

# Series de Tiempo: Notas modelos ARIMA

Martín Prado

November 27, 2024

## 1 Repaso de Estimación de modelos ARMA

### 3.4 Función de Autocorelación Parcial (pacf)

Para una serie de tiempo estacionaria  $(X_t)_{t \in \mathbb{Z}}$  de media 0, función de autocovarianza  $\gamma(h) = \mathbf{E}[X_{t+h} X_t] = \mathbf{E}[X_h X_0]$  y función de autocorrelación  $\rho(h) = \gamma(h)/\gamma(0)$ , la función de autocorrelación parcial se define como

$$\alpha(1) = \text{corr}(X_2, X_1) = \rho(1)$$

$$\alpha(k) = \text{corr}(X_{k+1} - P_{\text{sp}\{1, X_2, \dots, X_k\}} X_{k+1}, X_{k+1} - P_{\text{sp}\{1, X_2, \dots, X_k\}} X_1)$$

Por otro lado, si definimos  $\mathcal{H}_n := \overline{\text{sp}}\{X_1, \dots, X_n\}$  y  $\hat{X}_{n+1} := P_{\mathcal{H}_n}(X_{n+1})$ , del sistema de ecuaciones

$$R_k \alpha_k = \rho_k,$$

en donde  $[R_k]_{ij} = \rho(i-j)$ ,  $\rho_k = (\rho(1), \dots, \rho(k))'$ , se sigue que el vector de soluciones  $\alpha_k = (\alpha_{k1}, \dots, \alpha_{kk})'$  satisface  $\alpha(k) = \alpha_{kk}$ .

### 7.2 Estimación de la función de autocorrelación

El estimador de la media es  $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$  y el de la función de autocovarianza es

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n).$$

El estimador de la función de autocorrelación se define como

$$\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0),$$

y el de la matriz de covarianza es

$$\hat{\Gamma}_n = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-2) \\ \vdots & & & \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \cdots & \hat{\gamma}(0) \end{bmatrix}$$

## 8.1 Ecuaciones Yule-Walker

Para un proceso  $\mathbf{AR}(p)$  causal de media 0

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t, \quad Z_t \sim \mathbf{WN}(0, \sigma^2),$$

y función de autocovarianza  $\gamma(h) = \mathbf{E}[X_h X_0]$ , tenemos que

$$\Gamma_p \Phi = \gamma_p,$$

en donde  $[\Gamma_p]_{ij} = \gamma_{ij}(i - j)$ ,  $\Phi = (\phi_1, \dots, \phi_p)'$  y  $\gamma_p = (\gamma(1), \dots, \gamma(p))'$ . En particular, si tomamos los estimados de la función de autocovarianza muestral  $\hat{\gamma}$ , podemos estimar los coeficientes del proceso  $\mathbf{AR}(p)$  como la solución del vector  $\hat{\Phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)'$  en la ecuación

$$\hat{\Gamma}_p \hat{\Phi} = \hat{\gamma}_p, \quad \hat{\gamma}_p = (\hat{\gamma}(1), \dots, \hat{\gamma}(p))'.$$

## 8.3 Estimados de coeficientes de Innovación

Para una muestra  $X_t$  del modelo  $\mathbf{MA}(q)$

$$X_t = Z'_t + \theta_1 Z'_{t-1} + \dots + \theta_q Z'_{t-q}, \quad Z'_t \sim \mathbf{WN}(0, \sigma^2),$$

se puede realizar un ajuste de los coeficientes con el algoritmo de innovaciones para obtener

$$X_t = Z_t + \hat{\theta}_{m1} Z_{t-1} + \dots + \hat{\theta}_{mm} Z_{t-m}, \quad Z_t \sim \mathbf{WN}(0, \hat{v}_m)$$

con  $\hat{\theta}_{ij}$  y  $\hat{v}_m$  definidos a partir de la relación recursiva

$$\hat{\theta}_{m,m-k} = \hat{v}_k^{-1} \left[ \hat{\gamma}(m-k) - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j} \hat{\theta}_{k,k-j} \hat{v}_j \right], \quad k = 0, \dots, m-1,$$

y

$$\hat{v}_m = \hat{\gamma}(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}^2 \hat{v}_j.$$

