

Regresión con errores correlacionados, predicción con errores correlacionados y variables rezagadas.

María Guadalupe López Salomón¹ and José Sebastián Ramírez Jiménez¹

Keywords: Regression, Correlated Errors, Time Series, Prediction, Lagged Variables

Resumen

Este trabajo se enfoca en aspectos críticos de la regresión y la predicción en el contexto de errores autocorrelacionados y el uso de variables rezagadas. En el análisis de series temporales y datos longitudinales, la correlación entre errores y la inclusión de valores pasados de las variables son aspectos que no pueden ser ignorados. Primero, abordamos la regresión con errores correlacionados, destacando cómo estos pueden comprometer la eficiencia y la precisión de los modelos de regresión tradicionales. Se discuten métodos para detectar y corregir la autocorrelación en los errores.

Posteriormente, exploramos la predicción en contextos donde los errores están correlacionados. En estas circunstancias, las técnicas de predicción estándar pueden ser insuficientes, por lo que se examinan enfoques avanzados para mejorar la precisión predictiva en presencia de errores correlacionados.

Finalmente, la importancia de las variables rezagadas se examina en detalle. Estas variables, que reflejan los valores de periodos anteriores, son esenciales para capturar la dependencia temporal y la autocorrelación en los modelos de series temporales. Discutimos cómo la inclusión adecuada de variables rezagadas no solo mejora la comprensión de la dinámica temporal, sino que también refina la capacidad de estimación y predicción de los modelos.

Este trabajo subraya la importancia de considerar la correlación de errores y el uso de variables rezagadas en el análisis estadístico, proporcionando una visión integral y práctica de estas complejidades en el modelado de datos.

Índice

1. Introducción	3
2. Marco Teorico	3
2.1. Autocorrelación y estimación de la correlación	3
2.2. Regresión clásica en el contexto de Series de Tiempo	8
2.3. Regresión con errores autocorrelacionados	11
2.4. Modelos de regresión con errores correlacionados	12
2.5. Pruebas y Medidas de solución para los errores autocorrelacionados	14
2.5.1. Prueba de Durbin-Watson	14
2.5.2. Algoritmo Cochrane-Orcutt	15
2.6. Estimación y propiedades de los estimadores	16
2.6.1. Propiedad 1	17
2.6.2. Propiedad 2	17
2.6.3. Propiedad 3	17
2.6.4. Propiedad 4	17
2.7. Predicción	17
2.8. Modelos de regresión con rezagos	18
Bibliografía ²⁴	

1. Introducción

Dentro del mundo de la estadística y el análisis de datos, nos enfrentamos constantemente al desafío de interpretar y modelar datos que no siempre siguen patrones ideales o simplificados. Un área particularmente compleja y fundamental en este campo es la regresión y predicción en presencia de errores correlacionados, así como el uso de variables rezagadas. El presente reporte se centrará en explorar estos aspectos cruciales, abordando tanto los fundamentos teóricos como algunas de las aplicaciones prácticas.

Primero, examinaremos la regresión con errores correlacionados, una situación común en series temporales y datos longitudinales donde las observaciones no son independientes entre sí. Esta correlación entre errores puede conducir a estimaciones ineficientes y sesgadas si no se manejan adecuadamente. Discutiremos cómo identificar la presencia de errores correlacionados y las técnicas para ajustar los modelos de regresión en este contexto.

Posteriormente, abordaremos la predicción en escenarios donde los errores están correlacionados. En tales situaciones, las predicciones estándar basadas en modelos de mínimos cuadrados pueden no ser óptimas. Exploraremos métodos avanzados como la estimación por Mínimos Cuadrados Generalizados (GLS) y sus variantes para mejorar la precisión de las predicciones en estos casos.

Finalmente, abordaremos el concepto de variables rezagadas y su importancia en el análisis de series temporales y modelos econométricos. Las variables rezagadas, que representan valores de variables en periodos anteriores, son esenciales para capturar la dinámica temporal y la autocorrelación en los datos. Discutiremos cómo la inclusión de estas variables puede enriquecer nuestros modelos y mejorar tanto la estimación como la capacidad predictiva.

A lo largo del presente trabajo, combinaremos teoría y algunos ejemplos prácticos para proporcionar una comprensión integral de estos conceptos clave en el análisis de regresión y predicción con errores correlacionados.

2. Marco Teorico

2.1. Autocorrelación y estimación de la correlación

En el contexto del análisis de series temporales, surge un interés particular en describir el comportamiento de los errores cuando estos exhiben cierta estructura de dependencia. Para abordar este aspecto, se utilizan conceptos fundamentales como la autocorrelación y la estimación de la correlación. Estas herramientas proporcionan una comprensión más profunda de las relaciones temporales presentes en los datos y son esenciales para modelar y analizar adecuadamente las series temporales.

La falta de independencia entre dos valores adyacentes x_s y x_t se puede evaluar numéricamente, como en estadística clásica, utilizando las nociones de covarianza y correlación. Suponiendo que la varianza de x_t es finita, tenemos la siguiente definición.

Definición 1. La función de autocovarianza $\gamma_x(s, t)$ mide la dependencia lineal entre dos puntos en la misma serie observados en diferentes momentos. Para todos los s y t , se define como el producto del segundo momento:

$$\gamma_x(s, t) = \text{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]. \quad (2.1)$$

Cuando no hay ambigüedad, utilizamos $\gamma(s, t)$ en lugar de $\gamma_x(s, t)$.

La autocovarianza mide la dependencia lineal entre dos puntos en la misma serie observados en diferentes momentos. Las series muy suaves exhiben funciones de autocovarianza que permanecen grandes incluso cuando t y s están muy separados, mientras que las series irregulares tienden a tener funciones de autocovarianza que son casi cero para separaciones grandes. Cabe resaltar que de la estadística clásica si $\gamma_x(s, t) = 0$, entonces x_s y x_t no están linealmente relacionados, pero aún puede haber alguna estructura de dependencia entre ellos. Sin embargo, si x_s y x_t son normales bivariadas, que $\gamma_x(s, t) = 0$ implica su independencia.

Es claro que, para $s = t$, la autocovarianza se reduce a la varianza porque

$$\gamma_x(t, t) = E[(x_t - \mu_t)^2] = \text{var}(x_t)$$

Definición 2. La función de autocorrelación (ACF) se define como:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}. \quad (2.2)$$

La ACF mide la previsibilidad lineal de la serie en el tiempo t , denotada como x_t , utilizando solo el valor x_s . Podemos demostrar fácilmente que $-1 \leq \rho(s, t) \leq 1$ utilizando la desigualdad de Cauchy-Schwarz. Si podemos predecir x_t perfectamente a partir de x_s a través de una relación lineal, $x_t = \beta_0 + \beta_1 x_s$, entonces la correlación será +1 cuando $\beta_1 > 0$ y -1 cuando $\beta_1 < 0$. Por lo tanto, tenemos una medida aproximada de la capacidad para pronosticar la serie en el tiempo t a partir del valor en el tiempo s .

Las definiciones anteriores de las funciones de autocovarianza y autocorrelación son completamente generales. Ahora a menudo, nos gustaría medir la predecibilidad de otra serie y_t a partir de la serie x_s . Suponiendo que ambas series tienen varianzas finitas, tenemos la siguiente definición.

Definición 3. La función de autocovarianza cruzada entre dos series, x_t e y_t , es

$$\gamma_{xy}(s, t) = \text{cov}(x_s, y_t) = E[(x_s - \mu_{xs})(y_t - \mu_{yt})]. \quad (2.3)$$

También hay una versión escalada de la función de autocovarianza cruzada.

Definición 4. La función de autocorrelación cruzada (CCF) se da por

$$\rho_{xy}(s, t) = \frac{\gamma_{xy}(s, t)}{\sqrt{\gamma_x(s, s)\gamma_y(t, t)}}. \quad (2.4)$$

Podemos extender fácilmente las ideas anteriores al caso de más de dos series, digamos, $x_{t1}, x_{t2}, \dots, x_{tr}$; es decir, series temporales multivariadas con r componentes. Por ejemplo, la extensión de 2.20 en este caso es

$$\gamma_{jk}(s, t) = E[(x_{sj} - \mu_{sj})(x_{tk} - \mu_{tk})] \quad j, k = 1, 2, \dots, r. \quad (2.5)$$

En las definiciones anteriores, las funciones de autocovarianza y autocorrelación cruzada pueden cambiar a medida que nos movemos a lo largo de la serie porque los valores dependen tanto de s como de t , las ubicaciones de los puntos en el tiempo.

A continuación se introduce la noción de regularidad utilizando los conceptos de estacionariedad y de estacionariedad débil, cuando la media es constante, que son fundamentales para permitirnos analizar datos de series temporales de muestras cuando solo está disponible una serie.

