^I F \left(\left\{ X_\tau^{(i)} \right\}_{\tau=1}^T \right),$$

Esta estimación converge al valor exacto en el límite de un número infinito de realizaciones:

$$\langle F \rangle \rightarrow \mathbf{E} [F], \quad \text{cuando } I \rightarrow \infty.$$

- Media:

$$\mathbf{E}[X_t] = \mu_t.$$

- Varianza: Definiendo

$$\hat{X}_t = X_t - \mu_t,$$

la varianza es

$$\mathbf{E}[\hat{X}_t^2] = \sigma_t^2.$$

- Autocovarianza:

$$\mathbf{E}[\hat{X}_{t+\tau}\hat{X}_t] = \gamma(t; \tau).$$

- Autocorrelación:

$$\rho(t; \tau) = \frac{\gamma(t; \tau)}{\sigma_t^2}.$$

- Ejemplo: Ruido blanco gaussiano.

- $\mathbf{E}[\epsilon_t] = 0$
- $\mathbf{E}[\epsilon_{t+\tau}\epsilon_t] = \sigma^2\delta_{\tau,0}$
- $\epsilon_t \approx \mathcal{N}(0, \sigma)$

Procesos estacionarios.

- Un proceso $X_0, X_1, \dots, X_t, \dots$ es **estacionario en sentido estricto** si se cumple

$$P(X_{t_1}, X_{t_2}, \dots, X_{t_r}) = P(X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_r+\tau}).$$

- Un proceso es **débilmente estacionario**, o estacionario con respecto a la covarianza si cumple las condiciones

$$\begin{aligned}\mathbf{E}[X_t] &= \mu \\ \mathbf{E}[\hat{X}_{t+\tau}\hat{X}_t] &= \gamma_\tau\end{aligned}$$

La condición de estacionaridad estricta implica la débil si existen los dos primeros momentos de la distribución.

- Un proceso estacionario es **ergódico** respecto a la media si

$$\langle X \rangle = \frac{1}{T} \sum_{\tau=1}^T X_\tau \rightarrow \mu, \quad T \rightarrow \infty$$

- Un proceso estacionario es **ergódico** respecto a la varianza si

$$\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \hat{X}_{t+\tau}\hat{X}_t \rightarrow \gamma_\tau, \quad T \rightarrow \infty$$

Función de autocovarianza/autocorrelación.

Para un proceso estacionario (a partir de este momento, se utilizará la condición débil), se define la función de covarianza como

$$\gamma_{\tau} = \mathbf{E} [X_{t+\tau} X_t] .$$

El coeficiente de correlación es

$$\rho_{\tau} = \frac{\gamma_{\tau}}{\gamma_0} .$$

El valor de este coeficiente está acotado

$$-1 \leq \rho_{\tau} \leq 1 .$$

Se define el operador de retrasos

$$LX_t = X_{t-1};$$

con las propiedades

$$\begin{aligned} L^0 X_t &= X_t; \\ L^{-1} X_t &= X_{t+1}; \\ L^{\tau} X_t &= X_{t-\tau}; \end{aligned}$$

Ecuaciones de diferencias de primer orden.

Consideremos la ecuación de diferencias de primer orden

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

La ecuación se resuelve mediante un método recursivo

$$X_t = \phi^t X_0 + \sum_{\tau=0}^{t-1} \phi^\tau \epsilon_{t-\tau}$$

Los comportamientos posibles de la solución son

- Con el valor $\phi > 1$ la solución es explosiva.
- Con el valor $\phi < -1$ la solución es explosiva y presenta oscilaciones.
- Con el valor $0 \leq \phi < 1$ la solución decae exponencialmente.
- Con el valor $-1 < \phi \leq 0$ la solución decae exponencialmente con oscilaciones.

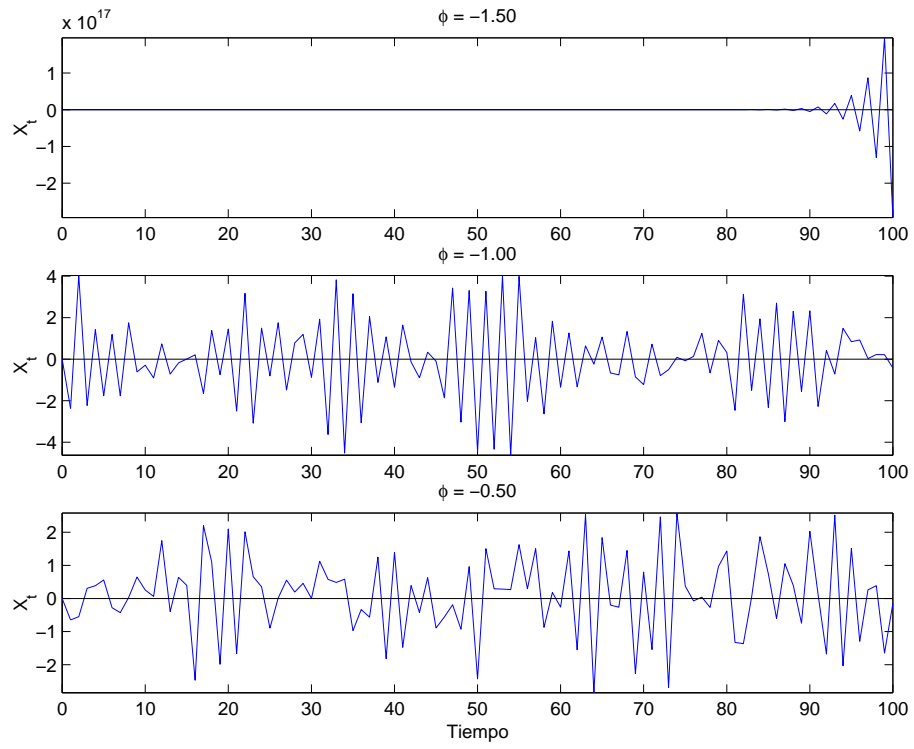


Figure 1: Simulaciones ($\sigma = 1$).

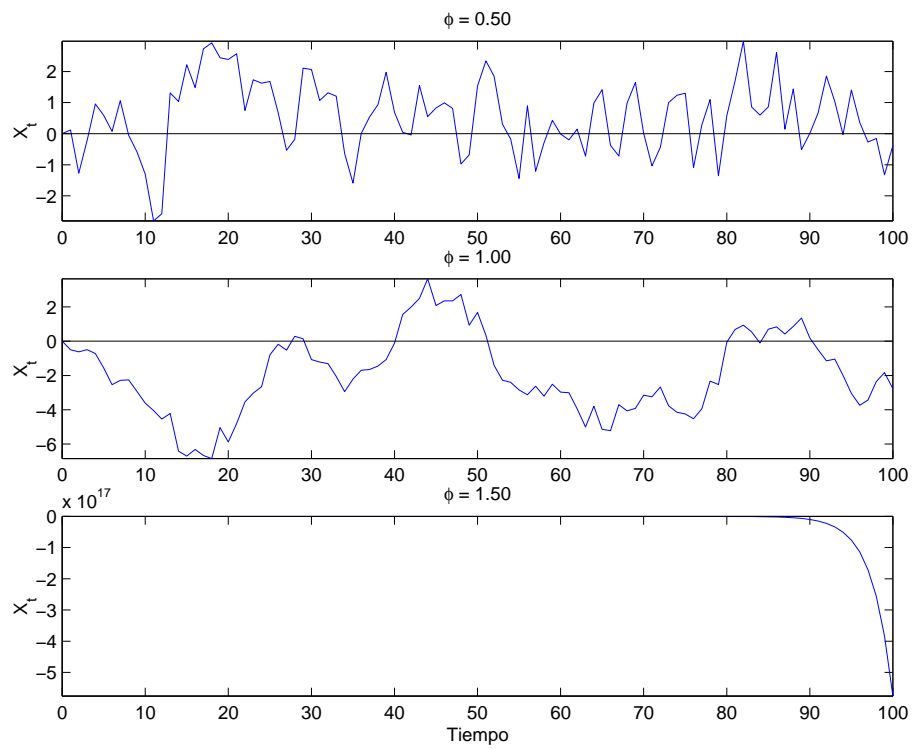


Figure 2: Simulaciones ($\sigma = 1$).

