FACULTAD DE CIENCIAS
ECONÓMICAS Y EMPRESARIALES

Ingeniería Informática + ADE

Universidad de Granada (UGR)

Autor: Ismael Sallami Moreno

Asignatura: Formulario y Resumen Econometría

Índice

1 Tema 2. Modelos de Regresión Lineal

1.1. Modelo Lineal Simple

$$Y_i \sim a + bx_i \tag{1}$$

1.2. Modelo Lineal Múltiple

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + u_i \tag{2}$$

- El subíndice i indica la observación.
- El subíndice k indica el número de variables explicativas.
- El subíndice u indica el término de error.
- Nota: β_1 es la constante, lleva asociado el valor 1 = x_{i1} (por eso no se pone).

1.3. Notación MLP

$$\vec{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \tag{3}$$

$$X = \begin{pmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{pmatrix} \tag{4}$$

$$\vec{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \tag{5}$$

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \tag{6}$$

1.3.1. Notación Extra

Matriz $X'X$

$$\begin{pmatrix} n & \sum_{i=0}^{n} X_{1i} & \cdots & \sum_{i=0}^{n} X_{ki} \\ \sum_{i=0}^{n} X_{1i} & \sum_{i=0}^{n} X_{1i}^2 & \cdots & \sum_{i=0}^{n} X_{1i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{n} X_{ki} & \sum_{i=0}^{n} X_{ki}X_{1i} & \cdots & \sum_{i=0}^{n} X_{ki}^2 \end{pmatrix} \tag{7}$$

Matriz $X'Y$

$$\begin{pmatrix} \sum_{i=0}^{n} Y_i \\ \sum_{i=0}^{n} X_{2i}Y_i \\ \vdots \\ \sum_{i=0}^{n} X_{ki}Y_i \end{pmatrix} \tag{8}$$

1.4. Supuestos

1.4.1. Linealidad

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + u_i \tag{9}$$

1.4.2. Rango Completo Por Columnas

$$rango(X) = k \tag{10}$$

1.4.3. Exogeneidad

$$E(u_i|X_i) = 0 \tag{11}$$

- El valor esperado de la esperanza que toman las variables explicativas es 0.

Por lo que si tenemos que: $Y_i = \vec{X}_i^t \vec{\beta}$, tenemos de forma inmediata que:

$$E(Y_i|X_i) = \vec{X}_i^t \vec{\beta} \tag{12}$$

1.4.4. Causalidad

- Relación de causalidad es unidireccional.
- Las variables aparecen en el lado derecho de la ecuación, pero no al contrario.
- La variable dependiente Y_i adoptará un carácter aleatorio.

1.4.5. Perturbación

- Esta centrada, es decir, $E(u_i) = 0 \forall i = 1, 2, \ldots, n$
- Es homocedástica, es decir, $Var(u_i) = E[u_i^2] = \sigma^2 \forall i = 1, 2, \ldots, n$
- Si es incorrelada implica que $cov[u_i u_j] = E[u_i u_j] = 0$. Cuando es $\neq 0$ decimos que hay un problema de autocorrelación.

Como consecuencia tenemos que:

$$var[\vec{u}] = E[(\vec{u} - E[\vec{u}])(\vec{u} - E[\vec{u}])^t] = E[\vec{u}\vec{u}^t] \tag{13}$$

Para la demostración de esta operación pincha aquí para verla.

1.4.6. Normalidad de la perturbación

Se considera que la perturbación aleatoria u_i se distribuye normalmente con media 0 y varianza σ^2, es decir, $u_i \sim N(0, \sigma^2) \forall i = 1, \ldots, n$.

En este caso supondremos que $\vec{u}$ se distribuye normalmente, es decir, $\vec{u} \sim N(\vec{0}, \sigma^2 I_n)$[1].

1.5. Estimación Míminos Cuadrados Ordinarios (MCO)

Para estimar los parámetros β utilizaremos el MCO, denotando a $\hat{\vec{y}}$ el valor ajustado de $\vec{y}$, es decir, $\hat{\vec{y}} = X\hat{\vec{\beta}}$.