## 8.4 Estimación de Modelos ARMA

En particular, para procesos  $\mathbf{ARMA}(p, q)$  de media cero y causales

$$\phi(B)X_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} = \theta(B)Z_t,$$

podemos encontrar una función  $\psi$  que satisfice

$$X_t = \psi(B)Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

Los coeficientes  $\psi_j$  satisfacen

$$\psi_0 = 1, \quad \psi_j = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i},$$

y por ende, con la fórmula del capítulo 8.3 podemos estimar  $\psi_1, \dots, \psi_{p+q}$  con  $\hat{\theta}_{m1}, \dots, \hat{\theta}_{m,p+q}$ , para luego obtener estimadores  $\hat{\phi}$  y  $\hat{\theta}$  con la fórmula

$$\hat{\theta}_{mj} = \hat{\theta}_j + \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\theta}_{m,j-i},$$

y también el estimador de la varianza del proceso **WN** con

$$\hat{\sigma}^2 = \hat{v}_m.$$

## 8.7 Estimadores de Máxima Verosimilitud y Mínimos Cuadrados

Para una realización  $X_1, \dots, X_n$  del proceso **ARMA**  $(p, q)$  definido en la sección anterior (con  $Z_t$  Gaussianas), definimos  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ . Para los coeficientes del algoritmo de Innovación  $\theta_{ij}$  que se obtienen con el proceso **ARMA** de parámetros  $\boldsymbol{\phi}$  y  $\boldsymbol{\theta}$ , definimos:

$$\begin{cases} \hat{X}_{i+1} = \sum_{j=1}^i \theta_{ij} (X_{i+1-j} - \hat{X}_{i+1-j}), & 1 \leq i < m = \max(p, q) \\ \hat{X}_{i+1} = \sum_{j=1}^p \phi_j X_{i+1-j} + \sum_{j=1}^i \theta_{ij} (X_{i+1-j} - \hat{X}_{i+1-j}), & i \geq m \end{cases}$$

y también,

$$r_i := \mathbf{E} [(X_{i+1} - \hat{X}_{i+1})^2] / \sigma^2.$$

La función de verosimilitud de los parámetros del proceso **ARMA** es

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} (r_0 \cdots r_{n-1})^{-1/2} \exp \left[ -\frac{1}{2} \sigma^{-2} \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1} \right].$$

Después de derivar con respecto a  $\sigma^2$ , tenemos que el estimador de máxima verosimilitud de la varianza es

$$\tilde{\sigma}^2 := \tilde{\sigma}^2(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}) = n^{-1} S(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}),$$

en donde

$$S(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}) = \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1}$$

y  $\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}$  son los valores de los parámetros  $\boldsymbol{\phi}, \boldsymbol{\theta}$  que minimizan la función

$$l(\boldsymbol{\phi}, \boldsymbol{\theta}) = \ln(n^{-1} S(\boldsymbol{\phi}, \boldsymbol{\theta})) + n^{-1} \sum_{j=1}^n \ln r_{j-1}.$$

## 2 Modelos ARIMA y SARIMA

Recuerde la definición de los operadores  $\nabla^d = (I - B)^d$  y  $\nabla_s^D = (I - B^s)^D$ .

## 9.1 Definición Modelos ARIMA

Para un número natural  $d \geq 0$ , se dice que  $\{X_t\}$  es un proceso **ARIMA**  $(p, d, q)$  (Auto-regressive integrated moving average) si al definir  $Y_t := \nabla^d X_t$  obtenemos un proceso **ARMA**  $(p, q)$

$$\phi(B)Y_t = \phi(B)(1 - B)^d X_t = \theta(B)Z_t, \quad Z_t \sim \mathbf{WN}(0, \sigma^2),$$

con  $\phi$  y  $\theta$  polinomios de grado  $p$  y  $q$  respectivamente.

## 9.6 Definición Modelos SARIMA

Note que si hay un componente estacional en el proceso estocástico, donde cada ciclo o temporada dura  $s$  unidades. Entonces aplicar el operador  $\nabla_s$  debería remover en alguna medida el efecto de este componente como si discutió en el capítulo 1. Sin embargo, el patrón que sigue de un ciclo a otro podría ser aleatorio.