Ecuaciones de diferencias de orden p .

Consideremos la ecuación de diferencias de primer orden

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t.$$

Definiendo un vector de estado

$$\boldsymbol{\xi}_t^\dagger = (X_t \ X_{t-1} \ \dots \ X_{t-p+1}),$$

la ecuación se puede reescribir como

$$\boldsymbol{\xi}_t = \mathbf{F} \cdot \boldsymbol{\xi}_{t-1} + \boldsymbol{\epsilon}_t,$$

con las definiciones

$$\boldsymbol{\epsilon}_t^\dagger = (\epsilon_t \ 0 \ 0 \ \dots \ 0 \ 0)$$
$$\mathbf{F} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

El caracter explosivo de la evolución depende de los autovalores de \mathbf{F} .

Ejemplo:

Consideremos la ecuación de diferencias de orden 2

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t.$$

La ecuación se puede escribir de manera equivalente como

$$\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}.$$

El comportamiento de las soluciones depende de los autovalores de \mathbf{F}

$$\begin{aligned} \lambda_1 &= \frac{1}{2} \left(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2} \right), \\ \lambda_2 &= \frac{1}{2} \left(\phi_1 - \sqrt{\phi_1^2 + 4\phi_2} \right). \end{aligned}$$

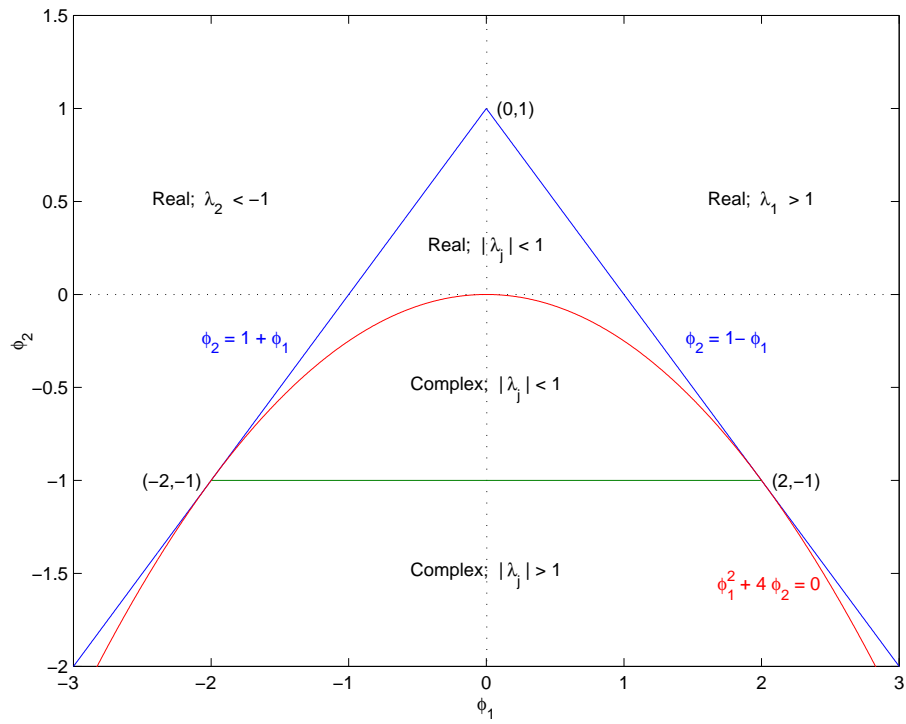


Figure 3: Diagrama de valores.

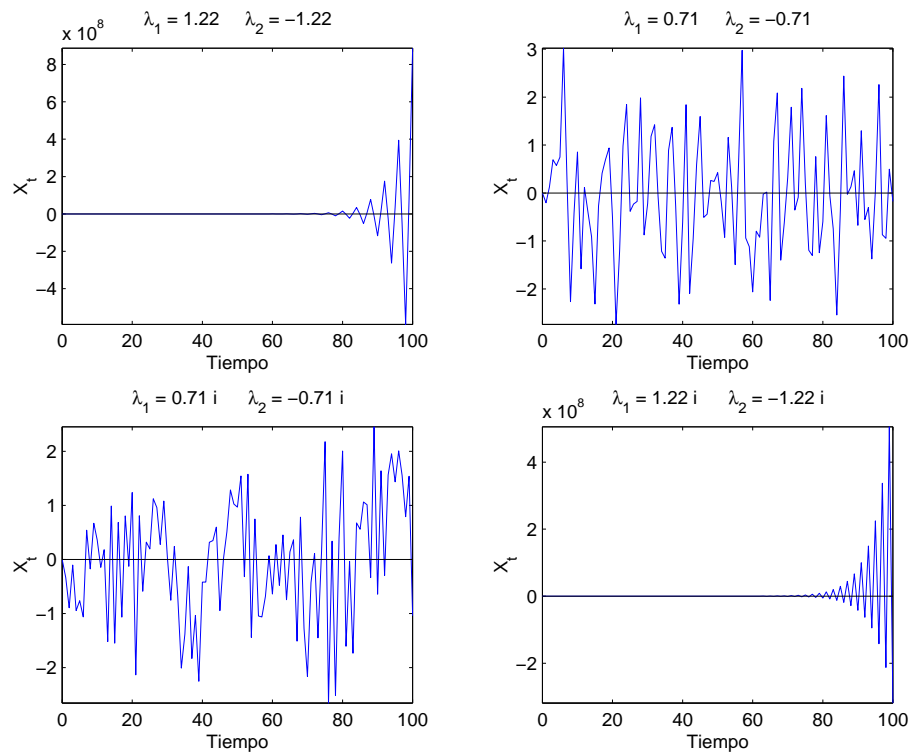


Figure 4: Simulaciones ($\sigma = 1$).

Predicción en series económicas.

Las series económicas presentan una serie de características comunes

- Generalmente las series temporales disponibles son cortas (excepto las series de “alta frecuencia”).
- Los valores de la serie temporal son generalmente difíciles de medir con precisión (definiciones imprecisas: inflación, desempleo, PNB, etc).
- No son estacionarias (tendencias a largo plazo, tendencias estacionales en la media y en la varianza, cambios en el paradigma económico, memoria a largo plazo, etc.)
- Series no-lineales con componentes estocásticos.

En el problema de predicción hay diversos elementos que entran en juego:

- Conjunto de datos de partida.
 - Determinar variables relevantes al problema.
 - Determinar ventana de tiempo óptima.
 - Determinar frecuencia óptima de los datos.
 - Obtener los datos.

Esta elección está condicionada a si el horizonte de predicción es a largo / corto plazo.

- Elección del modelo a utilizar.
 - Modelos paramétricos
 - * Lineales.
 - * No lineales (ej. modelos con cambio de régimen).
 - Modelos no paramétricos
 - * Redes neuronales.
 - * Mezclas jerárquicas de expertos.
 - * Árboles de regresión.
- Estimación y evaluación del modelo seleccionado

El conjunto de datos se divide en un conjunto de entrenamiento que sirve para estimar los parámetros del modelo propuesto, y en un conjunto de prueba que se utiliza para evaluar la capacidad de predicción (generalización) del modelo. Generalmente se utiliza el **Error cuadrático medio** como criterio para la evaluación de la predicción, aunque puede haber casos en los que ésta no sea una medida razonable.

