Modelo lineal General: Estructura AR(1) y Ponderada

Profesor: Alex de la Cruz H.

MODELOS LINEALES 1

EST631

Contenido

Modelo con estructura AR(1)

Un modelo de regresión lineal general es expresado por:

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times p}{\boldsymbol{X}}\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

- $\boldsymbol{y}$ es el vector de observaciones $(n \times 1)$
- $\boldsymbol{X}$ es la matriz de diseño $(n \times p)$
- $\boldsymbol{\beta}$ es el vector de coeficientes $(p \times 1)$
- $\boldsymbol{\varepsilon}$ es el vector de errores con $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ y $\mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$

Definición 1 (Modelo con estructura AR(1)).

Un modelo lineal general con errores AR(1), se caracteriza por

$$\boldsymbol{\Sigma} = \sigma^2\mathbf{V} = \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

- $\rho = \text{Corr}(\varepsilon_i, \varepsilon_{i-1}), \quad \varepsilon_i = \rho\varepsilon_{i-1} + u_i, \quad u_i \sim \mathcal{N}(0, \sigma^2)$
- $\text{Var}(\varepsilon_i) = \sigma_e^2 = \dfrac{\sigma^2}{1-\rho^2}$. Además, $\text{Cov}(\varepsilon_i, \varepsilon_j) = \rho^{|i-j|}\sigma_e^2$

La log-verosimilitud es:

$$\ell(\boldsymbol{\beta}, \sigma^2, \rho) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Para un AR(1), se tiene:

$$\ell(\boldsymbol{\beta}, \sigma^2, \rho) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{n-1}{2}\log(1-\rho^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Estimación de $\boldsymbol{\beta}$: Dado σ^2 y ρ, se maximiza ℓ respecto a $\boldsymbol{\beta}$, lo cual equivale a minimizar el cuadrado generalizado:

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}$$

Estimación de σ^2: Condicional a ρ, la estimación de σ^2 es:

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Estimación de ρ: Se obtiene maximizando el perfil de verosimilitud:

$$\ell_p(\rho) = \ell\left(\hat{\boldsymbol{\beta}}(\rho), \hat{\sigma}^2(\rho), \rho\right)$$

Esta función, ignorando constantes, es:

$$\ell_p(\rho) = -\frac{n}{2}\log\hat{\sigma}^2 - \frac{n-1}{2}\log(1-\rho^2) \quad \Rightarrow \hat{\rho} = \underset{|\rho|<1}{\arg\max}\ \ell_p(\rho)$$

Algoritmo FGLS para AR(1)

El Algoritmo de Mínimos Cuadrados Generalizados Factibles (FGLS):

1. Calcular estimador inicial: $\hat{\boldsymbol{\beta}}^{(0)} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \Rightarrow \hat{\boldsymbol{\varepsilon}}^{(0)} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(0)}$
2. Estimar ρ mediante:
$$\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum_{t=1}^{n} \hat{\varepsilon}_t^2}$$
3. Construir $\hat{\mathbf{V}}$ con $\hat{\rho}$
4. Re-estimar $\boldsymbol{\beta}$:
$$\hat{\boldsymbol{\beta}}^{(1)} = (\mathbf{X}^\top\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{V}}^{-1}\mathbf{y}$$
5. Iterar hasta convergencia

Modelo lineal general con varianza ponderada (WLS)

Consideremos el modelo lineal general:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$$

Definición 2 (WLS: Modelo con varianza ponderada).

Un modelo lineal general con varianza ponderada, se caracteriza porque $\boldsymbol{\Sigma}$ tiene la forma:

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{W}^{-1}$$

donde $\mathbf{W} = \operatorname{diag}(w_1, w_2, \ldots, w_n)$ es una matriz diagonal conocida o estimable. En este caso:

$$\operatorname{Var}(\varepsilon_i) = \frac{\sigma^2}{w_i}, \quad i = 1, \ldots, n$$

Estimación de $\boldsymbol{\beta}$

El estimador de $\boldsymbol{\beta}$ por mínimos cuadrados generalizados (GLS), se obtiene minimizando:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\top}\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\top}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{W}\mathbf{y}$$

Este estimador es insesgado y eficiente bajo el teorema de Gauss-Markov generalizado.

Estimación de σ^2

La varianza residual se estima mediante:

$$\hat{\sigma}^2 = \frac{1}{n-p}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^\top \mathbf{W}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

Este estimador es insesgado bajo los supuestos normales del modelo.

ESTIMACIÓN DE σ^2

Estimación de Pesos (W) Existen tres métodos principales:

Método	Fórmula
Información externa	$w_i = n_i$ (para datos agrupados)
Residuos al cuadrado	$w_i = 1/\hat{\varepsilon}_i^2$
Modelado de varianza	$w_i = 1/\text{Var}(\varepsilon_i \mid \mathbf{x}_i)$

Implementación matricial:

$$\mathbf{X}^\top \mathbf{W}\mathbf{X} = \sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{x}_i^\top$$

$$\mathbf{X}^\top \mathbf{W}\mathbf{y} = \sum_{i=1}^{n} w_i \mathbf{x}_i y_i$$

Ejercicios y aplicación de AR(1) y WLS

Aplicación

Utilizaremos el conjunto de datos `Orange`, incluido en `R`, que registra el crecimiento circunferencial (en mm) de 5 árboles de naranja (`Tree`) a lo largo del tiempo (`age` en días).