En este caso el valor de los residuos se obtiene como:

$$\vec{e} = \vec{y} - \hat{\vec{y}} = \vec{y} - X\hat{\vec{\beta}} \tag{14}$$

Debemos de minimizar la suma de los cuadrados de los residuos:

$$\min_{\vec{\beta}} \sum_{i=1}^{n} e_i^2 \tag{15}$$

Tendremos como función a minimizar: $f(\hat{\vec{\beta}}) = \vec{e}^{\,t}\vec{e}$

1.5.1. Pasos a seguir

1. Definimos la expresión a minimizar. Se debe de definir la suma de los cuadrados de los residuos:
$$f(\vec{\beta}) = (\vec{y} - X\vec{\beta})^t(\vec{y} - X\vec{\beta}) \tag{16}$$
2. Derivamos con respecto a $\vec{\beta}$ e igualamos a 0:
$$\frac{\partial f(\vec{\beta})}{\partial \vec{\beta}} = 0 \tag{17}$$

[1] distribución normal multivariante con vector de medias nulo y matriz de varianzas-covarianzas $\sigma^2 I_n$

3. Despejamos el vector de parámetros, debemos de considerar el supuesto del rango completo por columnas, verificándose que $\rho(X^tX) = 1$, condición necesaria para que exista $(X^tX)^{-1}$

4. Para que la solución sea mínimo se ha de cumplir:

$$\text{Hess}(f) = 2X^tX > 0 \text{ and } (A > 0 \text{ si } z^tAz > 0\ \forall z \neq 0) \tag{18}$$

Si el modelo tiene término independiente entonces se tiene

$$\vec{\beta} = \begin{pmatrix} n & \sum_{i=1}^n X_{2i} & \cdots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{2i}^2 & \cdots & \sum_{i=1}^n X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ki} & \sum_{i=1}^n X_{ki}X_{2i} & \cdots & \sum_{i=1}^n X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{2i}Y_i \\ \vdots \\ \sum_{i=1}^n X_{ki}Y_i \end{pmatrix}$$

En el caso de que el modelo carezca del término independiente entonces:

$$\vec{\beta} = \begin{pmatrix} \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i}X_{2i} & \cdots & \sum_{i=1}^n X_{1i}X_{ki} \\ \sum_{i=1}^n X_{2i}X_{1i} & \sum_{i=1}^n X_{2i}^2 & \cdots & \sum_{i=1}^n X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ki}X_{1i} & \sum_{i=1}^n X_{ki}X_{2i} & \cdots & \sum_{i=1}^n X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n X_{1i}Y_i \\ \sum_{i=1}^n X_{2i}Y_i \\ \vdots \\ \sum_{i=1}^n X_{ki}Y_i \end{pmatrix}$$

1.6. Estimación Míminos Cuadrados Ordinarios Restringidos(MCO(R))

$$\widehat{\beta_R} = \widehat{\beta} - (R(X'X)^{-1})^t \times [R(X'X)^{-1}R']^{-1} \times R\widehat{\beta} \tag{19}$$

1.7. Propiedades Algebraicas de los MCO

Considerando que SI hay ti

- Las variables exógenas son ortogonales al vector de los residuos: $X^t\vec{e} = \vec{0}$

  Demostración

  1. Tenemos que $-2X^t\vec{y} + 2X^tX\vec{\hat{\beta}} = \vec{0}$
  2. Si operamos nos queda que $X^t(\vec{y} - X\vec{\hat{\beta}}) = \vec{0}$
  3. Aplicando que $\vec{\hat{y}} = X\vec{\hat{\beta}}$
  4. Queda demostrado que $X^t\vec{e} = \vec{0}$

- La suma de los residuos mínimos cuadráticos es 0: $\sum_{i=1}^n e_i = 0$

- La suma de los valores observados coincide con la suma de los valores estimados: $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$

- Suma Cuadrados
  - $SCT = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \vec{y}\,^t\vec{y} - n \cdot \bar{Y}^2$
  - $SCR = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \vec{y}\,^t\vec{y} - \vec{\beta}^t\vec{X}^t\vec{y}$
  - $SCE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \vec{\beta}^{t2}\vec{X}^t\vec{y} - n \cdot \bar{Y}^2$
- SCT = SCR + SCE
- $\vec{\hat{y}}\vec{e} = 0$

1.8. Bondad del Ajuste

$$R^2 = \frac{SCE}{SCT} = \frac{\sum_{i=1}^{n}(\hat{Y} - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{20}$$

NOTA IMPORTANTE:

Si el modelo tiene término independiente entonces R^2 se puede expresar como:

$$R^2 = 1 - \frac{SCR}{SCT} \tag{21}$$

1.9. Bondad del Ajuste Corregido

$$\bar{R^2} = 1 - \frac{SCR/(n-k)}{SCT/n-1} = 1 - (1 - R^2) * \frac{n-1}{n-k} \tag{22}$$