Definición 5. Una serie temporal estrictamente estacionaria es aquella para la cual el comportamiento probabilístico de cualquier conjunto de valores

$$\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$$

es idéntico al conjunto desplazado en el tiempo

$$\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$$

Es decir,

$$Pr\{x_{t_1} \leq c_1, \dots, x_{t_k} \leq c_k\} = Pr\{x_{t_1+h} \leq c_1, \dots, x_{t_k+h} \leq c_k\} \quad (2.6)$$

para todos los $k = 1, 2, \dots$, todos los puntos temporales t_1, t_2, \dots, t_k , todos los números c_1, c_2, \dots, c_k , y todos los desplazamientos temporales $h = 0, \pm 1, \pm 2, \dots$

Si una serie temporal es estrictamente estacionaria, entonces todas las funciones de distribución multivariada para subconjuntos de variables deben coincidir con sus contrapartes en el conjunto desplazado, para todos los valores del parámetro de desplazamiento h . Por ejemplo, cuando $k = 1$, 2.9 implica que

$$Pr\{x_s \leq c\} = Pr\{x_t \leq c\} \quad (2.7)$$

para cualquier punto temporal s y t . Esta afirmación implica, por ejemplo, que la probabilidad de que el valor de una serie temporal muestreada por hora sea negativo a la 1 a. m. es la misma que a las 10 a. m. Además, si existe la función de media, μ_t , de la serie, la ecuación 2.12 implica que $\mu_s = \mu_t$ para todos los s y t , y por lo tanto, μ_t **debe ser constante**.

Cuando $k = 2$, podemos expresar la ecuación 2.9 como

$$Pr\{x_s \leq c_1, x_t \leq c_2\} = Pr\{x_{s+h} \leq c_1, x_{t+h} \leq c_2\} \quad (2.8)$$

para cualquier punto temporal s, t , y desplazamiento h . Así, si la función de varianza del proceso existe, las ecuaciones 2.12 y 2.13 implican que la función de autocovarianza de la serie x_t satisface

$$\gamma(s, t) = \gamma(s + h, t + h)$$

para todos los s, t , y h . Podemos interpretar este resultado diciendo que la función de autocovarianza del proceso depende solo de la diferencia de tiempo entre s y t , y no de los tiempos reales.

La versión de estacionariedad en la definición 3 es demasiado fuerte para la mayoría de las aplicaciones. Además, es difícil evaluar la estacionariedad estricta a partir de un solo conjunto de datos, de una realización. En lugar de imponer condiciones a todas las posibles distribuciones de una serie temporal, utilizaremos una versión más suave que imponga condiciones solo sobre los dos primeros momentos de la serie. Para ello se considera la siguiente definición.

Definición 6. Una serie temporal débilmente estacionaria, x_t , es un proceso de varianza finita tal que:

- La función de valor medio, μ_t , es constante y no depende del tiempo t .
- La función de autocovarianza, $\gamma(s, t)$, definida en 2.1, depende de s y t solo a través de su diferencia $|s - t|$.

De ahora en adelante, se usará el término “estacionaria” para referirnos a estacionariedad débil; si un proceso es estacionario en el sentido estricto, utilizaremos el término “estrictamente estacionario”.

Para el tiempo t , escribiremos $\mu_t = \mu$, dado que no depende del tiempo. Además, dado que la función de autocovarianza, $\gamma(s, t)$, de una serie temporal estacionaria x_t , depende de s y t solo a través de su diferencia $|s - t|$, podemos simplificar la notación. Sea $s = t + h$, donde h representa el desplazamiento o **rezago temporal**. Entonces,

$$\gamma(t + h, t) = \text{cov}(x_{t+h}, x_t) = \text{cov}(x_h, x_0) = \gamma(h, 0)$$

porque la diferencia de tiempo entre los momentos $t + h$ y t es la misma que la diferencia de tiempo entre h y 0. Por lo tanto, la función de autocovarianza de una serie temporal estacionaria no depende del argumento de tiempo t . De ahora en adelante, por conveniencia, omitiremos el segundo argumento de $\gamma(h, 0)$. Con esto en mente podemos simplificar la definición de la autocorrelación para cuando se trabaja con una serie de tiempo.

Definición 7. La función de autocorrelación (ACF) de una serie temporal estacionaria se escribirá utilizando 2.2 como

$$\rho(h) = \frac{\gamma(t + h, t)}{\sqrt{\gamma(t + h, t + h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)}. \quad (2.9)$$

La desigualdad de Cauchy-Schwarz muestra nuevamente que $-1 \leq \rho(h) \leq 1$ para todo h , lo que permite evaluar la importancia relativa de un valor de autocorrelación dado al compararlo con los valores extremos -1 y 1 .

Cuando se dispone de varias series, una noción de estacionariedad aún se aplica con condiciones adicionales. Consideremos las siguientes definiciones.

Definición 8. Dos series temporales, digamos x_t e y_t , se dicen conjuntamente estacionarias si cada una de ellas es estacionaria y la función de autocovarianza cruzada

$$\gamma_{xy}(h) = \text{cov}(x_{t+h}, y_t) = E[(x_{t+h} - \mu_x)(y_t - \mu_y)] \quad (2.10)$$

es una función solo del rezago h .

Definición 9. La función de autocorrelación cruzada (CCF) de series temporales conjuntamente estacionarias x_t e y_t se define como

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}} \quad (2.11)$$

Estimación de correlación

Aunque las funciones teóricas de autocovarianza y autocorrelación son útiles para describir las propiedades de ciertos modelos hipotéticos, la mayoría de los análisis deben realizarse utilizando datos muestrales. Esta limitación implica que solo se dispone de los puntos muestreados x_1, x_2, \dots, x_n para

estimar las funciones de media, autocovarianza y autocorrelación. Desde el punto de vista de la estadística clásica, esto plantea un problema porque típicamente no tendremos realizaciones independientes e idénticamente distribuidas de x_t disponibles para estimar las funciones de covarianza y correlación.

En la situación habitual con solo una realización, sin embargo, la suposición de estacionariedad se vuelve crítica. De alguna manera, debemos utilizar promedios sobre esta única realización para estimar las medias poblacionales y las funciones de covarianza.

En consecuencia, si una serie temporal es estacionaria, la función de la media $\mu_t = \mu$ es constante, por lo que podemos estimarla mediante la media muestral,

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \quad (2.12)$$

En este caso, $E(\bar{x}) = \mu$, y el error estándar de la estimación es la raíz cuadrada de $\text{var}(\bar{x})$, que se puede calcular utilizando las propiedades de la covarianza, y se expresa como:

$$\text{var}(\bar{x}) = \frac{1}{n^2} \sum_{t=1}^n \text{cov}(x_t, x_s) = \frac{1}{n} \sum_{h=-n}^{n-1} (1 - |h|/n) \gamma_x(h). \quad (2.13)$$

Si el proceso es ruido blanco, 2.13 se reduce al familiar σ_x^2/n , recordando que $\gamma_x(0) = \sigma_x^2$. Observa que, en el caso de dependencia, el error estándar de \bar{x} puede ser menor o mayor que en el caso de ruido blanco, dependiendo de la naturaleza de la estructura de correlación.

La función de autocovarianza teórica, 2.1, se estima mediante la función de autocovarianza muestral definida de la siguiente manera.

Definición 10. La función de autocovarianza muestral se define como

$$\hat{\gamma}(h) = \frac{1}{n-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}), \quad (2.14)$$

con $\hat{\gamma}(-h) = \hat{\gamma}(h)$ para $h = 0, 1, \dots, n-1$. Esta suma se realiza en un rango restringido porque x_{t+h} no está disponible para $t+h > n$. Se prefiere este estimador que al que se obtendría dividiendo por $n-h$ porque cuando dividimos por $n-1$ es una función no negativa definida. Cabe recalcar que la función de autocovarianza de un proceso estacionario se espera que sea no negativa definida, lo que asegura que las varianzas de las combinaciones lineales de las variables x_t nunca serán negativas. Y debido a que una varianza nunca es negativa, la estimación de esa varianza, dada por la siguiente ecuación,

$$\hat{\text{var}}(a_1 x_1 + \dots + a_n x_n) = \sum_{j=1}^n \sum_{k=1}^n a_j a_k \hat{\gamma}(j-k), \quad (2.15)$$

también debería ser no negativa. El estimador garantiza este resultado, pero no existe tal garantía si dividimos por $n-h$. Se puede probar que ni dividir por n ni por $n-h$ produce un estimador insesgado de $\gamma(h)$.

Definición 11. La función de autocorrelación muestral se define, de manera análoga a 2.9, como

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \quad (2.16)$$

La función de autocorrelación muestral tiene una distribución de muestreo que nos permite evaluar si los datos provienen de una serie completamente aleatoria o blanca, o si las correlaciones son estadísticamente significativas en ciertos rezagos.