Year	Month			
	1	2	...	12
1	$X_1$	$X_2$	...	$X_{12}$
2	$X_{13}$	$X_{14}$	...	$X_{24}$
3	$X_{25}$	$X_{26}$	...	$X_{36}$
$\vdots$	$\vdots$			$\vdots$
$r$	$X_{1+12(r-1)}$	$X_{2+12(r-1)}$	...	$X_{12+12(r-1)}$

Por ejemplo, la forma en la que se puede comportar una variable aleatoria de un año  $y \in \mathbb{Z}$  hacia atrás podría representar otro proceso **ARMA**  $(P, Q)$  en donde si  $X_y^{(m)} = X_{m+12y}$ , entonces para cada mes  $m = 1, \dots, 12$

$$X_y^{(m)} - \Phi_1 X_{y-1}^{(m)} - \dots - \Phi_P X_{y-P}^{(m)} = U_y^{(m)} + \Theta_1 U_{y-1}^{(m)} + \dots + \Theta_1 U_{y-Q}^{(m)},$$

para  $U_y^{(m)} \sim \mathbf{WN}(0, \sigma_U^2)$ .

Si en particular, después de aplicar  $\nabla^d \nabla_s^D$  para  $d, D \in \mathbb{N}$  obtenemos un proceso **ARMA**  $(p, q)$  causal  $Y_t := (I - B)^d (I - B^s)^D$ , entonces podemos escribir

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \quad Z_t \sim \mathbf{WN}(0, \sigma^2).$$

Si el proceso cumple las condiciones mencionadas anteriormente, entonces decimos que es un proceso **SARIMA**  $(p, d, q) \times (P, D, Q)_s$ .

## 9.2 Técnicas de Identificación

Los desafíos de este capítulo son:

- (a) Hallar una forma de transformar un proceso estocástico  $X_t$  para obtener una serie de tiempo estacionaria y consecuentemente un proceso **ARMA**.

- Aplicar (si es posible) alguna función  $f$  que convierta una serie de tiempo en un proceso **SARIMA**
  - Hallar el periodo  $s$  de componentes estacionales.
  - Hallar  $D$  que al diferenciar con  $\nabla_s^D$  elimine tendencias en el componente estacional.
  - Hallar  $d$  que al diferenciar con  $\nabla_d$  elimine tendencias globales. (9.1)
- (b) Para una serie de tiempo a la cual se le quiere ajustar un proceso **ARMA**  $(p, q)$ , la escogencia de  $p, q$  son fundamentales para la precisión de futuras predicciones
- Escoger  $p, q$  muy grandes puede llevar a sobreajustes, así que se requieren criterios de selección óptimos y eficientes de calcular. (9.3)
- (c) Estimar un modelo ARMA (8.4 y 8.7) y arreglar este ajuste con residuales (9.4)
- Lo preferible es obtener para  $Z_t \sim \mathbf{WN}(0, \hat{\sigma}^2)$  un ajuste  $\phi(B)X_t = \theta(B)Z_t$ . Queremos ver si este es el caso con los residuales
  - Si obtenemos  $\phi(B)X_t = \theta(B)Z_t$ , y los residuales nos indican que  $Z_t$  no es de ruido blanco, pero en cambio existe  $\phi_Z$  y  $\theta_Z$  tal que

$$\phi_Z(B)Z_t = \theta_Z(B)Z'_t, \quad Z'_t \sim \mathbf{WN}(0, \sigma'^2),$$

entonces reajustamos el proceso para obtener los coeficientes correctos del modelo **ARMA** para  $X_t$

$$\phi_Z(B)\phi(B)X_t = \phi_Z(B)\theta(B)Z_t = \theta_Z(B)\theta(B)Z'_t.$$

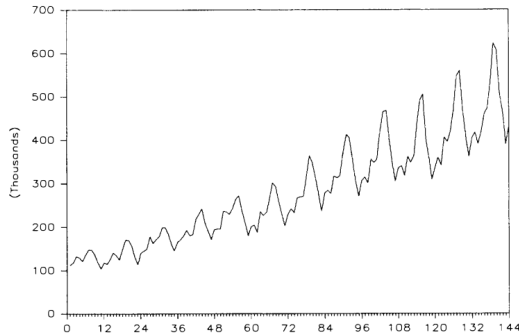


Figure 9.7. International airline passengers; monthly totals in thousands of passengers  $\{U_t, t = 1, \dots, 144\}$  from January 1949 to December 1960 (Box and Jenkins (1970)).

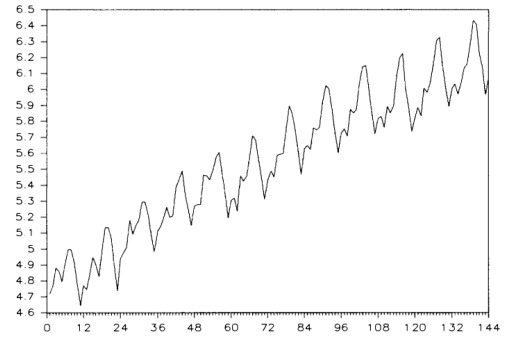


Figure 9.8. Natural logarithms,  $V_t = \ln U_t$ ,  $t = 1, \dots, 144$ , of the data in Figure 9.7.