En general se observa con este procedimiento que los modelos lineales simples son más robustos en sus predicciones que los no lineales. Este fenómeno puede ser debida a varias causas

- Los modelos no lineales “memorizan” la serie temporal (sobreajuste).
- La serie datos de prueba suelen ser más corta que la serie de entrenamiento, y los efectos de la no-linealidad pueden ser pequeños.
- El criterio de error elegido puede conducir a conclusiones erróneas sobre la calidad de la predicción.

Medias móviles: MA(q)

El proceso de medias móviles de orden q es generado mediante la ecuación de diferencias

$$X_t = \epsilon_t + \tilde{\boldsymbol{\theta}}^\dagger \cdot \tilde{\boldsymbol{\epsilon}}_t^{(q)},$$

donde ϵ_t es ruido blanco gaussiano con desviación estándar σ , el vector de retrasos de este ruido blanco es

$$\left[\tilde{\boldsymbol{\epsilon}}_t^{(q)} \right]^\dagger = \left(1 \quad \left[\boldsymbol{\epsilon}_t^{(q)} \right]^\dagger \right)$$

con la definición

$$\left[\tilde{\boldsymbol{\epsilon}}_t^{(q)} \right]^\dagger = (\epsilon_{t-1} \quad \epsilon_{t-2} \dots \epsilon_{t-q}).$$

El vector de parámetros es

$$\tilde{\boldsymbol{\theta}}^\dagger = (\theta_0 \quad \boldsymbol{\theta}^\dagger),$$

con

$$\boldsymbol{\theta}^\dagger = (\theta_1 \quad \theta_2 \dots \theta_q).$$

- La media del proceso es

$$\mu = \theta_0$$

- La función autocovarianza del proceso es

$$\gamma_\tau = \begin{cases} \theta_\tau \sigma + \sum_{j=1}^{q-\tau} \theta_j \theta_{\tau+j} & \tau \leq q \\ 0 & \tau > q \end{cases}$$

- Ejemplo: MA(1)

$$X_t = \mu + \epsilon_t + \theta \epsilon_{t-1}$$

- Varianza:

$$E \left[\hat{X}_t^2 \right] = (1 + \theta^2) \sigma^2.$$

- Autocovarianza:

$$\gamma_1 = \theta \sigma^2$$

$$\gamma_2 = \gamma_3 = \dots = 0$$

- Autocorrelación :

$$\rho_1 = \frac{\theta}{1 + \theta^2},$$

$$\rho_2 = \rho_3 = \dots = 0$$

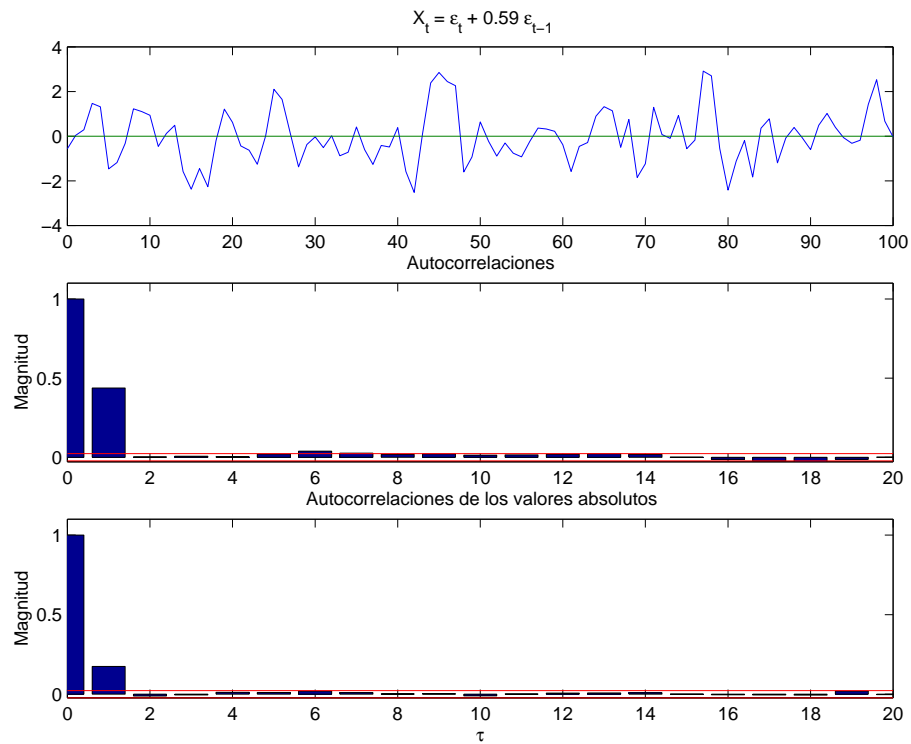


Figure 5: Simulación MA(1) ; $\sigma = 1$.

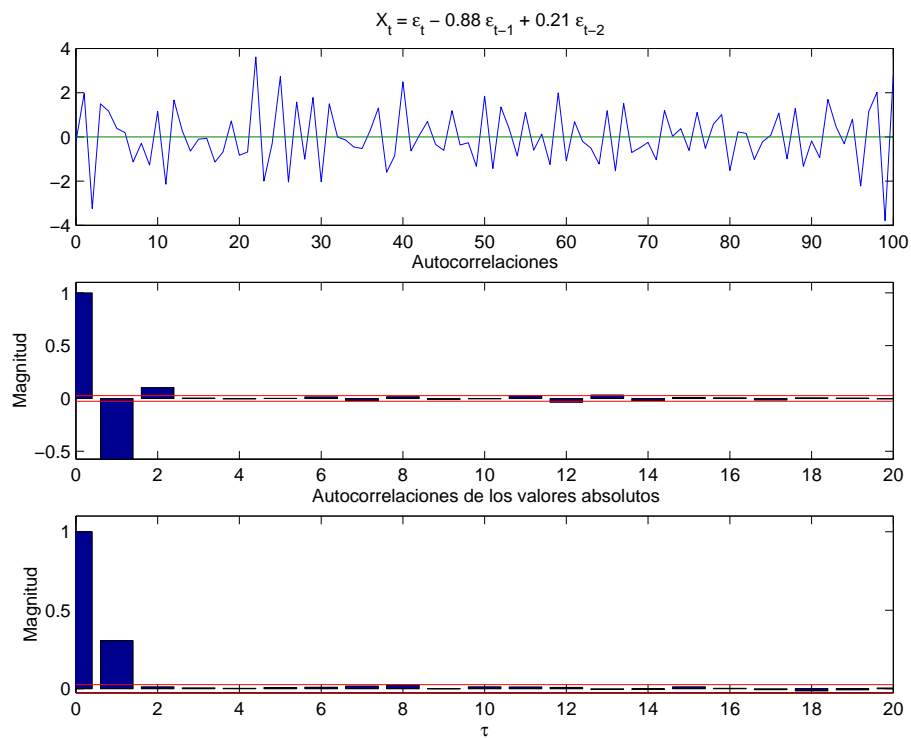


Figure 6: Simulación MA(2) ; $\sigma = 1$.