1.10. Criterio de Akaike

$$AIC = \ln(\frac{SCR}{n}) + \frac{2k}{n} \tag{23}$$

1.11. Criterio de Schwarz

$$BIC = \ln(\frac{SCR}{n}) + \frac{k\ln(n)}{n} \tag{24}$$

[2]En la práctica la simbología de t = traspuesta = *prima* = ′

2 Tema 3: Estimación y Contraste de Hipótesis

2.1. Valor Esperado y Varianza de los EMCO (Estimadores mínimos cuadrados ordinarios)

2.1.1. Valor Esperado de los EMCO

1. El estimador de $\vec{\beta}$ del modelo $\vec{y} = X\vec{\beta} + \vec{u}$ es:

$$\hat{\vec{\beta}} = (X^t X)^{-1} X^t \vec{y}$$

2. Si en esta sustituimos $\vec{y} = X\vec{\beta} + \vec{u}$ nos quedaría:

$$\hat{\vec{\beta}} = (X^t X)^{-1} X^t (X\vec{\beta} + \vec{u}) = \vec{\beta} + (X^t X)^{-1} X^t \vec{u}$$

3. Por lo que $E[\hat{\vec{\beta}}] = \vec{\beta}$

   - Debido a que $E[\vec{u}] = \vec{0} \rightarrow \hat{\vec{B}}$ **es un estimador insesgado de** $\vec{\beta}$

2.1.2. Varianza de los EMCO

1. $\text{var}(\hat{\vec{\beta}}) = \sigma^2 (X^t X)^{-1}$ Para ello debemos de seguir estos pasos:

   a) La varianza es igual a:

   $$\text{var}(\hat{\vec{\beta}}) = E[(\hat{\vec{\beta}} - E[\hat{\beta}])(\hat{\vec{\beta}} - E[\hat{\beta}])^t]$$

   b) Si en la ecuación anterior sustituiimos $\hat{\vec{\beta}} = \vec{\beta} + (X^t X)^{-1} X^t \vec{u}$ nos queda:

   $$\text{var}(\hat{\vec{\beta}}) = E[((X^t X)^{-1} X^t \vec{u})((X^t X)^{-1} X^t \vec{u})^t]$$

   c) Acto seguido, debemos de aplicar la propiedad de las matrices de $(AB)^t = B^t A^t$:

   $$\text{var}(\hat{\vec{\beta}}) = E[(X^t X)^{-1} X^t \vec{u}\vec{u}^t X (X^t X)^{-1}]$$

   d) Si aplicamos la propiedad de la esperanza de un producto de matrices $E[AB] = E[A]E[B]$:

   $$\text{var}(\hat{\vec{\beta}}) = (X^t X)^{-1} X^t E[\vec{u}\vec{u}^t] X (X^t X)^{-1}$$

   e) Si aplicamos la propiedad de la varianza de un vector $E[\vec{u}\vec{u}^t] = \sigma^2 I_n$:

   $$\text{var}(\hat{\vec{\beta}}) = \sigma^2 (X^t X)^{-1} X^t X (X^t X)^{-1} = \sigma^2 (X^t X)^{-1}$$

   f) Con esto quedaría demostrado que $\text{var}(\hat{\vec{\beta}}) = \sigma^2 (X^t X)^{-1}$

2.2. Teorema de Gauss-Markov

Según este teorema tenemos que los estimadores de los mínimos cuadrados ordinarios son:

- óptimos
- lineales
- insesgados

Nota:

Podemos decir que de todos los estimadores de $\vec{\beta}$, el EMCO es el que presenta *menor matriz de varianzas-covarianzas*.

2.2.1. Demos

- **Demo I**: $\hat{\vec{\beta}}$ es lineal con respecto a $\vec{y}$

  Demostración

  1. si definimos una matriz de dimensiones k*n $W = (X^tX)^{-1}X^t$
  2. es fácil demostrar que que $\vec{B} = W\vec{y}$ ya que $\vec{B} = (X^tX)^{-1}X^t\vec{y}$

- **Demo II**: $\hat{\vec{\beta}}$ es insesgado de $\vec{\beta}$[3]

- **Demo III**: $\hat{\vec{\beta}}$ es un estimador óptimo de $\vec{\beta}$

  Demostración

  - Esperanza

  Para la demostración debemos de seguir los siguientes pasos:

  1. Suponemos un estimador alternativo $\hat{\vec{\beta}}^*$
  2. Como se cumple que es lineal respecto de $\vec{\beta}$, entonces tenemos que $\hat{\vec{\beta}}^* = C\vec{y}$, es decir, existe un C que cumple la anterior condición de dimensiones k*n
  3. Además ha de ser insesgado, por lo que tenemos que $E[\hat{\vec{\beta}}^*] = \vec{\beta}$
  4. Si seguimos resolviendo tenemos que:

$$E[\hat{\vec{\beta}}^*] = E[C\vec{y}]$$

[3] Queda demostrado anteriormente ya que se verificaba que la esperanza del parámetro a estimar era igual al parámetro.