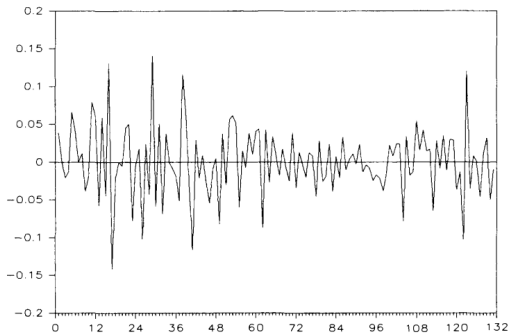
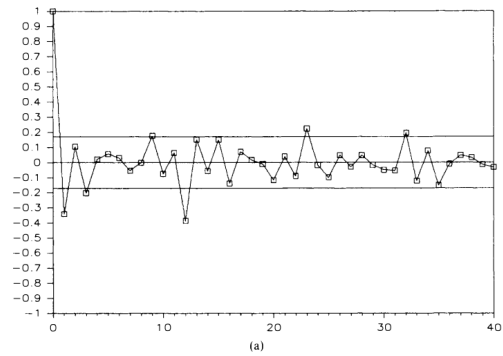


Figure 9.10. The differenced series  $\{\nabla_{12}V_{t+13}\}$  where  $\{V_t\}$  is the data shown in Figure 9.8.



La última gráfica corresponde a la función de autocorrelación de la serie  $\nabla\nabla_{12}V_{t+13}$ .

Si se observa un crecimiento de la varianza a lo largo del tiempo de forma lineal, se puede remover este crecimiento aplicando un logaritmo al proceso.

A partir de la diferenciación con el operador  $\nabla$  se puede eliminar tendencias lineales y con el operador  $\nabla_s$  se puede eliminar estacionalidad ("seasonality") de periodo  $s$  de un proceso estocástico  $X_t$ .

## 9.5 Predicciones de Modelos ARIMA

Sea  $X_t$  un proceso **ARIMA**  $(p, d, q)$ , queremos predecir futuros valores de  $X_t$ . Sin embargo, las herramientas que hemos desarrollado solo sirven para procesos **ARMA**, así que se requieren algunas modificaciones.

Si  $(1 - B)^d X_t =: Y_t$  en donde  $Y_t$  es un modelos **ARMA**  $(p, q)$ , entonces podemos reescribir la ecuación de la siguiente forma

$$X_t = Y_t - \sum_{j=1}^d \binom{d}{j} (-1)^j X_{t-j}. \quad (9.5.1)$$

Si  $X_{1-d}, X_{2-d}, \dots, X_n$  es una realización de  $X_t$ , entonces los valores observados de  $Y_t$  son  $Y_1, \dots, Y_n$ . Defina

$$S_n = \overline{\text{sp}}\{X_{1-d}, X_{2-d}, \dots, X_n\},$$

y note que el mejor predictor lineal de  $X_{n+h}$  basado en las  $n + d$  observaciones de  $X_t$  es  $P_n X_{n+h} := P_{S_n} X_{n+h}$ . Luego, definimos

$$P_n Y_{n+h} := P_{\overline{\text{sp}}\{Y_1, \dots, Y_n\}} Y_{n+h}$$

$$\hat{Y}_{n+1} := P_{\overline{\text{sp}}\{Y_1, \dots, Y_n\}} Y_{n+1} = P_n Y_{n+1}$$

y para continuar, debemos asumir que no hay correlación entre  $X_{1-d}, \dots, X_0$  y  $Y_t$  para  $t > 0$ , de esta forma

$$\overline{\text{sp}}\{X_{1-d}, X_{2-d}, \dots, X_0\} \perp \overline{\text{sp}}\{Y_1, Y_2, \dots, Y_n\},$$

y por ende, obtenemos con (9.5.1) que

$$S_n = \overline{\text{sp}}\{X_{1-d}, \dots, X_0, Y_1, \dots, Y_n\}.$$

Por lo tanto,

$$P_{S_n} Y_{n+h} = P_{S_0} Y_{n+h} + P_n Y_{n+h} = P_n Y_{n+h}. \quad (9.5.2)$$

Luego podemos recursivamente computar  $P_{S_n} X_{n+h}$  con la siguiente fórmula

$$P_{S_n} X_{n+h} = P_n Y_{n+h} - \sum_{j=1}^d \binom{d}{j} (-1)^j P_{S_n} X_{n+h-j}. \quad (9.5.3)$$