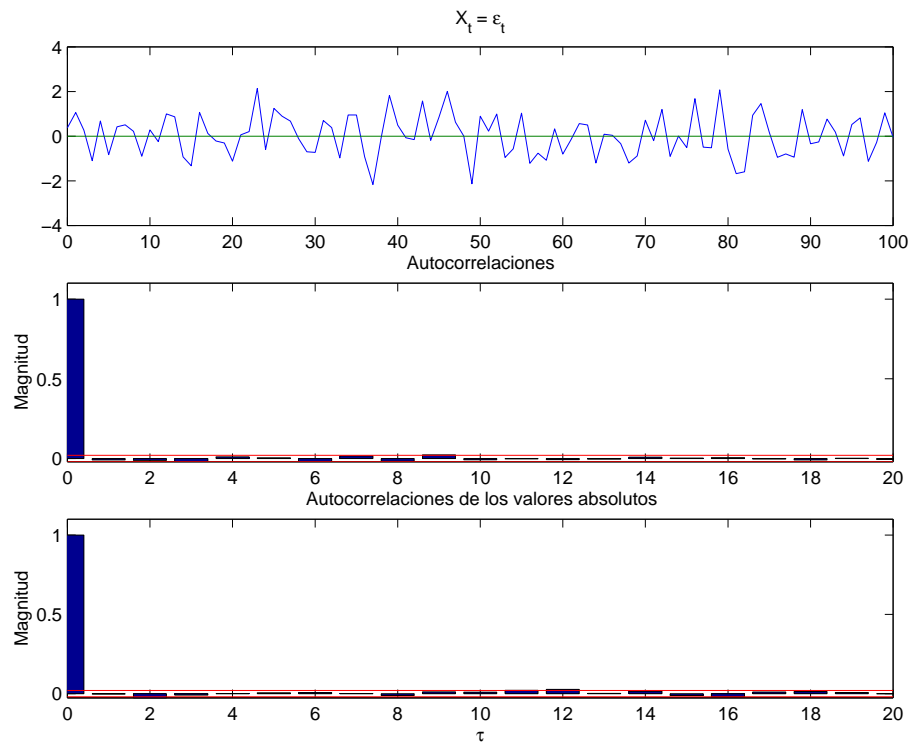


Figure 7: Simulación MA(0) ; $\sigma = 1$.

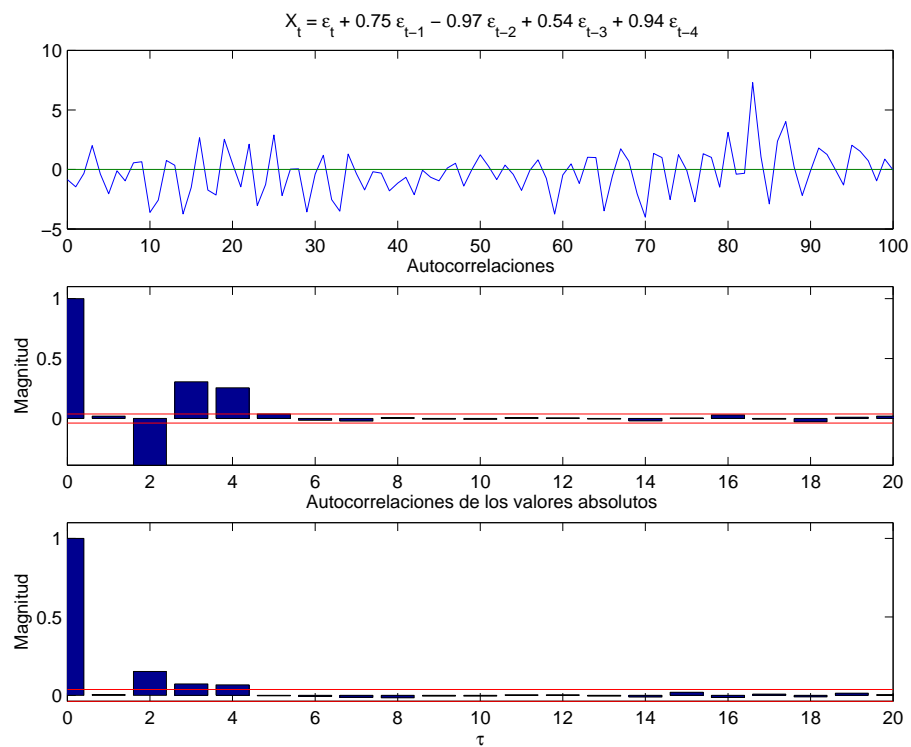


Figure 8: Simulación MA(4) ; $\sigma = 1$.

Proceso al Imite MA(∞)

Consideremos el modelo de medias mviles

$$X_t = \mu + \sum_{\tau=0}^{\infty} \psi_{\tau} \epsilon_{t-\tau},$$

donde ϵ_t es ruido blanco gaussiano con desviación estándar σ .

La condición suficiente para que el proceso sea estacionario es

$$\sum_{\tau=0}^{\infty} \psi_{\tau}^2 < \infty.$$

Normalmente, se requiere una condición más estricta que implica la anterior

$$\sum_{\tau=0}^{\infty} |\psi_{\tau}| < \infty.$$

En el caso de que el proceso sea estacionario,

- La media del proceso es

$$\mathbf{E}[X_t] = \mu$$

- La varianza es

$$\gamma_0 = \sum_{j=0}^{\infty} \psi_j^2$$

- La función de autocovarianza es

$$\gamma_{\tau} = \sum_{j=0}^{\infty} \psi_j \psi_{j+\tau}.$$

Procesos autorregresivos: AR(p)

El proceso autorregresivo de orden p es generado mediante la ecuación de diferencias

$$X_t = \tilde{\phi}^\dagger \cdot \tilde{\mathbf{X}}_t^{(p)} + \epsilon_t,$$

donde ϵ_t es ruido blanco gaussiano con desviación estándar σ , el vector de retrasos de X es

$$\begin{aligned} \left[\tilde{\mathbf{X}}_t^{(p)} \right]^\dagger &= \left(1 \left[\mathbf{X}_t^{(p)} \right]^\dagger \right), \\ \left[\mathbf{X}_t^{(p)} \right]^\dagger &= (X_{t-1} \ X_{t-2} \dots \ X_{t-p}), \end{aligned}$$

y el vector de parámetros es

$$\begin{aligned} \tilde{\phi}^\dagger &= (\phi_0 \ \phi^\dagger), \\ \phi^\dagger &= (\phi_1 \ \phi_2 \dots \ \phi_p). \end{aligned}$$

- El proceso es estacionario si las raíces de la ecuación

$$\sum_{i=1}^p \phi_i z^i = 1$$

se encuentran fuera del círculo unidad en el plano complejo.

- La media incondicional del proceso es

$$\mathbf{E}[X_t] = \frac{\phi_0}{1 - \sum_{i=1}^p \phi_i}$$

- La media condicional del proceso es

$$\mathbf{E}[X_t | X_{t-1} X_{t-2} \dots X_{t-p}] = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i}$$

- Ecuaciones de *Yule-Walker*: Las funciones de autocorrelación satisfacen la misma ecuación que el proceso autorregresivo

$$\gamma_\tau = \sum_{i=1}^p \phi_i \gamma_{\tau-i} + \sigma^2 \delta_{\tau,0}$$

$$\rho_\tau = \sum_{i=1}^p \phi_i \rho_{\tau-i}$$

- Ejemplo: AR(1)

$$X_t = \phi_0 + \phi_1 X_{t-1} + \epsilon_t, \quad \phi_1 < 1$$

– Media:

$$E[X_t] = \frac{\phi_0}{1 - \phi_1}.$$

– Varianza:

$$E[\hat{X}_t^2] = \frac{\sigma^2}{1 - \phi_1^2}$$

– Autocovarianza y autocorrelación:

$$\gamma_j = \frac{\phi_1^j}{1 - \phi_1^2} \sigma^2$$

$$\rho_j = \phi_1^j$$

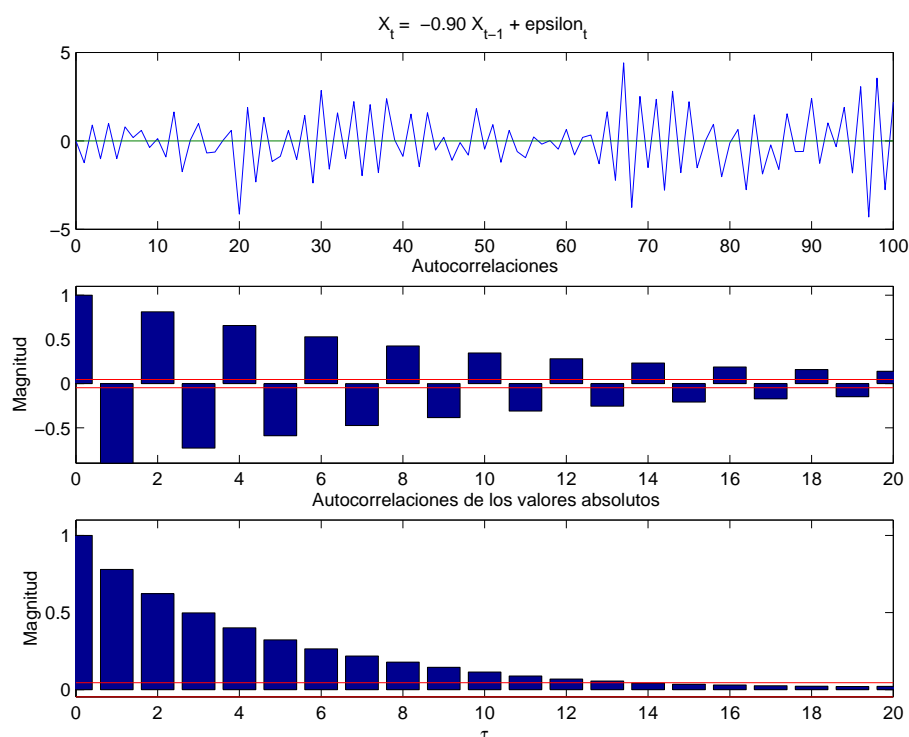


Figure 9: Simulaciones AR(1) ($\phi = -0.9$; $\sigma = 1$).

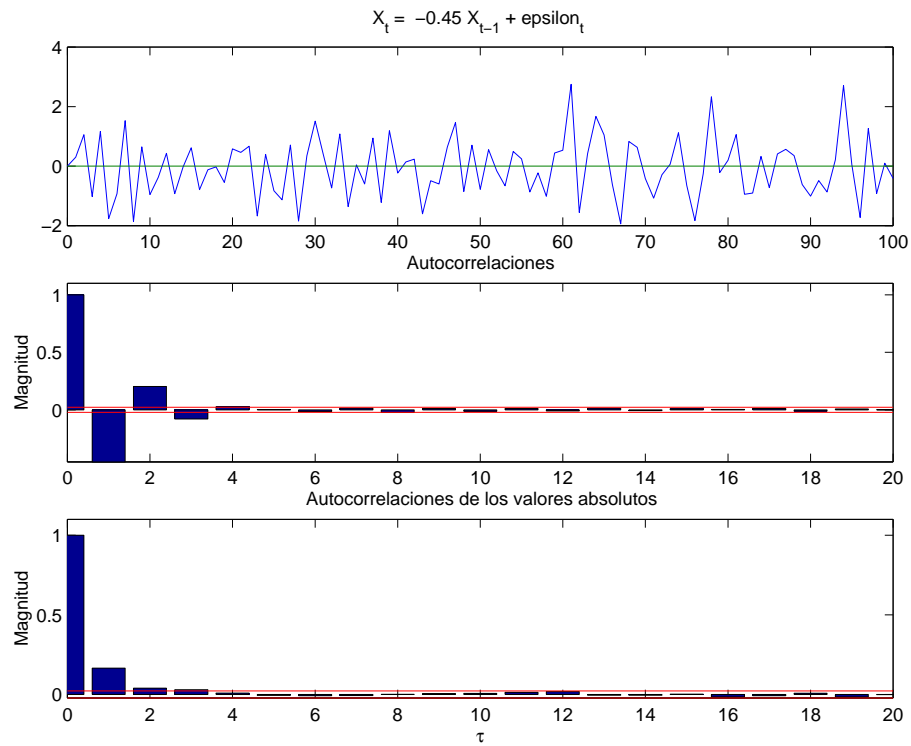


Figure 10: Simulaciones AR(1) ($\phi = -0.45$; $\sigma = 1$).

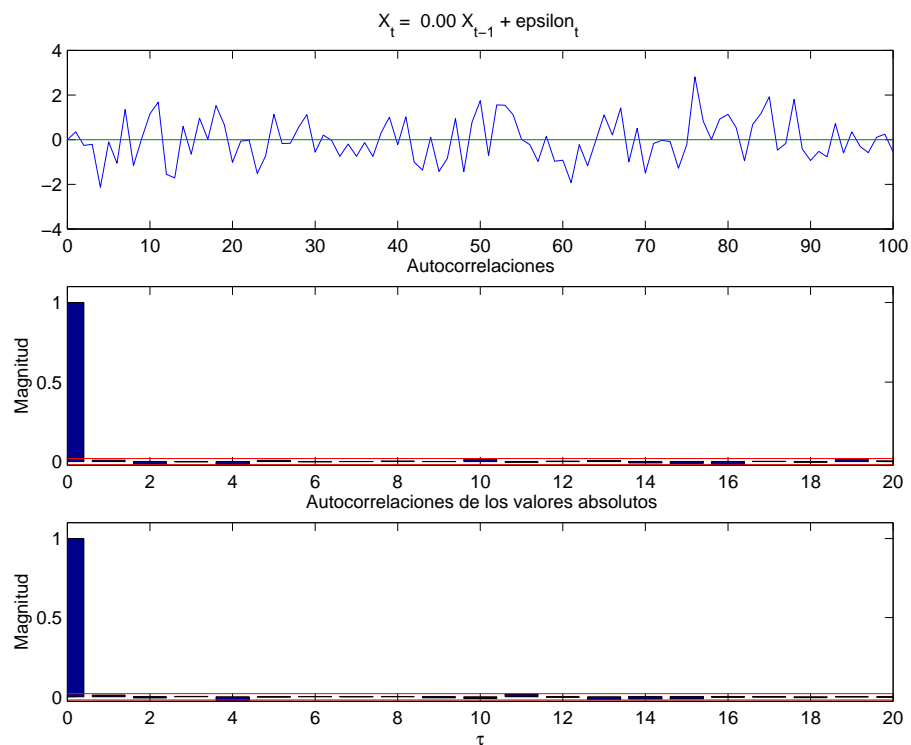


Figure 11: Simulaciones AR(1) ($\phi = 0$; $\sigma = 1$).

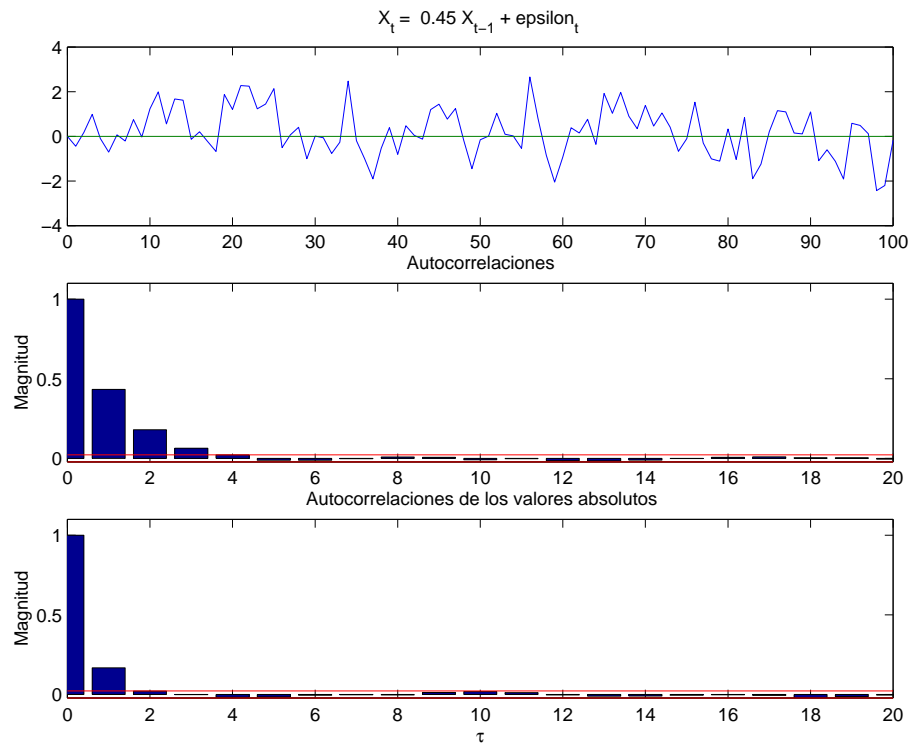


Figure 12: Simulaciones AR(1) ($\phi = 0.45$; $\sigma = 1$).

