Regresión con errores autocorrelacionados

Series de tiempo

Gerardo Ortega
Miguel Pluma
Luis Osorio
Johnatan García

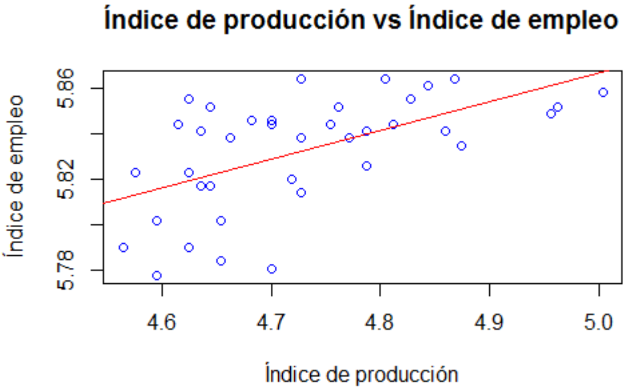
09 de diciembre de 2013

Contenido

Introducción

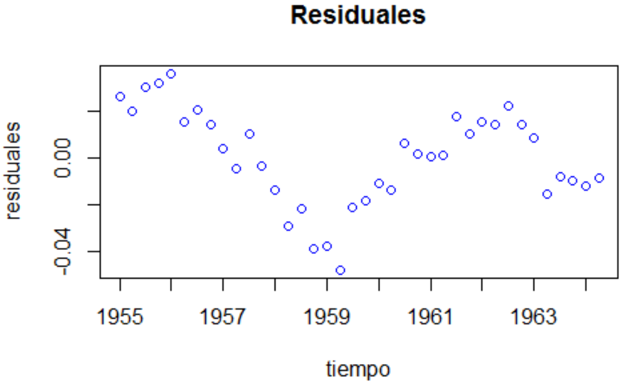
En el análisis de mínimos cuadrados, el modelo usual de regresión es

$$Y_t = \beta_0 + \sum_{i=1}^{r} \beta_i X_{it} + \epsilon_t, \;\; t = 1, 2, \ldots, n.$$



1

[1]Ejemplo 9.8.1, Fuller,pág. 557-558


Residuales
residuales
0.00
-0.04
1955
1957
1959
1961
1963
tiempo

Considere el modelo

$$y_t = \sum_{i=1}^{m} \beta_i x_{it} + e_t, , \;\; t = 1, 2, \ldots, n,$$

donde los errores no son independientes.
En forma matricial,

$$y = X\beta + e.$$

Idea intuitiva

Se quiere encontrar matriz de transformación A, tal que

$$\begin{aligned} AY &= A(X\beta + e) \\ &= AX\beta + Ae \\ &= U\beta + w \end{aligned}$$

donde $U = AX$, $A\Gamma A' = \sigma^2 I$ y $W \sim WN(0, \sigma^2)$.

Aplicando mínimos cuadrados o máxima verosimilitud al vector Ay se obtiene

$$\begin{aligned}
\hat{\beta}_w &= (U'U)^{-1}U'y \\
&= (X'A'AX)^{-1}X'A'Ay \\
&= (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}y.
\end{aligned}$$

Si $\{e_t\}$ son un proceso $AR(p)$, entonces

$$\phi(B)e_t = w_t.$$

De esta manera

$$u_t = \phi(B)y_t = \beta'\phi(B)x_t + w_t = \beta' v_t + w_t.$$

Supuestos

Si $\{e_t\}$ son ARMA(p,q), entonces

$$e_t = \sum_{j=1}^{p} \phi_j e_{t-j} - \sum_{k=1}^{q} \theta_k w_{t-k} + w_t, \;\; w_t \sim WN(0, \sigma^2).$$

Con el operador de retardo, B, se puede escribir como

$$\phi(B)e_t = \theta(B)w_t \;\; \text{o} \;\; e_t = \phi^{-1}(B)\theta(B)w_t = \frac{\theta(B)}{\phi(B)}w_t,$$

donde $\phi(B) = 1 - \sum \phi_j B^j, \;\; \theta(B) = 1 - \sum \theta_k B^k$, son los polinomios en B de grado p y q, respectivamente.

Por lo tanto,

$$\begin{aligned} y_t &= \sum_{i=1}^{m} \beta_i x_{it} + e_t \\ &= \sum_{i=1}^{m} \beta_i x_{it} + \frac{\theta(B)}{\phi(B)} w_t. \end{aligned}$$

Algoritmo

Paso 1

Aplica una regresión ordinaria de y_t con e_t.