Para el MSE "Mean squared error" defina

$$X_{n+1}^* = P_{S_n} X_{n+1},$$

y note que

$$X_{n+1} - X_{n+1}^* = Y_{n+1} - \widehat{Y}_{n+1}$$

Luego, al aplicar innovaciones en  $Y_t$  (ver 5.3.9) obtenemos para  $n > m = \max(p, q)$  y  $h \geq 1$  lo siguiente:

$$\begin{aligned} P_n Y_{n+h} &= \sum_{i=1}^p \phi_i P_n Y_{n+h-i} + \sum_{j=h}^1 \theta_{n+h-1,j} (Y_{n+h-j} - \widehat{Y}_{n+h-j}) \\ &= \sum_{i=1}^p \phi_i P_n Y_{n+h-i} + \sum_{j=h}^1 \theta_{n+h-1,j} (X_{n+h-j} - X_{n+h-j}^*) \end{aligned} \quad (9.5.4)$$

Por lo tanto, mezclando (9.5.2), (9.5.3) y (9.5.4) obtenemos

$$P_{S_n} X_{n+h} = \sum_{j=1}^{p+d} \psi_j^* P_{S_n} X_{n+h-j} + \sum_{j=h}^q \theta_{n+h-1,j} (X_{n+h-j} - X_{n+h-j}^*).$$

Luego, (Problema 9.9) se sigue que

$$\sigma_n^2(h) = \mathbf{E} [(X_{n+h} - P_{S_n} X_{n+h})^2] = \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1,j-r} \right)^2 v_{n+h-j-1}.$$

En donde  $\theta_{n0} = 1$ ,  $\chi(z) = \chi_0 + \chi_1 z + \dots = ((1-z)^d \phi(z))^{-1}$  para  $|z| < 1$ , y

$$v_{n+h-j-1} = \mathbf{E} [(X_{n+h-j} - X_{n+h-j}^*)^2] = \mathbf{E} [(Y_{n+h-j} - \widehat{Y}_{n+h-j})^2]$$

### 3 Ajustes y Pruebas de Ajuste para modelos ARMA

#### 9.3 Escogencia de Orden

Las 3 funciones que califican la verosimilitud de un modelo **ARMA** con parámetros  $(\beta, \sigma^2) = (\phi, \theta, \sigma^2)$  son

$$\begin{aligned} \text{AIC}(\beta) &:= -2 \ln L(\beta, S(\beta)/n) + 2(p + q + 1) \\ \text{AICC}(\beta) &:= -2 \ln L(\beta, S(\beta)/n) + \frac{2(p + q + 1)n}{(n - p - q - 2)} \\ \text{BIC}(\beta) &:= (n - p - q) \ln \left[ \frac{S(\beta)}{n - p - q} \right] + n(1 + \ln(\sqrt{2\pi})) \\ &\quad + (p + q) \ln \left[ \frac{\sum_{t=1}^n X_t^2 - S(\beta)}{p + q} \right], \end{aligned}$$

Para estimar  $\widehat{p}, \widehat{q}$  óptimos, calculamos el modelo de máxima verosimilitud  $\tilde{\beta}$  como en (8.7) para diferentes valores de  $p, q$  y evaluamos en alguna de estas 3 funciones hasta encontrar el que la minimiza. De acuerdo a la discusión de 9.3 las funciones tienen pros y contras al ser usadas como criterios:

- (1) AIC y AICC son estimadores de un estadístico llamado "Kullback-Leibler index" al cual al ser minimizado se maximiza la función de verosimilitud.
- (2) AIC y AICC son equivalentemente asintóticos. AIC tiende a sobreestimar  $\mathbf{p}$ , pero, para modelos de gran orden es mejor que AICC.
- (3) Entre los 3, el único que produce estadísticos consistentes  $\hat{\mathbf{p}} \rightarrow \mathbf{p}$ ,  $\hat{\mathbf{q}} \rightarrow \mathbf{q}$  es BIC. Sin embargo, a diferencia de AIC y AICC, el cálculo de BIC no es asintóticamente eficiente para modelos autorregresivos.