A continuación definimos las estimaciones muestrales de la autocovarianza y la autocorrelación.

Definición 12. Los estimadores para la función de autocovarianza cruzada, $\gamma_{xy}(h)$, como se da en 2.10, y la autocorrelación cruzada, $\rho_{xy}(h)$, dada en 2.11, son respectivamente, para la función de autocovarianza cruzada muestral

$$\hat{\gamma}_{xy}(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}) \quad (2.17)$$

donde $\hat{\gamma}_{xy}(-h) = \hat{\gamma}_{yx}(h)$ determina la función para rezagos negativos, y la función de autocorrelación cruzada muestral

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}} \quad (2.18)$$

Para el estudio de los modelos con variables rezagas y otras aplicaciones en series de tiempo es importante tener presente la siguiente propiedad.

Propiedad. Distribución en Muestras Grandes de la Cross-Correlation (correlación cruzada)

La distribución en muestras grandes de $\hat{\rho}_{xy}(h)$ es normal con media cero y

$$\sigma_{\hat{\rho}_{xy}} = \frac{1}{\sqrt{n}} \quad (2.19)$$

si al menos uno de los procesos es ruido blanco independiente.

2.2. Regresión clásica en el contexto de Series de Tiempo

Comenzamos nuestra discusión sobre regresión lineal en el contexto de series temporales al asumir que alguna serie temporal dependiente, digamos, x_t , para $t = 1, \dots, n$, está siendo influenciada por una colección de posibles entradas o series independientes, digamos, $z_{t1}, z_{t2}, \dots, z_{tq}$, donde primero consideramos que las entradas son fijas y conocidas. Esta suposición, necesaria para aplicar la regresión lineal convencional, se relajará más adelante. Expresamos esta relación a través del modelo de regresión lineal:

$$x_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + \dots + \beta_q z_{tq} + w_t, \quad (2.20)$$

donde $\beta_0, \beta_1, \beta_2, \dots, \beta_q$ son coeficientes de regresión fijos y desconocidos, y $\{w_t\}$ es un proceso de error aleatorio que consiste en variables normales independientes e idénticamente distribuidas (iid) con media cero y varianza σ_w^2 .

Para la regresión de series temporales, rara vez se da el caso de que el ruido sea blanco, y eventualmente tendremos que relajar esa suposición. Por ejemplo, consideremos los datos `chicken` disponibles en R sobre el precio de contado por mes de ave entera de corral (pollo), muelles de Georgia, centavos estadounidenses por libra, de agosto 2001 a julio de 2016.

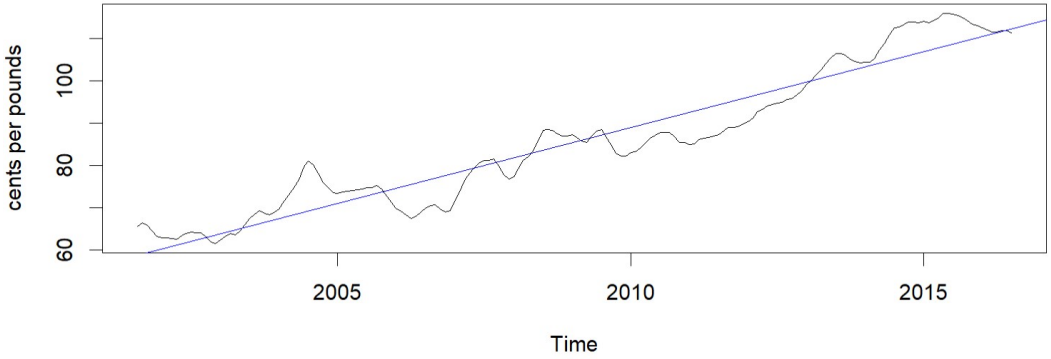


Figura 1. Precio del pollo, precio de contado por mes. Georgia US, 2001-2016 [1].

En la gráfica se observa la recta que ajusta el modelo de regresión como una estimación lineal del comportamiento del precio de contado de pollo por mes. Se ajustó el modelo dado en ecuación 2.20 con $q = 1$:

$$x_t = \beta_0 + \beta_1 z_t + w_t, \quad z_t = 2001\frac{7}{12}, 2001\frac{8}{12}, \dots, 2016\frac{6}{12}$$

Es importante destacar que estamos haciendo la suposición de que los errores, w_t , forman una secuencia normal independiente e idénticamente distribuida (iid), aunque esto puede no ser cierto; el problema de errores autocorrelacionados lo abordaremos más adelante. De hecho, este supuesto se viola con frecuencia en datos de series temporales.

En la técnica de mínimos cuadrados ordinarios (OLS), minimizamos la suma de cuadrados de errores:

$$Q = \sum_{t=1}^n w_t^2 = \sum_{t=1}^n (x_t - [\beta_0 + \beta_1 z_t])^2, \quad (2.21)$$

con respecto a β_i para $i = 0, 1$. En este caso, podemos usar cálculos simples para evaluar $\partial Q / \partial \beta_i = 0$ para $i = 0, 1$, y obtener dos ecuaciones para resolver los valores de β . Las estimaciones OLS de los coeficientes son explícitas y se están dadas por:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n (x_t - \bar{x})(z_t - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2},$$

$$\hat{\beta}_0 = \bar{x} - \hat{\beta}_1 \bar{z},$$

donde $\bar{x} = \frac{\sum_{t=1}^n x_t}{n}$ y $\bar{z} = \frac{\sum_{t=1}^n z_t}{n}$ son las medias muestrales respectivas.

Realizando la regresión por mínimos cuadrados obtenemos el coeficiente de la pendiente estimada dada por $\hat{\beta}_0 = 3,592$, con un error estándar de 3.59, lo que significa que se está produciendo un aumento estimado significativo de aproximadamente 3.6 centavos por mes.

El modelo de regresión lineal múltiple descrito en la ecuación 2.20 puede expresarse de manera más general mediante la definición de los vectores de columna $z_t = (1, z_{t1}, z_{t2}, \dots, z_{tq})'$ y $\beta = (\beta_0, \beta_1, \dots, \beta_q)'$, donde $'$ denota la transposición, por lo que 2.20 se puede expresar en la forma alternativa:

$$x_t = \beta_0 + \beta_1 z_{t1} + \dots + \beta_q z_{tq} + w_t = \beta^T z_t + w_t \quad (2.22)$$

donde $w_t \sim iid \mathcal{N}(0, \sigma_w^2)$. Al igual que en el ejemplo anterior, la estimación de OLS encuentra el vector de coeficientes β que minimiza la suma de cuadrados de errores:

$$Q = \sum_{t=1}^n w_t^2 = \sum_{t=1}^n (x_t - \beta^T z_t)^2, \quad (2.23)$$

con respecto a $\beta_0, \beta_1, \dots, \beta_q$. Esta minimización se puede lograr diferenciando 2.23 con respecto al vector β . De cualquier manera, la solución debe satisfacer $\sum_{t=1}^n (x_t - \hat{\beta}^T z_t) z_t = 0$. Este procedimiento proporciona las ecuaciones normales: donde se indica el resultado de aplicar la transposición a $\hat{\beta}$.

$$\left(\sum_{t=1}^n z_t z_t^T \right) \hat{\beta} = \sum_{t=1}^n z_t x_t. \quad (2.24)$$

Si $\sum_{t=1}^n z_t z_t^T$ no es singular, la estimación de mínimos cuadrados de β es

$$\hat{\beta} = \left(\sum_{t=1}^n z_t z_t^T \right)^{-1} \sum_{t=1}^n z_t x_t. \quad (2.25)$$

La suma de cuadrados del error (SSE) minimizada de 2.23, denotada como SSE, se puede expresar como

$$SSE = \sum_{t=1}^n (x_t - \hat{\beta}^T z_t)^2. \quad (2.26)$$

Los estimadores de mínimos cuadrados ordinarios son insesgados, es decir, $E(\hat{\beta}) = \beta$, y tienen la menor varianza dentro de la clase de estimadores lineales insesgados.

Si los errores w_t están distribuidos normalmente, $\hat{\beta}$ es también el estimador de máxima verosimilitud para β y sigue una distribución normal con covarianza:

$$\text{cov}(\hat{\beta}) = \sigma_w^2 C, \quad (2.27)$$

donde

$$C = \left(\sum_{t=1}^n z_t z_t^T \right)^{-1}, \quad (2.28)$$

es una notación conveniente. Un estimador insesgado para la varianza σ_w^2 es:

$$s_w^2 = \text{MSE} = \frac{\text{SSE}}{n - (q + 1)}, \quad (2.29)$$

donde MSE denota el error cuadrático medio. Bajo la suposición normal,

$$t = \frac{(\hat{\beta}_i - \beta_i)}{s_w \sqrt{c_{ii}}}, \quad (2.30)$$

donde β_i es el i -ésimo componente de β y c_{ii} es el i -ésimo elemento diagonal de la matriz C .