5. Aplicamos que $\vec{y} = X\vec{\beta} + \vec{u}$:
$$E[C\vec{y}] = E[C(X\vec{\beta} + \vec{u})] = E[CX\vec{\beta} + C\vec{u}] = CX\vec{\beta} + CE[\vec{u}]$$
6. Como sabemos que $E[\vec{u}] = 0$, nos queda:
$$CX\vec{\beta} + CE[\vec{u}] = CX\vec{\beta}$$
7. Volvemos al tema de insesgadez, para que sea insesgado (que se debe de cumplir en este punto) debe de que $CX = I_k$
8. Por lo que tenemos de nuevo nuestro estimador anterior, podemos concluir la demostración afirmando que $\vec{\hat{\beta}}$ es un estimador óptimo de $\vec{\beta}$

- Varianza

- Si la calulamos de igual manera a la anterior y usando la matriz C y aplicando que $E[(\widehat{\beta^*}^t) - E[\widehat{\beta^*}^t]] = C\vec{u}$, nos queda que:
$$= E[C\vec{u} * \vec{u}^t C^t] = C\sigma^2 I_n C^t = \sigma^2 CC^{t4}$$
- Sabiendo que $\vec{\hat{\beta}} = W\vec{y}$ y $\vec{\hat{\beta}}^* = C\vec{y}$
- Denotamos D = C-W
- Tenemos que:
$$DX = (C - W)X = CX - WX = CX - (X^tX)^{-1}X^tX = I_k - I_k = 0_k$$
- podemos afirmar además, que en base a lo definido anteriormente:
$$\text{var}(\vec{\hat{\beta}}^*) = \sigma^2 CC^t = \sigma^2 (W + D)^5 (W + D)^t = \sigma^2 (WW^t + WD^t + DW^t + DD^t)^6$$

Comparación de varianzas

  - $WW^t = ((X^tX)^{-1}X^t)((X^tX)^{-1}X^t)^t = (X^tX)^{-1}X^tX(X^tX)^{-1} = (X^tX)^{-1}$
  - $WD^t = ((X^tX)^{-1}X^t)(D)^t = (X^tX)^{-1}X^t(C - W)^t = (X^tX)^{-1}X^t(C^t - W^t) = (X^tX)^{-1}X^tC^t - (X^tX)^{-1}X^tW^t = 0$
  - $DW^t = D((X^tX)^{-1}X^t)^t = 0$
- Aplicando esto nos queda que:
$$\text{var}(\vec{\hat{\beta}}^*) = \sigma^2 DD^t + \text{var}(\vec{\hat{\beta}})$$
- Como sabemos el primer término es siempre positivo, por lo que podemos afirmar que:
$$\text{var}(\vec{\hat{\beta}}^*) - \text{var}(\vec{\hat{\beta}}) = \sigma^2 DD^t \geqslant 0 \to \text{var}(\vec{\hat{\beta}}^*) \geqslant \text{var}(\vec{\hat{\beta}})$$
- Vemos que la varianza también es mayor, por lo que podemos concluir que $\vec{\hat{\beta}}^*$ es un estimador peor que $\vec{\hat{\beta}}$

[4] Hemos usado que $E[\vec{u}\vec{u}^t] = \sigma^2 I_n$

[5] Usamos que $C = W + D$

[6] De la anterior forma también es válida esta es más extensa, esta sirve para comparar las varianzas como vamos a ver a continuación

2.3. Estimación de σ_u^2

Sabemos que el vector de los residuos es;

$$\vec{e} = \vec{y} - \hat{\vec{y}} = \vec{y} - X\hat{\vec{\beta}}$$

Si desarrollamos nos queda que:

$$(I_n - X(X^tX)^{-1}X^t)\vec{y} = M_x\vec{y}$$

Para ello debemos de aplicar que:

$$\hat{\vec{y}} = X\hat{\vec{\beta}} = X(X^tX)^{-1}X^t\vec{y}$$

Y sacar factor común $\vec{y}$.