## 9.4 Test de Residuales

Para un modelo **ARMA**  $X_t$  con parámetro  $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$  se observan datos  $X_1, \dots, X_n$ . A partir de estas observaciones se calculan predicciones en base a un parámetro  $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ . Para las predicciones en base a los parámetros reales  $\hat{X}_{n+1} = \hat{X}_{n+1}(\boldsymbol{\phi}, \boldsymbol{\theta}) = P_n X_{n+1}$ , los residuos "reales":

$$W_n(\boldsymbol{\phi}, \boldsymbol{\theta}) = (X_n - \hat{X}_n) / \sqrt{r_{n-1}(\boldsymbol{\phi}, \boldsymbol{\theta})}$$

debería comportarse, de acuerdo con la discusión del capítulo 8, como ruido blanco. De la misma forma, para los residuos de los parámetros estimados

$$\widehat{W}_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) = (X_n - \hat{X}_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})) / \sqrt{r_{n-1}(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})},$$

debería suceder lo mismo. En particular,  $\mathbf{E}[(W_n(\boldsymbol{\phi}, \boldsymbol{\theta}) - Z_n)^2] \rightarrow 0$  cuando  $n \rightarrow \infty$  y por lo tanto,

- (1) Si  $Z_t \sim \mathbf{WN}(0, \sigma^2)$ , entonces el proceso  $\widehat{W}_n$  debería tener covarianzas pequeñas.
- (2) Si  $Z_t \sim \text{IID}(0, \sigma^2)$ , entonces el proceso  $\widehat{W}_n$  es independiente.
- (3) Si  $Z_t \sim \mathcal{N}(0, \sigma^2)$ , entonces la distribución del proceso  $\widehat{W}_n$  debería parecerse a una gaussiana.

La función de autocovarianzas muestrales de  $\widehat{W}_t$  se define de la siguiente forma

$$\hat{\rho}_W(h) = \frac{\sum_{t=1}^{n-h} (\widehat{W}_t - \overline{W})(\widehat{W}_{t+h} - \overline{W})}{\sum_{t=1}^n (\widehat{W}_t - \overline{W})^2},$$

en donde  $\overline{W} = n^{-1} \sum_{t=1}^n \widehat{W}_t$ .



Let  $\hat{\rho}_w = (\hat{\rho}_w(1), \dots, \hat{\rho}_w(h))'$  where  $h$  is a fixed positive integer. If  $\{X_t\}$  is the causal invertible ARMA process  $\phi(B)X_t = \theta(B)Z_t$ , define

$$\tilde{\phi}(z) = \phi(z)\theta(z) = 1 - \tilde{\phi}_1 z - \dots - \tilde{\phi}_{p+q} z^{p+q}, \quad (9.4.2)$$

and

$$a(z) = (\tilde{\phi}(z))^{-1} = \sum_{j=0}^{\infty} a_j z^j. \quad (9.4.3)$$

It will be convenient also to define  $a_j = 0$  for  $j < 0$ .

Assuming  $h \geq p + q$ , set

$$\begin{aligned} T_h &= [a_{i-j}]_{1 \leq i \leq h, 1 \leq j \leq p+q}, \\ \tilde{\Gamma}_{p+q} &= \left[ \sum_{k=0}^{\infty} a_k a_{k+|i-j|} \right]_{i,j=1}^{p+q}, \end{aligned} \quad (9.4.4)$$

and

$$Q = T_h \tilde{\Gamma}_{p+q}^{-1} T_h' = [q_{ij}]_{i,j=1}^h. \quad (9.4.5)$$

Note that  $\tilde{\Gamma}_{p+q}$  is the covariance matrix of  $(Y_1, \dots, Y_{p+q})$  where  $\{Y_t\}$  is an AR( $p + q$ ) process with autoregressive polynomial given by  $\tilde{\phi}(z)$  in (9.4.2) and with  $\sigma^2 = 1$ . Then using the argument given in Box and Pierce (1970), it can be shown that

$$\hat{\rho}_w \text{ is AN}(0, n^{-1}(I_h - Q)), \quad (9.4.6)$$

where  $I_h$  is the  $h \times h$  identity matrix. The asymptotic variance of  $\hat{\rho}_w(i)$  is thus  $n^{-1}(1 - q_{ii})$ .