Algoritmo

Paso 1

Aplica una regresión ordinaria de y_t con e_t.

Paso 2

Ajustar un modelo ARMA a los residuales $\hat{e}_t = y_t - \hat{\beta}' x_t$ y

$$\hat{\Phi}(B)\hat{e}_t = \hat{\theta}(B) w_t.$$

Paso 3

Aplicar la transformación ARMA a

$$y_t = \beta' x_t + e_t,$$

si

$$u_t = \frac{\hat{\Phi}(B)}{\hat{\theta}(B)} y_t \text{ y } v_t = \frac{\hat{\Phi}(B)}{\hat{\theta}(B)} x_t,$$

se obtiene el modelo de regresión transformado

$$U = V\beta + w.$$

Paso 4

Aplicando mínimos cuadrados en una regresión ordinaria, asumiendo errores no correlacionados al modelo transformado, obteniendo

$$\hat{\beta}_w = \left(V'V\right)^{-1} V'U,$$

donde $V = [v_1, ..., v_n]'$ y $U = [U_1, ..., U_n]'$ son los correspondientes componentes transformados.

Paso 4

Aplicando mínimos cuadrados en una regresión ordinaria, asumiendo errores no correlacionados al modelo transformado, obteniendo

$$\hat{\beta}_w = \left(V'V\right)^{-1} V'U,$$

donde $V = [v_1, ..., v_n]'$ y $U = [U_1, ..., U_n]'$ son los correspondientes componentes transformados.

Paso 5

El procedimiento se repite hasta convergencia y se aproxima al EMV bajo normalidad de los errores.

Estimadores

Dado el modelo queremos estimar los parámetros. Existen varios métodos de estimación: OLS,GLS, EGLS,Least Square Residuals

Estimación

$$\begin{aligned} \hat{\beta}_{OLS} &= (X'X)^{-1}X'\tilde{y}, \\ \hat{\beta}_{G} &= (X'V^{-1}X)^{-1}X'V^{-1}\tilde{y}, \\ \hat{\beta}_{E} &= (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\tilde{y}, \end{aligned}$$

donde V es la matriz de varianzas y covarianzas de $\tilde{e}$.

Propiedades

Propiedad 1

Los estimadores OLS son consistentes, asintóticamente normales. Pueden no ser eficientes o tener tasas de convergencia distintas.

- Puede no ser insesgado.
- Las coordenadas de $\hat{\beta}$ puede converger a diferentes tasas.
- Se subestima la verdadera varianza.

Propiedades

Propiedad 1

Los estimadores OLS son consistentes, asintóticamente normales. Pueden no ser eficientes o tener tasas de convergencia distintas.

- Puede no ser insesgado.
- Las coordenadas de $\hat{\beta}$ puede converger a diferentes tasas.
- Se subestima la verdadera varianza.

Propiedad 2

El mejor estimador lineal insesgado es $\hat{\beta}_G$, ademas de consistente y asintóticamente normal. Aunque se requiere conocer V.

Propiedad 3

Los residuales de OLS se pueden usar para estimar $Y_t - X_t\hat{\beta} = \hat{e}_t$.

Propiedad 3

Los residuales de OLS se pueden usar para estimar $Y_t - X_t\hat{\beta} = \hat{e}_t$.

Propiedad 4

Un estimador de GLS de β se obtiene reemplazando V por $\hat{V} = V(\alpha)$. El estimador es consistente, asintóticamente normal y asintóticamente equivalente a GLS.

Estimador de la ACF

Se puede demostrar que

$$\widehat{\gamma}_{\widehat{e}_t}(h) = \widehat{\gamma}_{e_t}(h) + O_p(n^{-1}),$$

es decir $\widehat{\gamma}_{\widehat{e}_t}(h)$ y $\widehat{\gamma}_{e_t}(h)$ son asintóticamente equivalentes y como

$$\widehat{\gamma}_{e_t}(h) \to \gamma_{e_t}(h) \qquad h \geq 1,$$

entonces

$$\widehat{\gamma}_{\widehat{e}_t}(h) \to \gamma_{e_t}(h) \qquad h \geq 1.$$

Estadístico de Durbin Watson

Para investigar la naturaleza del sesgo en $\widehat{\gamma}_{\widehat{e}_t}(h)$ se definió el estadístico de Durbin Watson

$$
\begin{aligned}
d &= \frac{\sum_{i=1}^{n-1}(\widehat{e}_{i+1} - \widehat{e}_i)^2}{\sum_{i=1}^{n} \widehat{e}_i^2} \\
&= 2(1 - \widehat{\rho}_{\widehat{e}_t}(1)) - \frac{\widehat{e}_1^2 + \widehat{e}_n^2}{\sum_{i=1}^{n} \widehat{e}_t^2}.
\end{aligned}
$$