La matriz M_x se conoce como la matriz complemento del proyector ortogonal, cuyas propiedades son:

- simétrica[7]
- idempotente[8]
- traza(M_x) = n - k
- M_xX = X - X = 0

Si sustituimos $\vec{y} = X\vec{\beta} + \vec{u}$ en la ecuación anterior nos queda que:

$$\vec{y} = M_x\vec{y} = M_xX\vec{\beta} + M_x\vec{u} = M_x\vec{u}^9$$

Si tomamos valores esperados obtenemos que ($E[\vec{e^t}\vec{e}] =$):

1. $E[(M_x\vec{u})^t(M_x\vec{u})]$
2. = $E[\vec{u}^tM_x\vec{u}]^{10}$
3. = $E[\text{traza}(\vec{u^t}M_x\vec{u})]$
4. = $\text{traza}(M_xE[\vec{u^t}\vec{u}])$
5. = $\text{traza}(\sigma^2M_x)$
6. = $\sigma^2\text{traza}(M_x)$
7. = $\sigma^2(n-k)$

Finalmente, nos queda que $E[\vec{e^t}\vec{e}] = \sigma^2(n-k)$, por lo que como $E[\frac{\vec{e^t}\vec{e}}{n-k}] = \sigma^2$ podemos afirmar que $\hat{\sigma}^2 = \frac{\vec{e^t}\vec{e}}{n-k} = \frac{SCR}{n-k}$ es un estimador insesgado de σ_u^2

[7] Una matriz es simétrica si es igual a su traspuesta $A = A^t$

[8] Una matriz es idempotente si al elevarla al cuadrado nos da la misma matriz $A^2 = A$

[9] Ya que $M_xX = 0$

[10] Ya que M_x es simétrica e idempotente

2.4. Estimación de la varianza de los EMCO

Como sabemos la varianza del EMCO $\vec{\beta}$ es $\sigma^2(X^tX)^{-1}$, por lo que depende de la varianza de la perturbación y como sabemos el estimador de la varianza de la perturbación es: $\hat{\sigma}^2 = \frac{SCR}{n-k}$.

Por lo que nos queda que la varianza del EMCO es:

$$\operatorname{var}(\hat{\vec{\beta}}) = \frac{SCR}{n-k} * (X^tX)^{-1} \tag{25}$$

2.5. Supuesto de Normalidad e Inferencia

Supuesto de normalidad: Supongamos que $\vec{u} \sim N(\vec{0}, \sigma^2 I_n)$.
Propiedades del estimador:

- Dado que $\hat{\vec{\beta}} = \vec{\beta} + (X^tX)^{-1}X^t\vec{u}$, por el Teorema de Gauss-Markov, se tiene que:

$$\hat{\vec{\beta}} \sim N(E[\hat{\vec{\beta}}], \operatorname{var}(\hat{\vec{\beta}})) = N(\vec{\beta}, \sigma^2(X^tX)^{-1}).$$

- Equivalentemente, se cumple:

$$(\hat{\vec{\beta}} - \vec{\beta}) \sim N(\vec{0}, \sigma^2(X^tX)^{-1}).$$

Distribución del residuo:

- Como se verifica que $\vec{e}\,^t\vec{e} = \vec{u}^tM_X\vec{u}$, y dado que $\vec{u} \sim N(\vec{0}, \sigma^2 I_n)$, y que M_X es simétrica, idempotente y tiene rango $\rho(M_X) = n - k$, concluimos que:

$$\frac{1}{\sigma^2}\vec{u}^tM_X\vec{u} \sim \chi^2_{n-k}.$$

- Por tanto, se tiene que:

$$\frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}.$$

Más desarrollado, paso a paso:

Supuesto de Normalidad e Inferencia

Supuesto de normalidad: Supongamos que $\vec{u} \sim N(\vec{0}, \sigma^2 I_n)$.
Paso 1: Propiedades del estimador $\hat{\vec{\beta}}$

- Recordemos que el estimador $\hat{\vec{\beta}}$ se define como:

$$\hat{\vec{\beta}} = \vec{\beta} + (X^tX)^{-1}X^t\vec{u}.$$

- Dado que $\vec{u} \sim N(\vec{0}, \sigma^2 I_n)$, el término $(X^tX)^{-1}X^t\vec{u}$ también tiene distribución normal, ya que es una combinación lineal de $\vec{u}$.