*The Portmanteau Test.* Instead of checking to see if each  $\hat{\rho}_w(i)$  falls within the confidence bounds  $\pm 1.96(1 - q_{ii})^{1/2} n^{-1/2}$ , it is possible to consider instead a single statistic which depends on  $\hat{\rho}_w(i)$ ,  $1 \leq i \leq h$ . Throughout this discussion  $h$  is assumed to depend on the sample size  $n$  in such a way that (i)  $h_n \rightarrow \infty$  as  $n \rightarrow \infty$ , and (ii) the conditions of Box and Pierce (1970) are satisfied, namely

- (a)  $\psi_j = O(n^{-1/2})$  for  $j \geq h_n$  where  $\psi_j$ ,  $j = 0, 1, \dots$  are the coefficients in the expansion  $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ , and
- (b)  $h_n = O(n^{1/2})$ .

Then since  $h_n \rightarrow \infty$ , the matrix  $\tilde{\Gamma}_{p+q}$  may be approximated by  $T_h' T_h$  and so the matrix  $Q$  in (9.4.5) and (9.4.6) may be approximated by the projection matrix (see Remark 2 of Section 2.5),

$$T_h(T_h' T_h)^{-1} T_h',$$

which has rank  $p + q$ . Thus if the model is appropriate, the distribution of  $\hat{\boldsymbol{\rho}}_w = (\hat{\rho}_w(1), \dots, \hat{\rho}_w(h))'$  is approximately  $N(\mathbf{0}, n^{-1}(I_h - T_h(T_h' T_h)^{-1} T_h'))$ . It then follows from Problem 2.19 that the distribution of

$$Q_w = n \hat{\boldsymbol{\rho}}_w' \hat{\boldsymbol{\rho}}_w = n \sum_{j=1}^h \hat{\rho}_w^2(j)$$

is approximately chi-squared with  $h - (p + q)$  degrees of freedom. The adequacy of the model is therefore rejected at level  $\alpha$  if

$$Q_w > \chi_{1-\alpha}^2(h - p - q).$$

## Kullback-Leibler Index

If  $\mathbf{X}$  is an  $n$ -dimensional random vector whose probability density belongs to the family  $\{f(\cdot; \psi), \psi \in \Psi\}$ , the Kullback–Leibler discrepancy between  $f(\cdot; \psi)$  and  $f(\cdot; \theta)$  is defined as

$$d(\psi | \theta) = \Delta(\psi | \theta) - \Delta(\theta | \theta),$$

where

$$\begin{aligned} \Delta(\psi | \theta) &= E_\theta(-2 \ln f(\mathbf{X}; \psi)) \\ &= \int_{\mathbb{R}^n} -2 \ln(f(\mathbf{x}; \psi)) f(\mathbf{x}; \theta) d\mathbf{x}, \end{aligned}$$

is the Kullback–Leibler index of  $f(\cdot; \psi)$  relative to  $f(\cdot; \theta)$ . (In general  $\Delta(\psi | \theta) \neq \Delta(\theta | \psi)$ .) Applying Jensen's inequality, we see that

$$\begin{aligned} d(\psi | \theta) &= \int_{\mathbb{R}^n} -2 \ln\left(\frac{f(\mathbf{x}; \psi)}{f(\mathbf{x}; \theta)}\right) f(\mathbf{x}; \theta) d\mathbf{x} \\ &\geq -2 \ln\left(\int_{\mathbb{R}^n} \frac{f(\mathbf{x}; \psi)}{f(\mathbf{x}; \theta)} f(\mathbf{x}; \theta) d\mathbf{x}\right) \\ &= -2 \ln\left(\int_{\mathbb{R}^n} f(\mathbf{x}; \psi) d\mathbf{x}\right) \\ &= 0 \end{aligned}$$

with equality holding if and only if  $f(\mathbf{x}; \psi) = f(\mathbf{x}; \theta)$  a.e.  $[f(\cdot, \theta)]$ .

Suppose therefore that our observations  $X_1, \dots, X_n$  are from a Gaussian ARMA process with parameter vector  $\theta = (\boldsymbol{\beta}, \sigma^2)$  and assume for the moment that the true order is  $(p, q)$ . Let  $\hat{\theta} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$  be the maximum likelihood estimator of  $\theta$  based on  $X_1, \dots, X_n$  and let  $Y_1, \dots, Y_n$  be an independent realization of the true process (with parameter  $\theta$ ). Then

$$-2 \ln L_Y(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = -2 \ln L_X(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) + \hat{\sigma}^{-2} S_Y(\hat{\boldsymbol{\beta}}) - n,$$

so that

$$\begin{aligned} E_{\theta}(\Delta(\hat{\theta}|\theta)) &= E_{\boldsymbol{\beta}, \sigma^2}(-2 \ln L_Y(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)) \\ &= E_{\boldsymbol{\beta}, \sigma^2}(-2 \ln L_X(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)) + E_{\boldsymbol{\beta}, \sigma^2}\left(\frac{S_Y(\hat{\boldsymbol{\beta}})}{\hat{\sigma}^2}\right) - n. \end{aligned} \quad (9.3.3)$$