Se espera que $d \Rightarrow 2$, si no es así, se sospecha que e_t no es ruido blanco.

- El estadístico d está probando $\rho_{e_t}(1) = 0$, es posible que se cumpla $\rho_{e_t}(1) = 0$ pero no necesariamente e_t es ruido blanco, por ejemplo

$$e_t = \alpha e_{t-2} + w_t \quad w_t \sim \mathrm{WN}(0, \sigma^2).$$

- $$H_0: \quad e_t \quad \text{es ruido blanco} \quad \text{vs} \quad H_1 \; \rho_{e_t}(1) > 0,$$
$$H_0: \quad e_t \quad \text{es ruido blanco} \quad \text{vs} \quad H_2 \; \rho_{e_t}(1) < 0.$$

  Bajo H_0 se tiene que

$$\sqrt{n}(d-2) = 2\widehat{\gamma}_{\hat{e}_t}(1)\sqrt{n} + O_p(n^{-\frac{1}{2}}) \to N(0,4).$$

- La distribución exacta de d depende de X, sin embargo Durbin y Watson encontraron dos variables

$$d_{L,\alpha} \leq d \leq d_{U,\alpha}$$

que no dependen de X y tabularon sus percentiles.

$$H_0: \quad e_t \quad \text{es ruido blanco} \quad \text{vs} \quad H_1: \; \rho_{e_t}(1) > 0$$

- Si $d < d_{L,\alpha}$, rechazar H_0 a favor de H_1.
- Si $d > d_{U,\alpha}$, no rechazar H_0.
- Si $d_{L,\alpha} \leq d \leq d_{U,\alpha}$ la prueba no puede concluir.

$$H_0: \quad e_t \quad \text{es ruido blanco} \quad \text{vs} \quad H_2: \; \rho_{e_t}(1) < 0$$

- Si $4 - d < d_{L,\alpha}$, rechazar H_0 a favor de H_1.
- Si $4 - d > d_{U,\alpha}$, no rechazar H_0.
- Si $d_{L,\alpha} \leq 4 - d \leq d_{U,\alpha}$ la prueba no puede concluir.

Predicción

Si tenemos $Y_t = \tilde{X}_t\'\tilde{\beta} + e_t$ con e_t estacionaria de media 0 y función de autocovarianza $\Gamma_e(h)$ conocida y $\tilde{\beta}$ desconocida, queremos además de estimar $\tilde{\beta}$, predecir $Y_{n+1}, ..., Y_{n+s}$ dados $Y_1, ..., Y_n$, $\tilde{X}_{n+1}, ..., \tilde{X}_{n+s}$ y $\Gamma_e(h)$.

Sabemos que:

$$V = E[\tilde{e}\tilde{e}'] = \begin{bmatrix} \Gamma(0) & \Gamma(1) & ,..., & \Gamma(n-1) \\ \Gamma(1) & \Gamma(0) & ,.., & \Gamma(n-2) \\ \vdots & & \ddots & \vdots \\ \Gamma(n-1) & \Gamma(n-2) & ,.., & \Gamma(0) \end{bmatrix},$$

$\tilde{e} = (e_1, ..., e_n)'$.

Usando el modelo $Y = \tilde{X}\tilde{\beta} + \tilde{e}_t$, $\tilde{e}_t \sim (0, V)$, y el mejor estimador lineal insesgado para $\tilde{\beta}$, el cual esta dado por,

$$\hat{\beta} = (X\prime V^{-1}X)^{-1}X\prime V^{-1}\tilde{y},$$

(Estimador de minimos cuadrados generalizados), encontraremos es mejor predictor lineal insesgado.