2.3. Regresión con errores autocorrelacionados

Previamente, consideramos el modelo de regresión clásico con errores w_t **no correlacionados**, la independencia de los errores es uno de los supuestos al construir un modelo de regresión lineal. Sin embargo, en la práctica pueden aparecer dependencias en los errores debido a algún componente temporal. Los errores que se encuentran correlacionados en el tiempo son **errores autocorrelacionados** y la existencia de éstos indica que el modelo no es adecuado. Cuando los errores se encuentran autocorrelacionados pueden surgir algunas complicaciones al usar mínimos cuadrados ordinarios como:

- Los coeficientes estimados de regresión son insesgados pero ya no cumplen con la propiedad de mínima varianza
- El MSE puede subestimar seriamente la varianza real de los errores
- El error estándar de los coeficientes de regresión puede subestimar seriamente el valor real de la desviación estándar de los coeficientes de regresión estimados
- Intervalos de confianza y procedimientos de inferencia no son aplicables estrictamente

Si ignoramos los términos de los errores, tendremos una penalización en términos de intervalos de predicción más amplios. Por lo que para solucionar esta problemática modelaremos la autocorrelación para obtener intervalos de predicción más precisos.

En lo que sigue, mencionaremos las modificaciones que deben considerarse cuando los **errores están autocorrelacionados** o existe alguna dependencia entre ellos. Consideremos el siguiente modelo de regresión [1]:

$$y_t = \beta' x_t + \epsilon_t \quad (2.31)$$

$t = 1, \dots, n$, donde ϵ_t es un proceso con función de covarianza $\gamma(s, t)$, que en notación matricial (2.31) toma la forma:

$$y = X\beta + \epsilon \quad (2.32)$$

donde $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ es un vector de tamaño $(n \times 1)$ con matriz de covarianza de tamaño $n \times n$, $\Gamma = \{\gamma(s, t)\}$. Note que $X = [x_1, x_2, \dots, x_n]'$ de tamaño $n \times q$, es la matriz de variables de entrada. Si conociéramos a la matriz de covarianza Γ , sería posible encontrar una matriz de transformación, A , tal que $A\Gamma A' = \sigma^2 I$, donde I denota a la matriz identidad de tamaño $n \times n$. Por lo tanto, podemos transformar el modelo (2.32) a:

$$Ay = AX\beta + A\epsilon \quad (2.33)$$

$$= U\beta + w \quad (2.34)$$

donde $U = AX$ y w es un vector de ruido blanco con matriz de covarianza $\sigma^2 I$ como en (2.20). Luego, aplicando mínimos cuadrados o máxima verosimilitud al vector Ay obtenemos:

$$\hat{\beta}_w = (U'U)^{-1}U'Ay \quad (2.35)$$

$$= (X'A'AX)^{-1}X'A'Ay \quad (2.36)$$

$$= (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}y \quad (2.37)$$

dado que $\sigma^2\Gamma^{-1} = A'A$. Sin embargo, a menos que conozcamos la forma de Γ no podremos aplicar (2.37).

2.4. Modelos de regresión con errores correlacionados

Una estructura de error autocorrelacionada que podemos encontrar frecuentemente es un **proceso autorregresivo de primer orden AR(1)**. Este modelo considera la situación en la que el error en un tiempo específico está linealmente relacionado al tiempo anterior. Es decir que los errores siguen un modelo simple de regresión lineal que puede escribirse como:

$$\epsilon_t = \rho\epsilon_{t-1} + w_t, \quad |\rho| < 1 \quad (2.38)$$

donde $|\rho| < 1$ es el parámetro de autocorrelación, y w_t es el nuevo término de error, a veces llamado **random shock**, que sigue los supuestos de errores en regresión, esto es: $w_t \sim N(0, \sigma^2)$, *varianza constante, independientes e idénticamente distribuidos* [2].

- Si $\rho > 0$ los errores están autocorrelacionados positivamente. Es posible ver que si $\rho > 0$, un término de error positivo ϵ_{t-1} tiende a producir otro término de error positivo ϵ_t . Análogamente, un término de error negativo ϵ_{t-1} tiende a producir otro término de error negativo ϵ_t . En otras palabras, la autocorrelación positiva existe cuando términos de error positivos tienden a ser seguidos en el tiempo por términos de error positivos y términos de error negativo tienden a ser seguidos en el tiempo por términos de error negativos. La autocorrelación positiva en los términos del error significa que más del promedio de valores de y_t tienden a ser seguidos por valores más grandes que el promedio de y_t y valores más pequeños que el promedio de y_t , tienden a ser seguidos por valores más pequeños que el promedio de y_t .
- Si $\rho < 0$ los errores están autocorrelacionados negativamente y por lo tanto, un término de error positivo ϵ_{t-1} tiende a producir un término de error negativo ϵ_t . Sin embargo, un término de error negativo ϵ_{t-1} tiende a producir un término de error positivo ϵ_t .

Note que si ρ es cercano a 1, entonces los errores están muy correlacionados, en el caso donde $\rho = 0$ entonces regresamos al caso con independencia. El supuesto de independencia nos dice que los términos del error ordenados en el tiempo no muestran autocorrelación positiva ni autocorrelación negativa. Esto nos dice que los términos del error ocurren en un patrón aleatorio en el tiempo.

Este modelo (2.38) nos dice que el error al tiempo t es predecible a partir de una fracción del error al tiempo $t - 1$ más alguna nueva perturbación w_t .

Una de las razones por las que los errores pueden tener una estructura autorregresiva, es que tanto y como X al tiempo t están relacionadas con las mediciones de y y X al tiempo $t - 1$. Éstas relaciones están siendo absorbidas por el término del error en el modelo de regresión lineal múltiple que solo relaciona las mediciones de y y X realizadas en tiempos simultáneos.

Los términos del error, ϵ_t mantienen media 0 y varianza constante:

$$E(\epsilon_t) = 0 \quad (2.39)$$

$$Var(\epsilon_t) = \frac{\sigma^2}{1 - \rho^2} \quad (2.40)$$

Sin embargo, la covarianza entre errores adyacentes toma la forma:

$$Cov(\epsilon_t, \epsilon_{t-1}) = \rho \left(\frac{\sigma^2}{1 - \rho^2} \right) \quad (2.41)$$

de manera que el coeficiente de correlación entre los errores adyacentes será de la forma:

$$Corr(\epsilon_t, \epsilon_{t-1}) = \frac{Cov(\epsilon_t, \epsilon_{t-1})}{\sqrt{Var(\epsilon_t)Var(\epsilon_{t-1})}} = \rho \quad (2.42)$$

que es el parámetro de autocorrelación que introducimos en (2.38).

Podemos emplear el correlograma de la función de autocorrelación parcial (PACF) que nos ayuda a evaluar lags o retardos apropiados para los errores en el modelo de regresión con errores autorregresivos. Particularmente, primero ajustamos un modelo de regresión lineal a nuestros datos de la serie de tiempo y guardamos los residuales. Posteriormente podemos fijarnos en el correlograma de la PACF de los residuales contra los lags. Las autocorrelaciones parciales de muestras grandes que son significativamente diferentes de 0 indican términos rezagados de ϵ que pueden ser predictores útiles de ϵ_t .

Además del proceso autorregresivo de primer orden, existen otro tipo de estructuras autocorrelacionadas del error. Por ejemplo, el **proceso autorregresivo de orden p** que se define como [2]:

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \dots + \rho_p \epsilon_{t-p} + w_t \quad (2.43)$$

este proceso (2.43) relaciona ϵ_t el término del error en un periodo de tiempo t , a términos de errores previos $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-p}$. En este caso, $\rho_1, \rho_2, \dots, \rho_p$ son parámetros desconocidos y w_t es un término de error (random shock) con media cero, que satisface varianza constante, independencia y suposiciones de normalidad [2].

De manera general, es posible asumir una estructura estacionaria para la covarianza del proceso x_t del error que corresponde a un proceso lineal e intentar encontrar una representación ARMA para x_t .

Por ejemplo, para el caso del error que sigue un modelo autorregresivo de orden p (2.10), $AR(p)$, entonces:

$$\phi(B)e_t = w_t \quad (2.44)$$

y $\phi(B)$ es la transformación lineal que, cuando se aplica el proceso del error, produce ruido blanco w_t . De manera que si consideramos a esta transformación como la transformación A , podemos obtener:

$$\phi(B)y_t = \beta' \phi(B)x_t + w_t \quad (2.45)$$

que es el mismo modelo (2.34). Ahora, definiendo $u_t = \phi(B)y_t$, y $v_t = \phi(B)x_t$ obtenemos el problema de regresión simple:

$$u_t = \beta' v_t + w_t \quad (2.46)$$

considerado anteriormente en (2.34).