Making the local linearity approximation used in Section 8.11, we can write, for large  $n$ ,

$$\begin{aligned} S_Y(\hat{\boldsymbol{\beta}}) &\simeq S_Y(\boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \frac{\partial S_Y}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) + \frac{1}{2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \left[ \frac{\partial^2 S_Y(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j} \right]_{i,j=1}^n (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &\simeq S_Y(\boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' 2 \sum_{i=1}^n \frac{\partial Z_i}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) Z_i(\boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' D' D (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \end{aligned}$$

From Section 8.11, we know that  $n^{-1} D' D \xrightarrow{P} \sigma^2 V^{-1}(\boldsymbol{\beta})$ ,  $\hat{\boldsymbol{\beta}}$  is  $\text{AN}(\boldsymbol{\beta}, n^{-1} V(\boldsymbol{\beta}))$ , and that  $(\partial Z_i / \partial \boldsymbol{\beta})(\boldsymbol{\beta}) Z_i(\boldsymbol{\beta})$  has mean 0. Replacing  $D' D$  by  $n \sigma^2 V^{-1}(\boldsymbol{\beta})$  and assuming that  $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$  has covariance matrix  $V(\boldsymbol{\beta})$ , we obtain

$$\begin{aligned} E_{\boldsymbol{\beta}, \sigma^2}[S_Y(\hat{\boldsymbol{\beta}})] &\simeq E_{\boldsymbol{\beta}, \sigma^2}[S_Y(\boldsymbol{\beta})] + \sigma^2 E_{\boldsymbol{\beta}, \sigma^2}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' V^{-1}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ &\simeq \sigma^2 n + \sigma^2(p + q), \end{aligned}$$

since  $(\partial Z_i / \partial \boldsymbol{\beta})(\boldsymbol{\beta}) Z_i(\boldsymbol{\beta})$  is independent of  $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$  and  $E(\mathbf{U}' \boldsymbol{\Sigma}^{-1} \mathbf{U}) = \text{trace}(\boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1}) = k$  for any zero-mean random  $k$ -vector  $\mathbf{U}$  with nonsingular covariance matrix  $\boldsymbol{\Sigma}$ . From the argument given in Section 8.9,  $n \hat{\sigma}^2 = S_X(\hat{\boldsymbol{\beta}})$  is distributed approximately as  $\sigma^2 \chi^2(n - p - q)$  for large  $n$  and is asymptotically independent of  $\hat{\boldsymbol{\beta}}$ . With the independence of  $\{X_1, \dots, X_n\}$  and  $\{Y_1, \dots, Y_n\}$ , this implies that  $\hat{\sigma}^2$  is asymptotically independent of  $S_Y(\hat{\boldsymbol{\beta}})$ .

Consequently,

$$\begin{aligned}
E_{\boldsymbol{\beta}, \sigma^2} \left( \frac{S_Y(\hat{\boldsymbol{\beta}})}{\hat{\sigma}^2} \right) - n &\simeq \sigma^2(n + p + q)(E_{\boldsymbol{\beta}, \sigma^2} \hat{\sigma}^{-2}) - n \\
&\simeq \sigma^2(n + p + q) \left( \sigma^2 \frac{n - p - q - 2}{n} \right)^{-1} - n \\
&= \frac{2(p + q + 1)n}{n - p - q - 2}.
\end{aligned}$$

Thus the quantity,  $-2 \ln L_X(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) + 2(p + q + 1)n/(n - p - q - 2)$ , is an approximately unbiased estimate of the expected Kullback–Leibler index  $E_\theta(\Delta(\hat{\theta}|\theta))$  in (9.3.3). Since the preceding calculations (and the maximum likelihood estimators  $\hat{\boldsymbol{\beta}}$  and  $\hat{\sigma}^2$ ) are based on the assumption that the true order is  $(p, q)$ , we therefore select the values of  $p$  and  $q$  for our fitted model to be those which minimize  $\text{AICC}(\hat{\boldsymbol{\beta}})$ , where

$$\text{AICC}(\boldsymbol{\beta}) := -2 \ln L_X(\boldsymbol{\beta}, S_X(\boldsymbol{\beta})/n) + 2(p + q + 1)n/(n - p - q - 2). \quad (9.3.4)$$