Consideremos a $\tilde{c}_i'\tilde{y}$ como dicho predictor de Y_{n+i}, veamos que es insesgado:

$$
\begin{aligned}
E(\tilde{c}_i'\tilde{y}) &= E(\tilde{c}_i'\left(\tilde{X}\tilde{\beta} + \tilde{e}_t\right)) = \tilde{c}_i'\tilde{X}\tilde{\beta} = E[Y_{n+i}] = \tilde{X}_{n+i}\tilde{\beta}\ \forall\tilde{\beta} \\
&\Longrightarrow \tilde{c}_i'\tilde{X}\tilde{\beta} = \tilde{X}_{n+i}\tilde{\beta} \\
&\Longrightarrow \tilde{c}_i'\tilde{X} = \tilde{X}_{n+i}'.
\end{aligned}
$$

Ahora, para que quede determinado el predictor, debemos de decir quien es $\tilde{c}_i$, y para ello debe cumplir

$$\min\left(E\left[Y_{n+1}-\tilde{c}_i'\tilde{y}\right]^2\right),$$

sujeto a $\tilde{c}_i\tilde{x}=\tilde{x}_{n+i}'$.

Entonces para Y_{n+i},

$$\begin{aligned}
\hat{Y}_{n+i} &= \tilde{c}_i' \tilde{Y} \\
&= \tilde{b}_i' \tilde{Y} + \left[\tilde{X}_{n+i} - X'\tilde{b}_i\right] (X'V^{-1}X)^{-1} X'V^{-1} \tilde{y} \\
&= \tilde{b}_i' \tilde{Y} + \left[\tilde{X}_{n+i} - X'\tilde{b}_i\right] \hat{\beta} \\
&= \tilde{b}_i' \left[\tilde{Y} - X'\hat{\beta}\right] + \tilde{X}_{n+i}' \hat{\beta}.
\end{aligned}$$

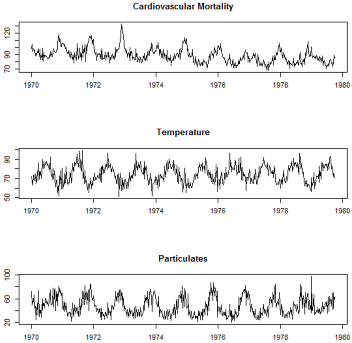
Y por tanto, el error de predicción esta dado por,

$$
\begin{aligned}
y_{n+i} - \hat{y}_{n+i} &= \tilde{X}'_{n+i}(\tilde{\beta} - \hat{\beta}) - \tilde{b}'_i \left[\tilde{y} - x\hat{\beta} \right] + e_{n+i} \\
&= \tilde{X}'_{n+i}(\tilde{\beta} - \hat{\beta}) - \tilde{b}'_i \tilde{e} - \tilde{b}'_i \tilde{X} (\tilde{\beta} - \hat{\beta}) + e_{n+i} \\
&= \left[\tilde{b}'_i \tilde{X} - \tilde{X}'_{n+i} \right] (\tilde{\beta} - \hat{\beta}) + (e_{n+i} - \hat{e}_{n+i}) \, .
\end{aligned}
$$

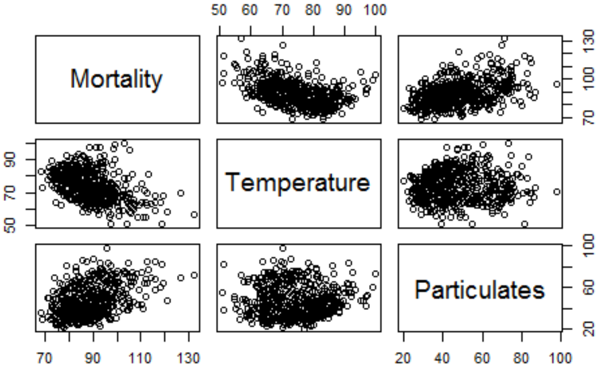
Regresión con errores corelacionados

- Analizaremos datos de muertes por daños cardiovasculares.
- Variables regresoras: tiempo, temperatura, nivel de partículas de contaminantes.
- Dichos datos se encuentran en el libro de Shumway

Datos


Cardiovascular Mortality
120
90
70
1970
1972
1974
1976
1978
1980
Temperature
90
70
50
1970
1972
1974
1976
1978
1980
Particulates
100
60
20
1970
1972
1974
1976
1978
1980