Si el error es un $ARMA(p, q)$, entonces:

$$e_t = \sum_{j=1}^p \phi_j e_{t-j} - \sum_{k=1}^q \theta_k w_{t-k} + w_t, \quad w_t \sim WN(0, \sigma^2) \quad (2.47)$$

con el operador de retardo, B , se puede escribir (2.47) como:

$$\phi(B)e_t = \theta(B)w_t \quad \text{o} \quad e_t = \phi^{-1}(B)\theta(B)w_t = \frac{\theta(B)}{\phi(B)}w_t \quad (2.48)$$

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,

$$y_t = \sum_{i=1}^m \beta_i x_{it} + \epsilon_t \quad (2.49)$$

$$= \sum_{i=1}^m \beta_i x_{it} + \frac{\theta(B)}{\phi(B)} w_t \quad (2.50)$$

2.5. Pruebas y Medidas de solución para los errores autocorrelacionados

De manera usual, asumimos que los términos de los errores son independientes a menos que exista una razón específica para creer que no es el caso. Usualmente, la violación a esta suposición ocurre debido a que existe un componente temporal conocido en cómo se obtuvieron las observaciones. Dado que los residuales son puntos estimados de los términos del error, una gráfica de los residuales contra el tiempo puede usarse de manera informal para detectar violaciones del supuesto de independencia. Si existe independencia, entonces los residuales deben encontrarse aleatoriamente dispersos alrededor del cero. Sin embargo, si se observa que surge un patrón, particularmente uno cíclico, entonces es probable que la dependencia sea un problema y los términos del error esten autocorrelacionados [2].

2.5.1. Prueba de Durbin-Watson

Recordamos que en el caso de tener un $AR(1)$ con los errores, éstos pueden modelarse como en (2.38). Si sospechamos una autocorrelación de primer orden con los errores, entonces existe una prueba formal con respecto al parámetro ρ . En particular, la prueba de **Durbin Watson**, la cual se construye como sigue:

$$H_0 : \rho = 0 \quad (2.51)$$

$$H_A : \rho \neq 0 \quad (2.52)$$

La hipótesis nula $\rho = 0$ significa que $\epsilon_t = w_t$, o que el término del error en un periodo de tiempo no está autocorrelacionada con el término de error en el periodo previo, mientras que la hipótesis alternativa de $\rho \neq 0$ significa que el término del error en un periodo está autocorrelacionado positiva o negativamente con el término del error en el periodo previo.

El estadístico de prueba de Durbin-Watson para un conjunto de datos de tamaño n , está dado por [2]:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (2.53)$$

donde $e_t = y_t - \hat{y}_t$ son los residuales del ajuste de mínimos cuadrados ordinarios. La prueba del estadístico de Durbin-Watson varía de 0 a 4, con valores entre 0 y 2 indicando autocorrelación positiva, 2 indicando autocorrelación cero y valores entre 2 y 4 indicando autocorrelación negativa. Resulta complicado obtener los valores críticos exactos, por lo que se puede hacer uso de las *tablas de significancia de Durbin-Watson*.

Durbin y Watson han mostrado que si existen puntos denotados como $d_{L,\alpha/2}$ y $d_{U,\alpha/2}$ tal que α es la probabilidad de un *error Tipo I*, entonces [2]:

1. Si $d < d_{L,\alpha/2}$ o si $(4 - d) < d_{L,\alpha/2}$ rechazamos la hipótesis nula H_0 .
2. Si $d > d_{U,\alpha/2}$ y si $(4 - d) > d_{U,\alpha/2}$ no rechazamos H_0 .
3. Si $d_{L,\alpha/2} \leq d \leq d_{U,\alpha/2}$ o $d_{L,\alpha/2} \leq (4 - d) \leq d_{U,\alpha/2}$, la prueba es inconclusa.

Algunas observaciones del test de Durbin-Watson [2]:

- La validez de la prueba de Durbin-Watson depende del supuesto de que la población de todos los posibles residuales en cualquier tiempo t tengan una distribución normal.
- La prueba de Durbin-Watson solo es válida para estructuras de errores del tipo $AR(1)$ (autocorrelacionados de orden 1, positiva o negativamente).
- Es más frecuente encontrar autocorrelaciones de errores positivas.
- Los datos de series temporales pueden exhibir estructuras de errores autocorrelacionadas más complicadas, en tales casos la autocorrelación puede detectarse usando la **función de autocorrelación muestral ACF**.

2.5.2. Algoritmo Cochrane-Orcutt

El algoritmo propuesto por Cochrane y Orcutt en 1949 permite ajustar un modelo de regresión con errores autocorrelacionados [1]:

1. Inicialmente, se aplica una regresión ordinaria de y_t en x_t (actuando como si los errores no estuvieran autocorrelacionados). Se guardan los valores de los residuales.
2. Se ajusta un modelo ARMA a los residuales $\hat{\epsilon}_t = y_t - \hat{\beta}'x_t$, digamos

$$\hat{\phi}(B)\hat{\epsilon}_t = \hat{\theta}(B)w_t \quad (2.54)$$

3. Posteriormente, aplicamos la transformación ARMA a ambos lados de (2.5), esto es:

$$u_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)}y_t \quad (2.55)$$

y

$$v_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)}x_t \quad (2.56)$$

para obtener el modelo de regresión transformado (2.45)

4. Aplicar un modelo de regresión de mínimos cuadrados asumiendo errores no correlacionados en el modelo de regresión transformado (2.45), obteniendo:

$$\hat{\beta}_w = (V'V)^{-1}V'u \quad (2.57)$$

donde $V = [v_1, \dots, v_n]'$ y $u = (u_1, \dots, u_n)'$ son los componentes transformados correspondientes

Este procedimiento se repite hasta la convergencia, esto es hasta que se observe una diferencia muy pequeña en las estimaciones entre iteraciones, y hasta que se alcance la solución de máxima verosimilitud bajo normalidad de los errores [1].

2.6. Estimación y propiedades de los estimadores

Las propiedades deseables de los estimadores en el contexto de la regresión lineal son:

- **Insesgados:** Un estimador es insesgado si su esperanza matemática es igual al verdadero valor del parámetro.

$$E(\hat{\beta}) = \beta$$

- **Consistencia:** Un estimador es consistente si, a medida que el tamaño de la muestra se incrementa, converge en probabilidad al verdadero valor del parámetro.

$$P(\lim_{n \rightarrow \infty} \hat{\beta} = \beta) = 1$$

- **Eficientes:** Un estimador es eficiente si tiene la menor varianza posible entre todos los estimadores insesgados.

$$Var(\hat{\beta}) < Var(\tilde{\beta}) \quad \forall \tilde{\beta} \neq \hat{\beta}$$

- **Mínima Varianza:** Entre todos los estimadores insesgados y consistentes, el estimador eficiente es el que tiene la mínima varianza.

$$Var(\hat{\beta}_{eficiente}) = \min_{\hat{\beta}} Var(\hat{\beta})$$

Dado el modelo, queremos estimar los parámetros. Existen varios métodos de estimación: OLS (Mínimos Cuadrados Ordinarios), GLS (Mínimos Cuadrados Generalizados), EGLS (Mínimos Cuadrados Generalizados Factibles), y Least Square Residuals.

Las fórmulas para los estimadores son las siguientes:

- **Estimador OLS:** Se usa para obtener una estimación de los coeficientes del modelo cuando los errores se consideran homocedásticos e independientes

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$$

- **Estimador GLS:** se utiliza cuando se conoce la estructura de la matriz de varianzas y covarianzas de los errores y esta no es una matriz escalar (lo que implicaría homocedasticidad y ausencia de autocorrelación)

$$\hat{\beta}_G = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

donde V es la matriz de varianzas y covarianzas de los errores.

- **Estimador EGLS:** es similar al GLS pero se utiliza cuando la matriz de varianzas y covarianzas no se conoce y se debe estimar a partir de los datos. Por eso se le llama “factible”, ya que en la práctica no siempre conocemos la matriz V . La notación es similar a la del GLS, con la diferencia de que la matriz V y su inversa son estimaciones basadas en los residuos de una estimación preliminar (usualmente OLS)

$$\hat{\beta}_E = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y$$

donde \hat{V} es la estimación de la matriz de varianzas y covarianzas de los errores.

2.6.1. Propiedad 1

Los estimadores OLS son conocidos por ser consistentes y asintóticamente normales. Sin embargo, pueden presentar ciertas particularidades como:

- Puede que no sean eficientes o que tengan tasas de convergencia distintas.
- Puede no ser insesgado.
- Las coordenadas de $\hat{\beta}$ pueden converger a diferentes tasas.
- Se subestima la verdadera varianza.