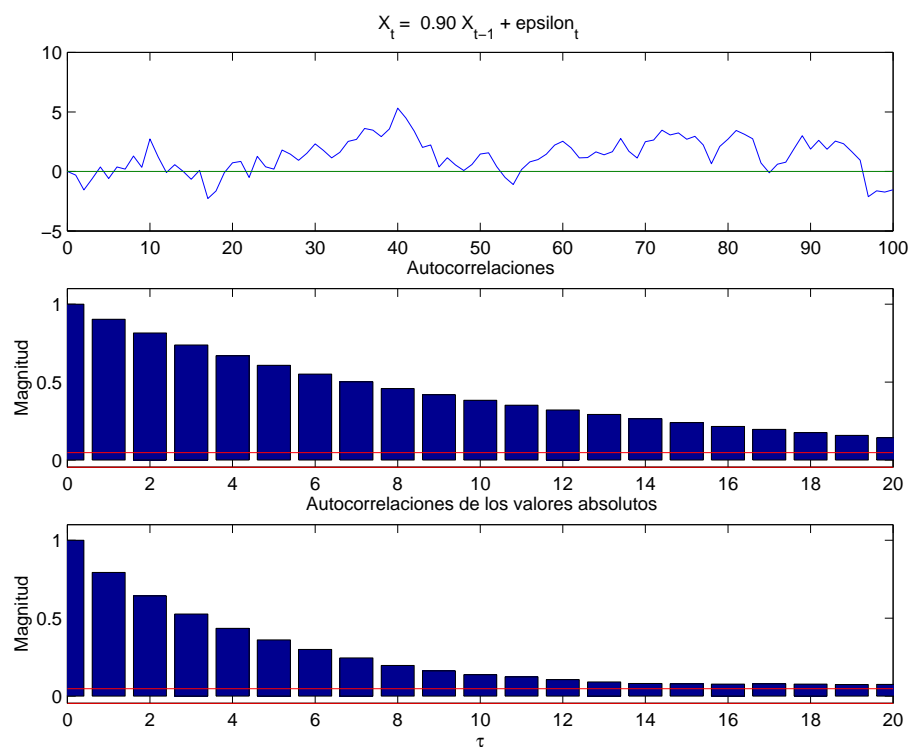


Figure 13: Simulaciones AR(1) ($\phi = 0.9$; $\sigma = 1$).

El proceso AR(p) como un proceso MA(∞)

En términos del operador de retardo $LX_t = LX_{t-1}$, el proceso AR(p) se puede escribir como

$$X_t = \phi_0 + \sum_{\tau=1}^p \phi_{\tau} L^{\tau} X_t + \epsilon_t$$

o, de manera equivalente

$$(1 - \sum_{\tau=1}^p \phi_{\tau} L^{\tau}) X_t = \phi_0 + \epsilon_t$$

Suponiendo que el inverso existe (lo cual est garantizado si el proceso es estacionario respecto a la covarianza)

$$X_t = \mu + \psi(L)\epsilon_t,$$

que corresponde a un proceso MA(∞) con

$$\mu = \frac{\phi_0}{1 - \sum_{\tau=1}^p \phi_{\tau}} \quad \psi(L) = \frac{1}{1 - \sum_{\tau=1}^p \phi_{\tau} L^{\tau}}.$$

Ejemplo: AR(1)

El proceso AR(1) con $|\phi_1| < 1$

$$X_t = \phi_0 + \phi_1 X_{t-1} + \epsilon_t$$

es equivalente a un proceso MA(∞)

$$\begin{aligned} X_t &= (1 - \phi_1 L)^{-1} [\phi_0 + \epsilon_t] = \\ &= \frac{\phi_0}{1 - \phi_1} + \frac{1}{1 - \phi_1 L} \epsilon_t = \frac{\phi_0}{1 - \phi_1} + \sum_{\tau=0}^{\infty} \phi_1^{\tau} L^{\tau} \epsilon_t = \\ &= \frac{\phi_0}{1 - \phi_1} + \sum_{\tau=0}^{\infty} \phi_1^{\tau} \epsilon_{t-\tau}. \end{aligned}$$

Procesos ARMA(p,q)

Se puede construir un modelo que incluya simultáneamente términos de medias móviles y autorregresivos

$$X_t = c + \phi^\dagger \cdot \mathbf{X}_t^{(p)} + \theta^\dagger \cdot \boldsymbol{\epsilon}_t^{(q)} + \epsilon_t.$$

El proceso es estacionario respecto a la varianza si las raíces de

$$1 - \sum_{i=1}^p \phi_i z^i = 0$$

se encuentran fuera del círculo unidad. Esta condición es independiente de los valores de los coeficientes del término correspondiente a medias móviles.

El proceso también se puede expresar como un modelo MA(∞)

$$X_t = \mu + \psi(L)\epsilon_t,$$

con las definiciones

$$\mu = \frac{c}{1 - \sum_{i=1}^p \phi_i}; \quad \psi(L) = \frac{\sum_{i=1}^q \theta_i L^i}{1 - \sum_{i=1}^p \phi_i L^i}.$$

De esta forma se puede también poner de manifiesto posibles problemas de modelos degenerados: Modelos con factores comunes en el denominador y numerador de $\psi(L)$ generan series temporales que son indistinguibles.

Predicción

La predicción del valor de X_{t+h} basado en un conjunto de variables explicativas $\mathcal{I}_t \equiv \{\mathbf{Y}_t - \tau; \tau \geq 0\}$ equivale a hacer una estimación de la distribución condicional

$$P(X_{t+h}|I_t).$$

Normalmente las estimaciones se restringen a los dos primeros momentos:

- La media condicional

$$\hat{X}_{t+h|t} = \mathbf{E}[X_{t+h}|I_t] =$$

- La varianza condicional

$$\mathbf{E}\left[\left(X_{t+h} - \hat{X}_{t+h|t}\right)^2 | I_t\right]$$

La media condicional es el estimador para X_{t+h} que minimiza el error cuadrático medio (varianza) de la estimación.