Graficos de dispersión



Gráficos de dispersión

Las graficás anteriores sugieren el siguiente modelo de regresión:

$$M_t = \beta_1 + \beta_2 t + \beta_3 T_t + \beta_4 T_t^2 + \beta_5 P_t + Z_t \tag{1}$$

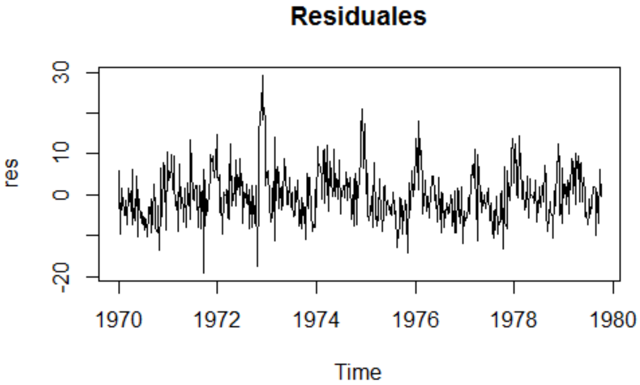
donde

- T_t: es la temperatura al tiempo t,
- P_t: es el nivel de particulas contaminantes al tiempo t,
- Z_t: son los errores (los cuales resultan ser correlacionados)
- β_i: son los coeficientes a determinar en la regresión.

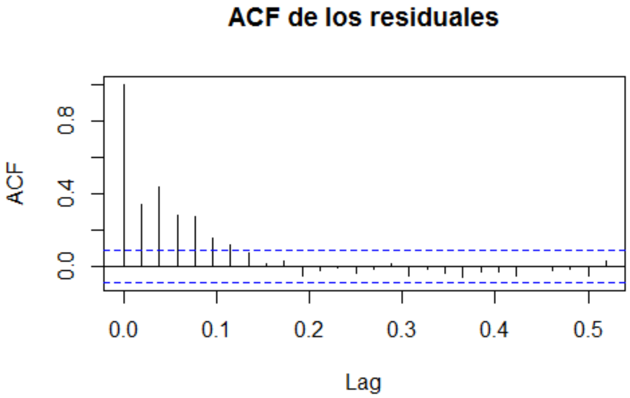
Errores estimados

Los errores estimados, que son los residuales con el modelo clásico de regresión, se muestran en la siguiente figura. Es claro que no es ruido blanco. Además de la prueba de Durbin Watson obtenemos el estadístico

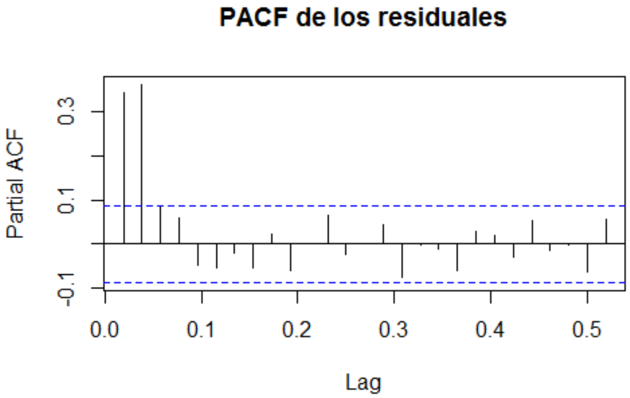
$$d = DW = 1.3109, \quad \text{con } p - valor = 1.523e - 15$$