- Por lo tanto, $\hat{\vec{\beta}}$ sigue una distribución normal:

$$\hat{\vec{\beta}} \sim N(\vec{\beta}, \sigma^2(X^tX)^{-1}).$$

Paso 2: Distribución de $\hat{\vec{\beta}} - \vec{\beta}$

- Al restar $\vec{\beta}$ del estimador $\hat{\vec{\beta}}$:

$$\hat{\vec{\beta}} - \vec{\beta} = (X^tX)^{-1}X^t\vec{u}.$$

- Este término sigue una distribución normal con media cero y varianza $\sigma^2(X^tX)^{-1}$:

$$(\hat{\vec{\beta}} - \vec{\beta}) \sim N(\vec{0}, \sigma^2(X^tX)^{-1}).$$

Paso 3: Distribución del residuo $\vec{e}$

- Recordemos que el residuo se define como:

$$\vec{e} = \vec{y} - \hat{\vec{y}} = \vec{y} - X\hat{\vec{\beta}}.$$

- Sustituyendo $\vec{y} = X\vec{\beta} + \vec{u}$:

$$\vec{e} = \vec{y} - X\hat{\vec{\beta}} = (X\vec{\beta} + \vec{u}) - X\hat{\vec{\beta}}.$$

- Simplificando:

$$\vec{e} = \vec{u} - X(\hat{\vec{\beta}} - \vec{\beta}).$$

- Como $\hat{\vec{\beta}}$ depende de $\vec{u}$, al aplicar propiedades de matrices y el operador $M_X = I_n - X(X^tX)^{-1}X^t$:

$$\vec{e} = M_X\vec{u}.$$

Paso 4: Distribución de $\vec{e}^t\vec{e}$

- El término $\vec{e}^t\vec{e}$ puede escribirse como:

$$\vec{e}^t\vec{e} = \vec{u}^t M_X \vec{u}.$$

- Dado que $\vec{u} \sim N(\vec{0}, \sigma^2 I_n)$ y que M_X es simétrica, idempotente, y tiene rango $n-k$:

$$\frac{1}{\sigma^2}\vec{u}^t M_X \vec{u} \sim \chi^2_{n-k}.$$

- Por lo tanto, podemos concluir que:

$$\frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k},$$

  donde $\hat{\sigma}^2 = \frac{\vec{e}^t\vec{e}}{n-k}$ es el estimador insesgado de σ^2.

2.6. Intervalos de Confianza

2.6.1. IC para β_i

$$\hat{\beta_i} \pm t_{n-k,1-\frac{\alpha}{2}} * \sqrt{\hat{var}\hat{\beta_i}}\ , i = 1, \ldots, k. \tag{26}$$

2.6.2. IC para σ^2

$$\left[\frac{(n-k)\sigma^2}{\chi^2_{n-k,1-\frac{\alpha}{2}}}, \frac{(n-k)\sigma^2}{\chi^2_{n-k,\frac{\alpha}{2}}}\right] \tag{27}$$

2.7. Contraste de hipótesis

$$H_0 : R^{11}\vec{\beta} = \vec{r}$$

donde:

$$R = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mk} \end{pmatrix}_{m\times k} \quad , \quad \vec{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix}_{m\times 1} \tag{28}$$

2.7.1. Estadístico de contraste

$$F_{exp} \sim F_{m,n-k}{}^{12} \tag{29}$$

Como F_{exp} tenemos:

$$F_{exp} = (R\vec{\hat{\beta}}) - \vec{r}^t \times \frac{[R(X^tX)^{-1}R^t]^{-1}}{m\hat{\sigma^2}} \times (R\vec{\hat{\beta}} - \vec{r}) \tag{30}$$

- Siendo m el número de restricciones.
- Siendo n el número de observaciones.
- Siendo k el número de variables explicativas.

¿Cuando se rechazará?

Cuando $F_{exp} > F_{m,n-k,1-\alpha}$.

[11] En la práctica hace referencia a la matriz de restricciones que se saca mediante el enunciado y/o los datos que se proporcionan

[12] Cabe destacar que el estadístico desarrollado se encuentra en las diapositivas de la asignatura, en específico en la diapositiva 15/34 del tema 3

2.7.2. Casos particulares

Contraste de significación individual

Si m = 1 y $r_i = 0 \ \forall i$, tenemos las siguientes hipótesis:

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases} \tag{31}$$

Con un estadístico de contraste:

$$t_{exp} = \frac{\hat{\beta_i}}{\sqrt{\hat{var}\hat{\beta_i}}} \tag{32}$$

Se rechazaría la hipótesis nula si $|t_{exp}| > t_{n-k,1-\frac{\alpha}{2}}$.