Predicción en procesos ARMA(p,q)

Consideremos un proceso ARMA(p,q) invertible

$$\left(1 - \sum_{j=1}^p \phi_j L^j\right) (X_t - \mu) = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \epsilon_t$$

Este proceso se puede reescribir como

$$X_t - \mu = \sum_{j=1}^p \phi_j L^j (X_t - \mu) + \left(1 + \sum_{j=1}^q \theta_j L^j\right) \epsilon_t.$$

La predicción con un horizonte $s = 1$ es

$$\hat{X}_{t|t-1} - \mu = \sum_{j=1}^p \phi_j L^j (X_t - \mu) + \sum_{j=1}^q \theta_j L^j \epsilon_t,$$

de donde se deduce

$$\epsilon_t = X_t - \hat{X}_{t|t-1}$$

La predicción con un horizonte s es

$$\hat{X}_{t+s|t} - \mu = \sum_{j=1}^p \phi_j (\hat{X}_{t+s-j|t} - \mu) + \sum_{j=s}^q \theta_j \epsilon_{t+s-j}.$$

Predicción con un número finito de observaciones

Las fórmulas de predicción del apartado anterior asumen que disponemos de un número infinito de observaciones hacia el pasado. En caso de que el número de observaciones sea finito $\{X_{t-m+1}, X_{t-m+2}, \dots, X_t\}$, podemos suponer que las innovaciones son nulas para los tiempos anteriores al inicial

$$\epsilon_\tau \tau \leq t - m.$$

Este procedimiento es razonable siempre que la memoria de la serie temporal sea a corto plazo y que $t \gg 0$.

Existe una predicción más precisa, basada en encontrar la proyección exacta de $(X_{t+1} - \mu)$ en los m valores más recientes

$$\mathbf{X}_t^\dagger = (X_t - \mu, X_{t-1} - \mu, \dots, X_{t-m+1} - \mu).$$

La predicción con un horizonte $s = 1$ es

$$\hat{X}_{t+1|t} - \mu = \sum_{i=1}^m \alpha_i (X_{t-i+1} - \mu)$$

tiene como solución

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{pmatrix}.$$

Numéricamente el sistema se puede resolver mediante una descomposición de Cholesky.

La predicción con un horizonte s es

$$\hat{X}_{t+1|t} - \mu = \sum_{i=1}^m \alpha_i^{(s)} (X_{t-i+1} - \mu)$$

tiene como solución

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1^{(s)} \\ \alpha_2^{(s)} \\ \vdots \\ \alpha_m^{(s)} \end{pmatrix} = \begin{pmatrix} \gamma_s \\ \gamma_{s+1} \\ \vdots \\ \gamma_{s+m-1} \end{pmatrix}$$

descomposición de Wold (Wold, 1938).

Un proceso cualquiera $\{X_t\}_{t=-\infty}^{\infty}$, estacionario respecto a la covarianza y de media cero, puede ser representado mediante el proceso

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + \kappa_t$$

con $\psi_0 = 1$ y $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$. La componente ϵ_t es ruido blanco y representa el error de predicción para X_t de la predicción lineal óptima a partir de toda la serie de valores anteriores de la variable

$$\epsilon_t = X_t - \hat{\mathbf{E}}[X_t | X_{t-1}, X_{t-2}, \dots]$$

La componente linealmente determinista κ_t no está correlacionada con ϵ_{t-j} para ning ún valor de j y puede ser predicha linealmente a partir de la secuencia de valores anteriores de X_t

$$\kappa_t = \hat{\mathbf{E}}[\kappa_t | X_{t-1}, X_{t-2}, \dots]$$

La componente $\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ es una componente linealmente no-determinista.

La descomposición de Wold no puede utilizarse directamente para proponer un modelo para la serie temporal, ya que contiene un número infinito de parámetros. Con la hipótesis adicional de que la función $\psi(L)$ puede ser aproximada con cierta precisión por un aproximante de Padé

$$\psi_L \equiv \sum_{j=0}^{\infty} \psi_j L^j \approx \frac{\theta(L)}{\phi(L)} = \frac{\sum_{i=0}^q \theta_i L^i}{\sum_{j=0}^p \phi_j L^j},$$

con $\theta_0 = \phi_0 = 1$.

Metodología Box-Jenkins.

La observación de que modelos con un número excesivo de parámetros conduce a modelos que generalizan de manera deficiente, nos lleva a preferir modelos con el número más pequeño de parámetros.

Box y Jenkins proponen una metodología para modelación de series temporales

1. Realizar las transformaciones necesarias sobre los datos de forma que la serie transformada sea estacionaria.

- Transformación logartmica

$$Y_t = \log \left(\frac{X_t}{X_{t-1}} \right)$$

- Transformación a rendimientos

$$r_t = \log \left(\frac{S_t}{S_{t-1}} \right) \approx \frac{S_t - S_{t-1}}{S_{t-1}}$$

- Substraer tendencias

$$Y_t = X_t - (\alpha + \delta t).$$

- Tomar diferencias

$$Y_t = X_t - X_{t-1}$$

- Eliminar componentes estacionales

$$Y_t = X_t - X_{t-12}$$

2. Elegir un modelo ARMA(p,q), con p y q pequeños.
3. Estimar los parámetros del modelo.
4. Hacer tests de diagnóstico para comprobar que el modelo es coherente con los datos de la serie temporal.

Modelo para una serie temporal no estacionaria.

Hasta este momento hemos descrito modelos de una variable para una serie temporal que pueden ser escritos como

$$X_t = \mu + \sum_{\tau=0}^{\infty} \psi_{\tau} \epsilon_{t-\tau} = \mu + \psi(L) \epsilon_t$$

con la propiedad

$$\sum_{\tau=0}^{\infty} |\psi_{\tau}| < \infty$$

y con las raíces de $\psi(z)$ fuera del círculo de radio unidad en el plano complejo. Para este tipo de series se cumple

$$\mathbf{E}[X_t] = \mu$$

$$\hat{X}_{t+s|t} \equiv \lim_{s \rightarrow \infty} \mathbf{E}[X_{t+s} | X_t, X_{t-1}, X_{t-2}, \dots] = \mu.$$

En el contexto de series económicas o financieras esto no tiene por qué ser correcto.

Hay dos formas de describir tendencias en una serie temporal

- Procesos que tienen un comportamiento estacionario una vez que se substraen un término de tendencia.

Ejemplo: Modelo con tendencia lineal

$$X_t = \alpha + \delta t + \psi(L)\epsilon_t$$

Las propiedades de este modelo son

- Predicción

$$\hat{X}_{t+s|t} = \alpha + \delta(t+s) + \sum_{\tau=0}^{\infty} \psi_{s+\tau} \epsilon_{t-\tau}$$

- Comportamiento en el límite: Dado que $\lim_{\tau \rightarrow \infty} \psi_{\tau} = 0$, se cumple

$$\lim_{s \rightarrow \infty} \mathbf{E} \left[\hat{X}_{t+s|t} - (\alpha + \delta(t+s)) \right] = 0.$$

- Error de predicción:

$$X_{t+s} - \hat{X}_{t+s|t} = \sum_{\tau=0}^{s-1} \psi_{\tau} \epsilon_{t+s-\tau}$$

El valor del error cuadrático medio es

$$\mathbf{E} \left[\left(X_{t+s} - \hat{X}_{t+s|t} \right)^2 \right] = \sigma^2 \sum_{\tau=0}^{s-1} \psi_{\tau}^2$$

Converge a una constante cuando $s \rightarrow \infty$.

- **Modelos integrados o de raízunidad:** Un modelo integrado de orden d es tal que al ser diferenciado d veces da lugar a un proceso estacionario.