2.6.2. Propiedad 2

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

2.6.3. Propiedad 3

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

2.6.4. Propiedad 4

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

2.7. Predicción

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

Sabemos que:

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

$$\hat{\epsilon} = (\epsilon_1, \dots, \epsilon_n)'$$

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

$$\hat{\beta} = (\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'V^{-1}\mathbf{Y},$$

(Estimador de mínimos cuadrados generalizados), encontraremos el mejor predictor lineal insesgado.

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

$$\begin{aligned} E(\hat{\mathbf{c}}'_i\mathbf{Y}) &= E(\hat{\mathbf{c}}'_i(\mathbf{X}\hat{\beta} + \hat{\epsilon}_t)) = \hat{\mathbf{c}}'_i\mathbf{X}\hat{\beta} = E[Y_{n+i}] = \mathbf{X}'_{n+i}\hat{\beta} \quad \forall \hat{\beta} \\ &\Rightarrow \hat{\mathbf{c}}'_i\mathbf{X}\hat{\beta} = \mathbf{X}'_{n+i}\hat{\beta} \\ &\Rightarrow \hat{\mathbf{c}}'_i\mathbf{X} = \mathbf{X}'_{n+i}. \end{aligned}$$

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

$$\text{mín}(E[(Y_{n+1} - \hat{\mathbf{c}}'_i\mathbf{Y})^2]),$$

sujeto a $\hat{\mathbf{c}}'_i\mathbf{X} = \mathbf{X}'_{n+i}$.

Entonces para Y_{n+i} ,

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

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

$$\begin{aligned}
 Y_{n+i} - \hat{Y}_{n+i} &= \mathbf{X}_{n+i}' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \hat{\mathbf{b}}_i' [\mathbf{Y} - \mathbf{X}' \hat{\boldsymbol{\beta}}] + \epsilon_{n+i} \\
 &= \mathbf{X}_{n+i}' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \hat{\mathbf{b}}_i' [\mathbf{Y} - \mathbf{X}' \hat{\boldsymbol{\beta}}] + \epsilon_{n+i} \\
 &= [\hat{\mathbf{b}}_i' \mathbf{X} - \mathbf{X}_{n+i}'] (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\epsilon_{n+i} - \hat{\epsilon}_{n+i}).
 \end{aligned}$$

2.8. Modelos de regresión con rezagos

Ahora se considera el problema básico para la descripción y modelado de la relación entre dos series temporales. Considere la relación entre dos series temporales y_t y x_t , la serie y_t puede estar relacionada con rezagos pasados de la serie x_t . La variable x_t se añade a la discusión dado que esto permite abonar información valiosa para la modelación e inferencia de errores en la serie de tiempo y_t .

Cuando se incorpora una variable explicativa rezagada en un modelo, se sugiere que dicha variable podría tener una relación estadística con la respuesta, pero con un “rezago” en esa relación. Esto puede ocurrir cuando la variable explicativa tiene un efecto causal en la respuesta, pero este efecto se manifiesta gradualmente y se refleja en cambios en la respuesta en momentos posteriores.

Por otro lado, cuando se utiliza una variable de respuesta rezagada en un modelo, se interpreta como una especie de representación para la autocorrelación en la variable de respuesta. Las variables explicativas adicionales se incluyen para determinar si existe alguna relación estadística remanente entre estas variables y la respuesta después de eliminar el efecto de la autocorrelación.

Ambas situaciones se consideran comunes en entornos econométricos, donde las variables suelen estar autocorrelacionadas y presentan efectos causales que se manifiestan gradualmente con el tiempo. En términos de cuándo incluir estos términos en los modelos, se destaca que esta decisión es compleja y está relacionada con consideraciones teóricas fundamentales y análisis diagnósticos de los datos. En la literatura, se menciona que, más allá de las consideraciones teóricas, se pueden buscar autocorrelaciones en los residuos de la regresión y correlaciones rezagadas entre variables explicativas y residuos para evaluar si un modelo existente podría beneficiarse de la adición de un término de rezago en el modelo.

En este punto es importante retomar la función de autocorrelación cruzada muestral (CCF) ya que es útil para identificar rezagos de la variable x que podrían ser predictores útiles de y . La CCF muestral (2.18) se define como el conjunto de correlaciones muestrales entre y_t y x_t para $h = 0, \pm 1, \pm 2, \pm 3$, y así sucesivamente. Un valor negativo para h representa una correlación entre la variable x en un tiempo anterior a t y la variable y en el tiempo t . Por ejemplo, considera $h = -2$. El valor de la CCF indicaría que si existe correlación entre x y y en ese caso. En R, la función que proporciona la función de autocorrelación cruzada muestral esta dada por la función `ccf`. Para especificar cuántos rezagos mostrar, se agrega ese número como un argumento de la función. Por ejemplo, `ccf(x, y, 50)` mostrará la CCF

para valores de $h = 0, \pm 1, \dots, \pm 50$.

Cuando uno o más rezagos negativos de x son predictores de y , a veces se dice que x precede a y . Cuando uno o más rezagos positivos de x son predictores de y , a veces se dice que x sigue a y .

En algunos problemas, el objetivo puede ser identificar qué variable está liderando y cuál está rezagada. En muchos problemas que consideramos, sin embargo, examinaremos la(s) variable(s) x como una variable líder de la variable y , porque queremos usar los valores de la variable x para predecir los valores futuros de y . Por lo tanto, generalmente estaremos observando lo que sucede en los valores negativos de h en el gráfico de la CCF.

La regresión clásica se puede extender para el análisis de modelos de regresión rezagados de la forma

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t \quad (2.58)$$

donde v_t es un proceso de ruido estacionario, x_t es la serie de entrada observada, e y_t es la serie de salida observada. Nos interesa estimar los coeficientes del filtro β_r que relacionan los valores rezagados adyacentes de x_t con la serie de salida y_t .

El modelo dado por 2.58 es útil en varios escenarios diferentes, correspondientes a diferentes suposiciones que se pueden hacer sobre los componentes. En el siguiente ejemplo se analiza el uso de la función de autocorrelación cruzada (CCF) para identificar algunas estructuras de regresión relativamente simples para modelar y , además de permitir comprender mejor cómo se escogen las variables de rezago y cómo realizamos la modelación correspondiente.

Ejemplo: *Southern Oscillation Index*(SOI) ¹

La imagen 2 muestra los valores mensuales de una serie ambiental llamada Índice de Oscilación del Sur (SOI, por sus siglas en inglés) y el Reclutamiento asociado (número de peces nuevos), los datos fueron recolectados por el Dr. Roy Mendelssohn del Grupo de Pesquerías Ambientales del Pacífico (comunicación personal). Ambas series abarcan un período de 453 meses, desde 1950 hasta 1987. El SOI mide cambios en la presión atmosférica relacionados con las temperaturas de la superficie del mar en el Pacífico central. El Pacífico central se calienta cada tres a siete años debido al efecto El Niño, que se ha relacionado con diversos eventos climáticos extremos a nivel mundial.

Ambas series en la figura 2 muestran un comportamiento repetitivo con ciclos regularmente recurrentes fácilmente visibles. Este comportamiento periódico es de interés porque los procesos subyacentes pueden ser regulares y la tasa o frecuencia de oscilación que caracteriza el comportamiento de la serie subyacente ayudaría a identificarlos. Las series muestran dos tipos básicos de oscilaciones: un ciclo anual evidente (caliente en verano, frío en invierno) y una frecuencia más lenta que parece repetirse aproximadamente cada 4 años.

Al final de la misma figura 2 se observa el CCF muestral entre ambas series. Las correlaciones cruzadas más dominantes ocurren en algún lugar entre -10 y aproximadamente -4 . Aunque es difícil leer exactamente los desfases desde el gráfico en el código que se adjunta a este reporte se encuentra el listado de todos los valores. Se puede observar claramente que hay valores máximos casi iguales en

¹Ejemplo tomado de [1].