Residuales
res
30
10
0
-20
1970
1972
1974
1976
1978
1980
Time

ACF de los errores estimados



PACF de los errores estimados

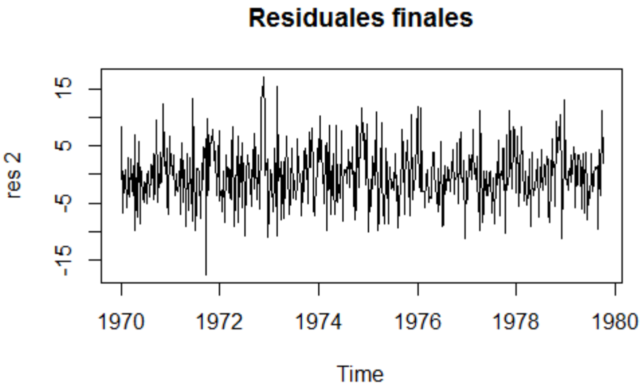


Modelo propuesto para los errores Z_t

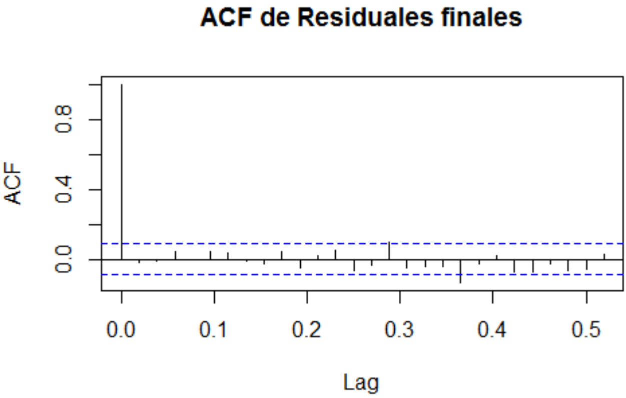
Los gráficos anteriores sugieren que los errores son un modelo $AR(2)$.
Estimando los coeficientes del $AR(2)$ y de la regresión se tiene

β_1	β_2	β_3	β_4	β_5	α_1	α_2
0.3848	0.4326	3075.1482	-1.5165	-0.0190	0.0154	0.1545

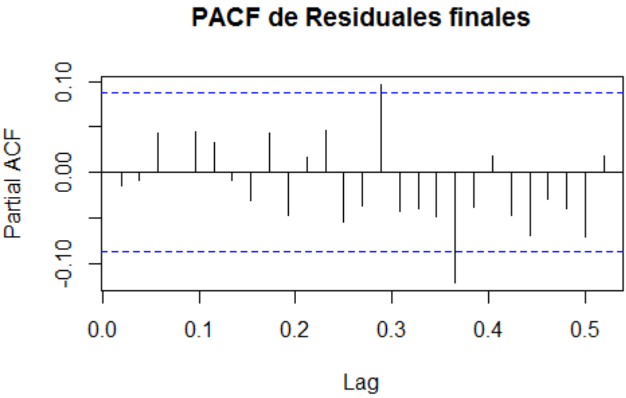
Residuales finales



ACF de los residuales finales



PACF de los residuales finales



PACF de los residuales finales

Además de la prueba de Durbin Watson obtenemos el estadístico

$$d = 2.0152, \quad \text{con } p - valor = 0.5538$$

Dicho valor, junto con las gráficas, muestran que hay un buen ajuste.

Ejercicio: Les enviaremos un archivo con datos de consumo bebidas alcohólicas en el Reino Unido de 1870 a 1938. La primera variable Y_t(primera columna) corresponde al consumo anual percápita, las variables explicativas son: $\phi_{1,t}$ (segunda columna) y $\phi_{2,t}$ (tercera columna) las cuales representan el ingreso percápita y el precio de las bebidas, respectivamente. El modelo que utilizaremos es:

$$Y_t = \beta_0 + \beta_1\phi_{1,t} + \beta_2\phi_{2,t} + \beta_3\phi_{3,t} + \beta_4\phi_{4,t} + Z_t$$

donde $\phi_{3,t} = 10^{-2}t$, $\phi_{4,t} = 10^{-4}(t-35)^2$, los valores para t son a partir de 1870, y asumimos Z_t es una serie de tiempo estacionaria. Deben estimar los parámetros de $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$, siguiendo el ejemplo mostrado en clase.

Bibliografía I

R. L. Anderson.
The problem of autocorrelation in regression analysis.
American Statistical Association, 49:113–29, 1954.

Frantisek atulajter.
Predictions in Time Series using Regression models.
Springer New York, 2002.

M .S. Bartlett.
On the theoretical specification and sampling properties of autocorrelated time series.
Royal Statistical Society, B 8:27–41, 1946.

Bibliografía II

J. Durbin.
Estimation of parameters in time-series regression model.
Royal Statitistal Society, B 22:139–53, 1960.

Dra. Graciela González Farías.
Notas.

W. A. Fuller.
Introduction to Statistical Time Series.
Wiley & Sons Inc, 2 edition, 1996.

Bibliografía III

D. A. Pierce.
Distribution of residual autocorrelations in the regression model with auto-regressive-moving average errors.
Royal Statistical Society, B 23, 1971.

Dr. Joaquín Ortega Sánchez.
Notas del curso.

Robert H. Shumway & David S. Stoffer.
Time series analysis and its applications.
Springer, 2000.