- Caminante aleatorio con deriva δ

$$X_t = X_{t-1} + \delta + \epsilon_t, \quad \psi(1) \neq 0.$$

- Ejemplo: Proceso integrado de orden 1

$$(1 - L)X_t = \delta + \psi(L)\epsilon_t, \quad \psi(1) \neq 0$$

- Consideremos el proceso con una raízunidad

$$X_t = \alpha + \delta t + u_t,$$

donde u_t es un proceso ARMA(p,q). El proceso diferenciado es estacionario

$$(1 - L)X_t = \delta + \psi(L)\epsilon_t,$$

donde

$$\psi(L)\epsilon_t = (1 - L)u_t.$$

Las propiedades de este modelo son

* Predicción:

La relación básica es la predicción de las primeras diferencias

$$\mathbf{E}[X_{t+s} - X_{t+s-1} | X_t, X_{t-1}, \dots] = \delta + \sum_{\tau=0}^{\infty} \psi_{s+\tau} \epsilon_{t-\tau}.$$

Con la observación

$$X_{t+s} = (X_{t+s} - X_{t+s-1}) + (X_{t+s-1} - X_{t+s-2}) + \dots + (X_{t+1} - X_t) + X_t$$

se llega a la relación

$$\hat{X}_{t+s|t} = s\delta + X_t + \sum_{\tau=0}^{\infty} \sum_{\theta=1}^s \psi_{\theta+\tau} \epsilon_{t-\tau}$$

* Errores de predicción

El error de predicción es

$$X_{t+s} - \hat{X}_{t+s|t} = \sum_{\tau=1}^s \sum_{\theta=0}^{\tau-1} \psi_{\tau} \epsilon_{t+s-\tau}.$$

El valor del error cuadrático medio es

$$\mathbf{E} \left[\left(X_{t+s} - \hat{X}_{t+s|t} \right)^2 \right] = \sigma^2 \sum_{\theta=1}^s \sum_{\tau=0}^{\tau-1} \psi_{\tau}^2$$

Crece linealmente cuando $s \rightarrow \infty$.

En resumen, la mayor diferencia entre procesos no estacionarios de estos dos tipos es la persistencia del efecto de las innovaciones: En los procesos con tendencia la influencia de las innovaciones disminuye con el tiempo. Sin embargo, en los procesos con raíz unidad, el efecto de las innovaciones se acumula en el tiempo. A pesar de las diferencias cualitativa, determinar cuál es el proceso subyacente a partir de una muestra finita puede ser problemático.

Procesos ARIMA(p,d,q)

Un caso especial de modelos integrados son los procesos ARIMA(p,d,q). La diferencia de orden d de un proceso ARIMA(p,d,q) da lugar a un proceso estacionario ARMA(p,q).

Consideremos el proceso ARIMA(0,1,1)

$$X_t = X_{t-1} + \delta + \epsilon_t + \theta\epsilon_{t-1}$$

- Predicción

$$\hat{X}_{t+s|t} = s\delta + X_t + \theta\epsilon_t.$$

- Suponiendo $\delta = 0$, $s = 1$

$$\hat{X}_{t+1|t} = X_t + \theta\epsilon_t.$$

Utilizando

$$\epsilon_t = X_t - \hat{X}_{t|t-1}.$$

obtenemos

$$\hat{X}_{t+1|t} = X_t + \theta(X_t - \hat{X}_{t|t-1}).$$

Iterando esta relación se obtiene la predicción correspondiente a un “exponential smoothing”

$$\hat{X}_{t+1|t} = (1 + \theta) \sum_{\tau=0}^{\infty} (-\theta)^{\tau} X_{t-\tau}.$$

heterocedasticidad: ARCH(q)

Este modelo intenta reflejar la estructura temporal de la volatilidad de una serie temporal, proponiendo un modelo autoregresivo para la serie temporal y otro distinto para las innovaciones

$$\begin{aligned}X_t &= \tilde{\phi}^\dagger \cdot \tilde{X}_t^{(m)} + u_t \\u_t &= \sqrt{h_t} \epsilon_t \\h_t &= \tilde{\alpha}^\dagger \cdot [\tilde{u}_t^2]^{(q)},\end{aligned}$$

donde ϵ_t es ruido blanco gaussiano con desviación estándar igual a 1. La condición de no negatividad

$$\alpha_i \geq 0$$

implica que el proceso es estacionario respecto a la covarianza si

$$\sum_{i=1}^q \alpha_i < 1$$

El vector de retrasos para las innovaciones es

$$\left([\tilde{u}_t^2]^{(q)}\right)^\dagger = (1 \ u_{t-1}^2 \ u_{t-2}^2 \ \dots \ u_{t-q}^2)$$

- Volatilidad incondicional

$$\sigma^2 = \mathbf{E} [u_t^2] = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i}$$

- Volatilidad condicional

$$E(u_t^2 | u_{t-1}u_{t-2} \dots u_{t-q}) = \boldsymbol{\alpha}^\dagger \cdot [\mathbf{u}_t^2]^{(q)}$$

- Estimación por máxima verosimilitud Suponiendo que las innovaciones tienen una distribución normal la probabilidad condicional es

$$p(X_t | \mathbf{X}_t^{(\max(q,m))}) = \frac{1}{\sqrt{2\pi h_t}} \exp \left\{ -\frac{(X_t - \tilde{\boldsymbol{\phi}}^\dagger \cdot \tilde{\mathbf{X}}_t^{(m)})^2}{2h_t} \right\},$$

donde

$$h_t = \tilde{\boldsymbol{\alpha}}^\dagger \cdot [\tilde{\mathbf{u}}_t^2]^{(q)}$$

y

$$u_t = X_t - \tilde{\boldsymbol{\phi}}^\dagger \cdot \tilde{\mathbf{X}}_t^{(q)}.$$

La función a optimizar, con las restricciones correspondientes es el logaritmo de la función de verosimilitud

$$\mathcal{LL} = \sum_{t=r}^T \log p(X_t | \mathbf{X}_t^{(r)}), \quad r = \max(q, m).$$

- Ejemplo: AR(1) / ARCH(1)

$$X_t = \phi_0 + \phi_1 X_{t-1} + u_t$$

$$u_t = \sqrt{h_t} \epsilon_t$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2$$

– Volatilidad incondicional:

$$E \left[u_t^2 \right] = \frac{\alpha_0}{1 - \alpha_1}.$$

– Volatilidad condicional:

$$E \left(u_t^2 \mid u_{t-1} \right) = \alpha_0 + \alpha_1 u_{t-1}^2$$

heterocedasticidad: GARCH(p,q)

Este modelo intenta reflejar la estructura temporal de la volatilidad de una serie temporal, proponiendo un modelo autoregresivo para la serie temporal y otro distinto para las innovaciones

$$\begin{aligned}X_t &= \tilde{\phi}^\dagger \cdot \tilde{\mathbf{X}}_t^{(m)} + u_t \\u_t &= \sqrt{h_t} \epsilon_t \\h_t &= \kappa + \beta^\dagger \cdot [\mathbf{h}_t]^{(p)} + \alpha^\dagger \cdot [\mathbf{u}_t^2]^{(q)},\end{aligned}$$

donde ϵ_t es ruido blanco gaussiano con desviación estándar igual a 1. El vector de retrasos para las innovaciones es

$$\left([\mathbf{u}_t^2]^{(q)}\right)^\dagger = (u_{t-1}^2 \ u_{t-2}^2 \ \dots \ u_{t-q}^2).$$

El vector de retrasos para h_t

$$\left([\mathbf{h}_t]^{(p)}\right)^\dagger = (h_{t-1} \ h_{t-2} \ \dots \ h_{t-p}).$$

Las restricciones para los parámetros son

- No-negatividad

$$\kappa > 0; \ \alpha_j \geq 0; \ \beta_j \geq 0$$

- Proceso estacionario (respecto a la covarianza)

$$\sum_{i=1}^r (\alpha_i + \beta_i) < 1; \quad r = \max\{p, q\}$$

- Volatilidad incondicional

$$\sigma^2 = \mathbf{E} [u_t^2] = \frac{\kappa}{1 - \sum_{i=1}^r (\beta_i + \alpha_i)}$$

- Ejemplo: AR(1) / GARCH(1,1)

$$X_t = \phi_0 + \phi_1 X_{t-1} + u_t$$

$$u_t = \sqrt{h_t} \epsilon_t$$

$$h_t = \kappa + \beta h_{t-1} + \alpha u_{t-1}^2$$

- Volatilidad incondicional:

$$E [u_t^2] = \frac{\kappa}{1 - \beta - \alpha}.$$