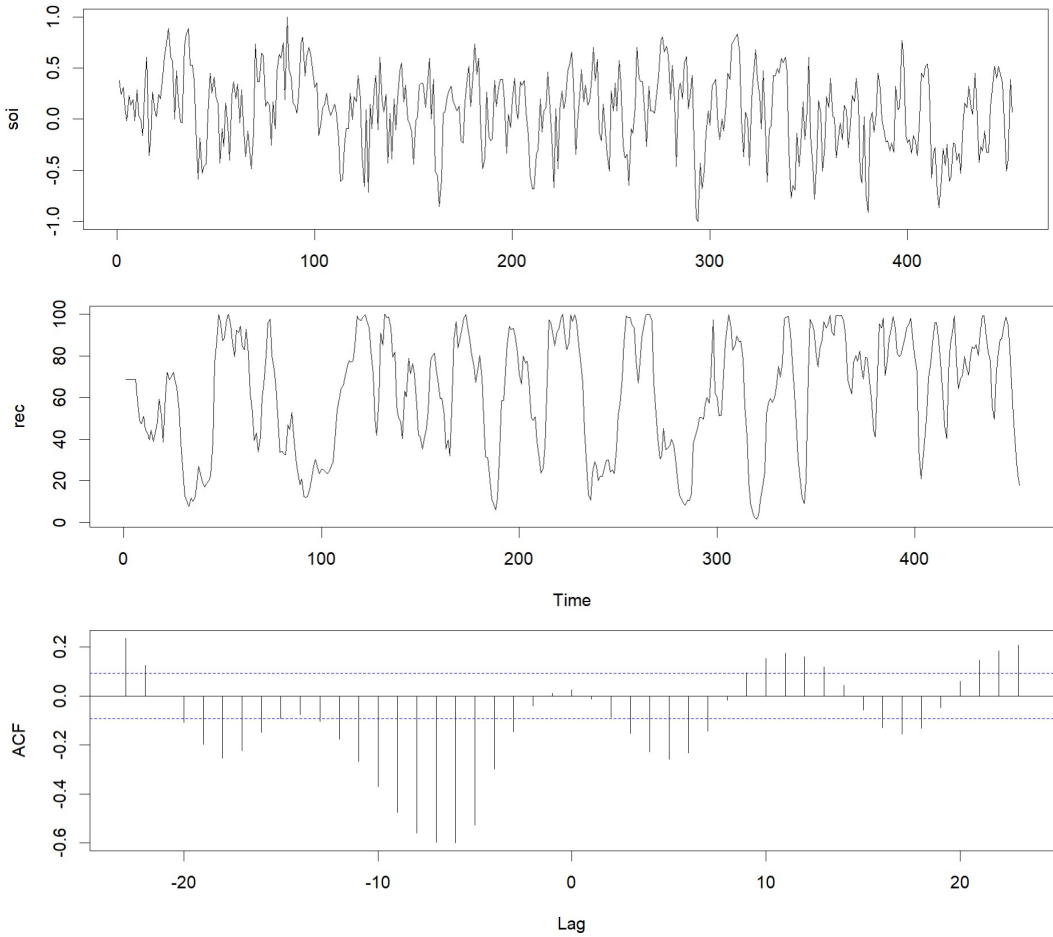


Figura 2. Índice de Oscilación del Sur y Población de peces en el hemisferio Sur, 1950–1987.

$h = -5, -6, -7$ y -8 ($-0.527, -0.599, -0.598$ y -0.560 , respectivamente) con un valores menos predominantes en ambas direcciones desde esos pico. El CCF muestral permite visualizar cierta desviación del componente cíclico de cada serie y hay un pico evidente en $h = -6$. Este resultado implica que el SOI medido en el tiempo $t - 6$ meses está asociado con la serie de reclutamiento en el tiempo t . Podríamos decir que el SOI precede a la serie de reclutamiento en seis meses. El signo del CCF es negativo, lo que lleva a la conclusión de que las dos series se mueven en direcciones opuestas; es decir, los aumentos en el SOI preceden a las disminuciones en el reclutamiento y viceversa.

Considera la siguiente regresión,

$$R_t = \beta_0 + \beta_1 S_{t-6} + w_t, \quad (2.59)$$

donde R_t denota el Reclutamiento para el mes t y S_{t-6} denota el SOI seis meses antes. Suponiendo que la secuencia w_t es ruido blanco, el modelo ajustado es

$$\hat{R}_t = 65,79 - 44,28 S_{t-6} \quad (2.60)$$

con un error estándar residual de $\hat{\sigma}_w = 22,5$ en 445 grados de libertad, además el error estándar para β_1 es de 2.78. Este resultado indica la fuerte capacidad predictiva del SOI para el Reclutamiento seis meses hacia adelante. Por supuesto, sigue siendo esencial verificar las suposiciones del modelo.

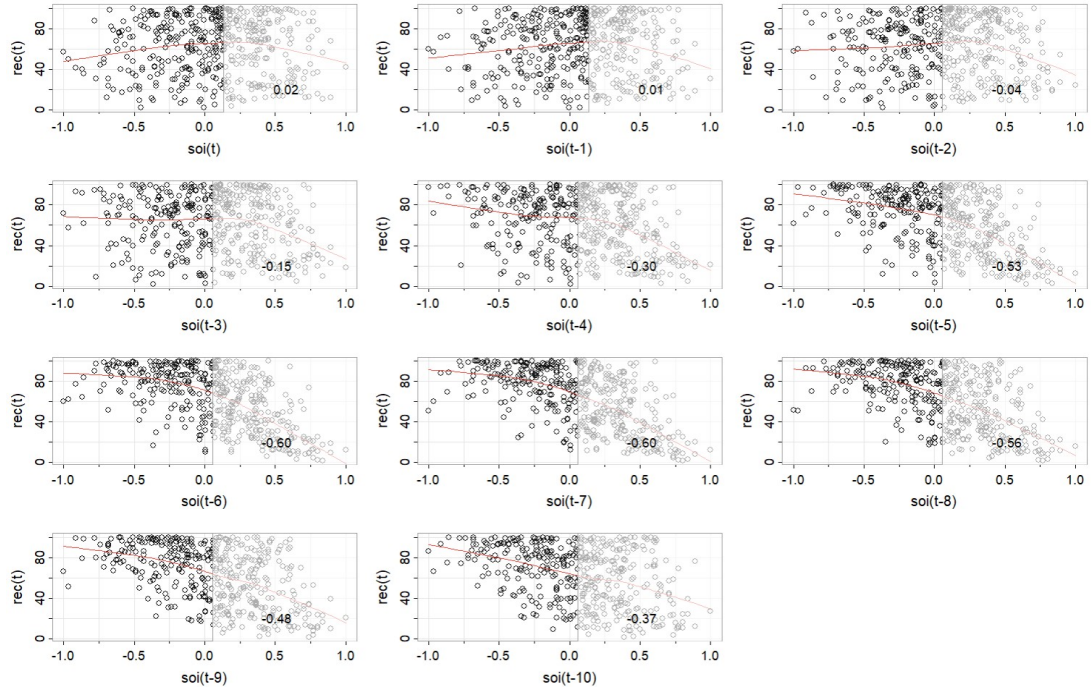


Figura 3. Diagramas de dispersión para el reclutamiento sobre distintos lags de SOI.

Del scatterplot del reclutamiento vs el SOI rezagado 6 meses se observa que la relación no es lineal y podría considerarse un modelo más complejo, por ejemplo, el siguiente:

$$R_t = \beta_0 + \beta_1 S_{t-6} + \beta_2 \cdot D_{t-6} + \beta_3 D_{t-6} S_{t-6} + w_t,$$

donde R_t es el reclutamiento, S_t es el SOI y D_t es una variable dummy que es 0 si $S_t < 0$ y 1 en caso contrario. Sin embargo, el análisis de residuos indica que los residuos no son ruido blanco. Además, la función de autocorrelación parcial (PACF) de los residuos indica que un modelo AR(2) podría ser apropiado. Asumiendo que los errores se pueden modelar como un proceso AR(2), ajustamos el modelo con dummies, y se obtiene un error estándar residual de $\hat{\sigma}_w = 21,84$ en 443 grados de libertad. El modelo ajustado sería el siguiente:

$$\hat{R}_t = 64,8028 + 1,3624 \cdot ar1 - 0,4703 \cdot ar2 + 8,6671 S_{t-6} - 2,5945 \cdot D_{t-6} - 10,3092 \cdot \text{intract} + w_t$$

donde ar1 y ar2 son los coeficientes estimados de los términos autoregresivos del modelo AR(2) para los residuos, S_{t-6} es la variable SOI rezagada seis meses, D_{t-6} es la variable dummy rezagada seis meses, intract es el término de interacción entre S_{t-6} y D_{t-6} , y w_t es el término de error (residuo).

Este modelo más complejo incluye una interacción entre S_{t-6} y D_{t-6} y muestra parámetros estimados para la tendencia autorregresiva de segundo orden (AR(2)). De nueva cuenta, el análisis de residuos y las pruebas de significancia de los coeficientes son aspectos importantes para validar este modelo.

Nos interesa una estructura más compleja para aprovechar que ambas series tienen correlación, considere los diagramas de dispersión de la variable respuesta contra las variables rezagadas de la serie x_t . En cada gráfico, la variable de reclutamiento está en el eje vertical y los distintos valores de rezago de SOI están en el eje horizontal, en este caso van desde 0 hasta un rezago de 10 (ya que son de nuestro interés aquellas que tienen rezago desde el lag $h - 5$ hasta $h = -8$. Los valores de correlación se proporcionan en cada gráfico y coinciden con los picos observados última gráfica de la figura 2 sobre la CCF muestral entre x_t e y_t .