Única restricción lineal

Cuando m = 1 y $r_i \neq 0 \ \forall i$, se tiene un contraste de hipótesis de la forma:

$$\begin{cases} H_0 : \beta_i = r_i \\ H_1 : \beta_i \neq r_i \end{cases} \tag{33}$$

Con un estadístico de contraste:

$$t_{exp} = \frac{\hat{\beta_i} - r_i}{\sqrt{\hat{var}\hat{\beta_i}}} \tag{34}$$

Se rechazaría la hipótesis nula si $|t_{exp}| > t_{n-k,1-\frac{\alpha}{2}}$.

Contraste de significación global

Si m = k-1 y $r_i = 0 \ i = 2, 3, \ldots, k$.

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \text{ con } i = 2, 3, \ldots, k \end{cases} \tag{35}$$

Con un estadístico de contraste:

$$F_{exp} = (R\vec{\hat{\beta}})^t \times \frac{[R(X^tX)^{-1}R^t]^{-1}}{(k-1) \times \sigma^2} \times (R\vec{\hat{\beta}}) \quad \text{ó} \quad F_{exp} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} \tag{36}$$

Se rechazaría la hipótesis nula si $F_{exp} > F_{k-1,n-k,1-\alpha}$.

Nota: Según la cronología de la asignatura, ahora tocaría la demostración del contraste de significación global, pero en clase se afirmó que no se iba a preguntar en el examen.

ANOVA

Teniendo en cuenta que se verifica:

1. $\vec{y}^{\,t}A\vec{y} = \hat{\vec{\beta}}_2^{\,t}X_2^tAX_2\hat{\vec{\beta}}_2 + \vec{e}^{\,t}\vec{e}$
2. SCT = SCR + SCE

ANOVA $\rightarrow$ Análisis de la varianza donde se estudia la significación global del modelo donde[13]:

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \text{ con } i = 2, 3, \ldots, k \end{cases} \qquad F_{exp} = \frac{SCE/k-1}{SCR/n-k} \quad \text{Se rechazará } H_0 \text{ si } F_{exp} > F_{k-1,n-k,1-\alpha} \tag{37}$$

2.8. Cota R^2

Se rechazará la hipótesis nula si $F_{exp} > F_{k-1,n-k,1-\alpha}$:

$$\frac{n-k}{k-1}\frac{R^2}{1-R^2} > F_{k-1,n-k,1-\alpha} \tag{38}$$

2.9. Predicción puntual

2.9.1. Para el valor individual

Suponiendo que tenemos permanencia estructural[14], podemos tomar como predictor para el valor medio y para el valor individual de la variable endógena:

$$\vec{x_0}^t\hat{\vec{\beta}} = \hat{y_0} = E[\widehat{y_0 \mid \vec{x_0}}]^{15} \tag{39}$$

Tenemos que:

- $E[e_0] = 0$[16]
- $\text{var}(e_0) = \sigma^2(1 + \vec{x_0}^t(X^tX)^{-1}\vec{x_0})$[17]

2.9.2. Para el valor esperado

$e_0^* \sim N(0, \sigma^2(1 + \vec{x_0}^t(X^tX)^{-1}\vec{x_0}))$[18]

[13] Para un extra se puede mirar la diapositiva 24/34 del tema 3

[14] En este contexto, la permanencia estructural se refiere a la estabilidad de los parámetros del modelo a lo largo del tiempo. El proceso de generación de los datos para la nueva observación $\vec{x_0}$ es el mismo que ha generado la información muestral

[15] El predictor mínimo cuadrático es lineal, insesgado y óptimo

[16] La demostración se encuentra en la diapositiva 27/34 del tema 3

[17] La demostración se encuentra en la diapositiva 28/34 del tema 3

[18] La demostración se encuentra en la diapositiva 29/34 del tema 3

2.9.3. Por intervalo

- Intervalo de confianza para el valor individual y_0:

$$IC_{y_0} = \hat{y_0} \pm t_{n-k,1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2(1 + \vec{x_0}^t (X^t X)^{-1} \vec{x_0})} \tag{40}$$

Teniendo en cuanta que podemos sacar el sigma de raíz y de que $\hat{y_0} = \vec{x_0}^t \vec{\hat{\beta}}$, nos queda:

$$IC_{y_0} = \vec{x_0}^{t[19]} \vec{\hat{\beta}} \pm t_{n-k,1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{1 + \vec{x_0}^t (X^t X)^{-1} \vec{x_0}} \tag{41}$$

- Intervalo de confianza para el valor esperado $E[y_0 \mid \vec{x_0}]$:

$$IC_{E[y_0|\vec{x_0}]} = \vec{x_0}^t \vec{\hat{\beta}} \pm t_{n-k,1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{\vec{x_0}^t (X^t X)^{-1} \vec{x_0}} \tag{42}$$