- Ajustar un modelo WLS que permita varianzas distintas según la edad del árbol.
- Ajustar un modelo que considere correlación temporal.
- Comparar con MCO tradicional a través de gráficos adecuados.

Desarrollo en clase

Ejercicio

Considere el modelo de regresión lineal simple

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, 2, 3$$

$$x = (1,\ 2,\ 3), \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{R}), \quad \sigma^2 = 1, \quad \rho = 0.5$$

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

Encuentre el valor de $Var(\hat{\beta}_1)$ para el modelo clásico y para el modelo propuesto (bajo autocorrelación). Compare los resultados, ¿cuál es la conclusión?.

Pruebas de heterocedasticidad y autocorrelación

Prueba de Breusch-Pagan

Considere el modelo de regresión lineal clásico:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n,$$

donde se asume que $\mathbb{E}[\varepsilon_i \mid \boldsymbol{x}_i] = 0$ y $\mathrm{Var}(\varepsilon_i \mid \boldsymbol{x}_i) = \sigma_i^2$. El supuesto de homocedasticidad establece que $\sigma_i^2 = \sigma^2$.

- $H_0 : \mathrm{V}(\varepsilon_i \mid \boldsymbol{x}_i) = \sigma^2 \qquad \text{(Homocedasticidad)}$
- $H_1 : \mathrm{V}(\varepsilon_i \mid \boldsymbol{x}_i) = \sigma^2 h(\boldsymbol{x}_i), \quad h(\boldsymbol{x}_i) = 1+\gamma_1 x_{i1}+\cdots+\gamma_k x_{ik} \quad \text{(Heterocedasticidad)}$

Prueba de Breusch-Pagan

1. Estimar el modelo por MCO y obtener los residuos $\hat{\varepsilon}_i$.
2. Calcular los residuos al cuadrado: $\hat{u}_i = \hat{\varepsilon}_i^2$.
3. Ajustar la siguiente regresión auxiliar: $\hat{u}_i = \alpha_0 + \alpha_1 x_{i1} + \cdots + \alpha_k x_{ik} + v_i$, y calcular el coeficiente de determinación R^2 de dicha regresión.
4. Calcular el estadístico de prueba de Breusch-Pagan:

$$\mathrm{BP} = nR^2 \xrightarrow{\mathcal{D}} \chi_k^2 \tag{1}$$

Se rechaza la hipótesis nula de homocedasticidad si:

$$\mathrm{BP} > \chi_{k,1-\alpha}^2.$$

Prueba de Durbin-Watson

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n,$$

donde se asume que los errores ε_i son no correlacionados y con media cero: $\mathbb{E}[\varepsilon_i] = 0$ y $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ para $i \neq j$.

- $H_0 : \rho = 0$ o equivale a decir $\text{Cov}(\varepsilon_i, \varepsilon_{i-1}) = 0 \quad \forall i$
- $H_1 : \rho \neq 0$ (o bien $\rho > 0$ si se sospecha autocorrelación positiva)

Prueba de Durbin-Watson

Sea $\hat{\varepsilon}_i$ el residuo del modelo estimado por mínimos cuadrados ordinarios.

$$DW = \frac{\sum_{i=2}^{n}(\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}. \tag{2}$$

El valor del estadístico DW se encuentra en el intervalo $[0, 4]$, con la siguiente interpretación:

- $DW \approx 2$: no hay autocorrelación.
- $DW < 2$: indicio de autocorrelación positiva.
- $DW > 2$: indicio de autocorrelación negativa.

Prueba de Durbin-Watson

Se compara el valor observado de DW con los valores críticos d_L (límite inferior) y d_U (límite superior), determinados por tablas específicas según n y k.

Condición sobre DW	Conclusión
$DW < d_L$	Hay evidencia de autocorrelación positiva
$d_L \leq DW \leq d_U$	Resultado inconcluso
$d_U < DW < 4 - d_U$	No hay evidencia de autocorrelación
$4 - d_U \leq DW \leq 4 - d_L$	Resultado inconcluso
$DW > 4 - d_L$	Hay evidencia de autocorrelación negativa

Tabla 1: Regla de decisión para el estadístico de Durbin-Watson

Aplicación de pruebas estadísticas

5. Ejercicio y aplicación

Para esta sección, usaremos dos conjuntos de datos para ilustrar las pruebas estadísticas:

- (1) `cars`: que registra 50 observaciones sobre la relación entre la velocidad de un automóvil en millas por hora (speed) y la distancia que necesita para detenerse en pies (dist).
- (2) `Orange`: contiene datos sobre el crecimiento de árboles de naranjo. 5 árboles (Tree), Edad del árbol en días desde el comienzo del estudio (age) y Circunferencia del tronco del árbol en milímetros (circumference).

Desarrollo en clase