Los valores de los rezagos $h = -5, -6, -7$ y -8 y las gráficas de dispersión reflejan que las dos series si están relacionadas, de hecho es fácil imaginar que la población de peces depende de la temperatura del océano. Esta posibilidad sugiere probar alguna versión de análisis de regresión como un procedimiento para relacionar las dos series.

Con esto en mente, sea x_t la serie SOI y y_t la serie de Reclutamiento hay muchos modelos que podríamos probar basándonos en la CCF y en los gráficos de dispersión rezagados para estos datos. Con fines demostrativos, se intentó una regresión múltiple en la que y_t , la variable de reclutamiento, es una función lineal de (rezagos pasados) 5, 6, 7, 8, 9 y 10 de la variable SOI. Ese modelo funciona tiene el R-cuadrado aproximadamente del 62 % y un error estándar residual de 17.42 con 436 grados de libertad, que mejora el modelo pasado, sin embargo se está ignorando por completo la estructura que siguen de los residuales.

La ecuación para el modelo basado en los coeficientes estimados sería:

$$\begin{aligned} \hat{R}_t = & 69,2743 - 23,8255 \cdot S_{t-5} - 15,3775 \cdot S_{t-6} - 11,7711 \cdot S_{t-7} \\ & - 11,3008 \cdot S_{t-8} - 9,1525 \cdot S_{t-9} - 16,7219 \cdot S_{t-10} + w_t \end{aligned} \quad (2.61)$$

donde R_t es la variable de reclutamiento en el tiempo t , S_t es la variable SOI en el tiempo t y w_t es el término de error en el tiempo t . Esta ecuación representa la relación lineal propuesta entre la variable de reclutamiento y varios rezagos pasados de la variable SOI.

Finalmente podríamos proporcionar un mejor modelo agregando el comportamiento de los residuales y eliminando las variables no significativas para conseguir un modelo más parsimonioso. La ecuación del modelo mejorado, considerando los coeficientes estimados significativos, sería:

$$\hat{R}_t = 8,78498 + 1,24575 \cdot ar1 - 0,37193 \cdot ar2 - 20,83776 \cdot S_{t-5} + 8,55600 \cdot S_{t-6} + w_t$$

donde R_t es la variable de reclutamiento en el tiempo t , $ar1$ y $ar2$ son rezagos pasados de la variable de reclutamiento, S_{t-5} y S_{t-6} son rezagos pasados de la variable SOI y w_t es el término de error en el tiempo t .

Este modelo mejorado ha eliminado variables no significativas y ha considerado el comportamiento de los residuales para obtener una representación más parsimoniosa y ajustada a los datos. La calidad del modelo se refleja en el alto valor del coeficiente de determinación ajustado de 0.937. Además se obtuvo un error estándar de de los residuos de 7.069 con 442 grados de libertad.

Aunque naturalmente pensamos en el SOI como la entrada y el Reclutamiento como la salida, dos configuraciones de entrada-salida son de interés. Ahora consideremos varios modelos del tipo como el presentado en 2.58. En el que entra en juego el manejo y estimación de filtros lineales. Con el SOI como entrada, el modelo es:

$$y_t = \sum_{r=-\infty}^{\infty} a_r x_{t-r} + w_t \quad (2.62)$$

mientras que un modelo que invierte los dos roles sería:

$$x_t = \sum_{r=-\infty}^{\infty} b_r y_{t-r} + v_t \quad (2.63)$$

donde w_t y v_t son procesos de ruido blanco. Aunque no hay una explicación ambiental plausible para el segundo de estos dos modelos, mostrar ambas posibilidades ayuda a decidir sobre un modelo de función de transferencia parsimonioso.

Para estos dos últimos modelos podemos aplicar el comando o función `LagReg` de la biblioteca `astsa` que permite realizar una regresión o modelado de la función de respuesta al impulso. En términos de suavizado (smoothing) y el número de términos utilizados en la regresión, los parámetros clave en la función `LagReg` son:

- **L**: Número de términos en la regresión (ancho de la regresión).
- **M**: Número de términos en la función de respuesta al impulso (ancho de la función de respuesta al impulso).
- **threshold**: Umbral para truncar coeficientes pequeños en la función de respuesta al impulso.
- **inverse**: Este parametro permite cambiar entre el modelo de la ecuación 2.62 y 2.63, si se coloca TRUE entonces se hace la regresión con el comportamiento inverso (en ambas direcciones).

Estos parámetros afectan la cantidad de suavizado y el número de términos considerados en la regresión. Un mayor valor de **L** y **M** podría conducir a un suavizado adicional y a una función de respuesta al impulso más detallada, pero también podría introducir más complejidad en el modelo. El umbral (threshold) es un criterio para ignorar coeficientes pequeños en la función de respuesta al impulso, lo que puede ayudar a simplificar el modelo.

Para modelar este ejemplo de SOI conforme a la ecuación 2.62, se utilizan $L=15$ y $M=32$, lo que implica que se están considerando 15 términos en la regresión y se están modelando 32 términos en la función de respuesta al impulso. El parámetro `threshold=6` indica que los coeficientes más pequeños que este valor se truncarán en la función de respuesta al impulso.

Basándonos en este modelo la función de respuesta al impulso para SOI, es: La ecuación completa basada en los coeficientes proporcionados por la función ‘`LagReg`’ con $M = 32$ y $L = 15$ sería:

$$\begin{aligned} \hat{y}_t &= \alpha_0 + \sum_r \alpha_r \cdot x_{t-r} \\ &= 65,97 - 18,4793 \cdot x_{t-5} - 12,263 \cdot x_{t-6} - 8,539 \cdot x_{t-7} - 6,984 \cdot x_{t-8} + w_t \end{aligned} \quad (2.64)$$

Esta ecuación representa el modelo de regresión para la serie de tiempo de reclutamiento (rec) en función de los rezagos pasados de la serie de tiempo del Índice de Oscilación del Sur (SOI), con $L = 15$ y $M = 32$, y considerando un umbral de truncamiento de 6 en la función de respuesta al impulso. El

error cuadrático medio (MSE) de este modelo es 414.08.

Para completar, si examinamos la relación inversa dada en 2.63, es decir, un modelo de regresión con la serie de reclutamiento y_t como la entrada, se usarán los parámetros $L=15$, $M=32$, $inverse=TRUE$ y un $threshold=.01$. Con este ajuste, la ecuación de predicción es:

$$\begin{aligned}\hat{x}(t) &= b_0 + \sum_r \alpha_r \cdot y_{t-r} \\ &= 0,41 + 0,016y_{t+4} - 0,02y_{t+5} + v_t\end{aligned}\quad (2.65)$$

El error cuadrático medio (MSE) es 0.07 y esta ecuación depende solo de dos coeficientes. Multiplicando ambos lados por $50B^5$ y reorganizando, obtenemos:

$$(1 - 0,8B)y_t = 20,5 - 50B^5x_t + \epsilon_t$$

Finalmente, verificamos si el ruido ϵ_t es blanco. Simplifica las cosas si volvemos a ejecutar la regresión con errores autocorrelacionados (usando la 2.62 y volvemos a estimar los coeficientes. El modelo se denomina modelo ARMAX (la X significa exógeno) se prefiere al obtener un $\sigma_w^2 = 49,93$ y el modelo es más parsimonioso. Al ajustar los errores con una estructura correlacionada se pueden modelar con un solo lag, por lo que los coeficientes del modelo ARMAX permiten este modelo final:

$$y_t = 12,33 + 0,8y_{t-1} - 2103 \cdot x_{t-5} + \epsilon_t$$

donde $\epsilon_t = 0,45\epsilon_{t-1} + w_t$ y w_t es ruido blanco con $\sigma_w^2 = 49,93$.

En este subtema centrado en modelos de regresión con rezagos, se abordó la problemática de describir la relación entre dos series temporales, y_t y x_t . Se enfatizó la relevancia de incorporar variables rezagadas en modelos estadísticos, ya que estas ofrecen información crucial para la modelación e inferencia de errores en la serie de tiempo y_t tal como pudimos observar en las diferentes modelaciones del ejemplo dado. Durante la exposición del ejemplo se trató de resaltar la importancia de mejorarlos mediante la incorporación del comportamiento de los residuales y la eliminación de variables no significativas, con el objetivo de obtener representaciones más parsimoniosas y ajustadas a los datos.

También se observa que el uso de las funciones de R son valiosas tanto para la exploración de datos como para exhibir comportamiento y validar supuestos.

Referencias

- [1] SHUMWAY, ROBERT H. & STOFFER DAVID S. (2017). *Time Series Analysis and Its Applications with R examples*. [4th. ed, Springer]
- [2] B. L. BOWERMAN, & D. S. STOFFER, & R. T. O'CONNELL, & A. B. KOEHLER (2005). *Forecasting, time series and regression. An applied approach*. [4th. ed, Thomson. Brooks Cole]