2.10. Construcción de las distribuciones

2.10.1. Distribución χ^2 de Pearson

Suma de los cuadrados de n variables aleatorias independientes e idénticamente distribuidas que siguen una distribución normal estándar[20].

$$\chi^2_n = \sum_{i=1}^{n} X_i^2 \quad \text{donde } X_i \sim N(0,1) \quad \forall i \tag{43}$$

- No simétrica
- No negativa
- No acotada
- Su media es n y su varianza es $2n$

2.10.2. Distribución t de Student

Cociente entre una variable aleatoria normal estándar y la raíz cuadrada de una χ^2 de Pearson dividida por sus grados de libertad, siendo ambas independientes.

$$t_n = \frac{Z}{\sqrt{\frac{X}{n}}} \quad \text{donde } Z \sim N(0,1) \text{ e } X \sim \chi^2_n \tag{44}$$

- Simétrica
- Para n>30 se puede aproximar a una normal

[19] Este término se refiere, por simplicidad, a algún punto en específico, es decir, en la práctica corresponde con unos valores que nos da el enunciado para sacar el IC en base a esos valores.

[20] $N(0,1)$

- Su media es 0
- Su varianza es $\frac{n}{n-2}$
- Acotada

2.10.3. Distribución F de Fisher-Snedecor

Cociente entre dos χ^2 de Pearson divididas por sus grados de libertad, siendo ambas independientes.

$$F_{m,n} = \frac{X_1/m}{X_2/n} \quad \text{donde } X_1 \sim \chi^2_m \text{ e } X_2 \sim \chi^2_n \tag{45}$$

- No simétrica
- $F_{m,n,1-\alpha} = \frac{1}{F_{n,m,\alpha}}$
- $t^2_n = F_{1,n}$

3 Tema 4: Multicolinealidadad

3.1. Concepto

Existencia de relaciones entre dos o más variables independientes del modelo lineal uniecuacional múltiple.

3.2. Clasificación

- **Multicolinealidadad exacta:** Existencia de una relación exacta entre dos o más variables independientes. Las **causas** son básicamente el incumplimientos de una de las hipótesis del modelo uniecuacional múltiple, la matriz X no es de rango completo por columnas: $\mathrm{rango}(X) < k$. Esto tiene como **consecuencias** que no se pueda invertir la matriz X^tX y que no se pueda estimar el modelo, haciendo que el sistema sea indeterminado y por ende, no haya una solución única para $\hat{\beta}$, sino que haya infinitas soluciones.
- **Multicolinealidadad aproximada:** Existencia de una relación aproximada entre dos o más variables independientes. Las causas son:
  - relación causal entre las variables explicativas del modelo.
  - escasa variabilidad de las observaciones.
  - tamaño de la muestra pequeño.

3.2.1. Clasificaciones de la multicolinealidad aproximada

- Según *Spanos y McGuirk*:
  - Sistemática: alta correlación lineal entre las variables exógenas.
  - Errática: mal condicionamiento de los datos considerados.
- Según *Marquardt y Snee*:
  - No esencial: relacion lineal de las variables exógenas, incluyendo la constante.
  - Esencial: relación lineal de las variables exógenas sin incluir la constante.

  Las **consecuencias** de la multicolinealidad aproximada son:
  - En este caso si se puede invertir la matriz X^tX, pero el determinante de la matriz es muy pequeño, lo que hace que la varianza de los estimadores sea muy grande, lo que hace que los estimadores sean poco precisos.
  - Contrastes individuales no rechazaran la hipótesis nula y los globales si.
  - Coeficientes estimados serán sensibles a pequeñas variaciones en los datos.
  - Coeficiente de determinación alto.

3.3. Factor de inflación de la varianza

Sirve para detectar el grado de multicolinealidad de las variables, se define como:

$$FIV_i = \frac{1}{1 - R_i^2} \quad i = 2, 3, \dots, p \tag{46}$$

Si esta medida es superior a 10, se considera que hay multicolinealidad de grado preocupante. Al no tener en cuenta la relación de las varibales exógenes con la constante no detecta si hay multicolinealidad no esencial.

Nota: No detecta la multicolinealidad entre variables cualitaticas ni dicotómicas.

3.4. Número de condición

Se define como:

$$\text{Número de condición} = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \tag{47}$$

Donde λ_{max} y λ_{min} son los autovalores máximos y mínimos de la matriz X^tX.Si el número de condición alcanza valores superiores a 20 es moderado y si el número de condición es mayor que 30, se considera que hay multicolinealidad de grado preocupante. *En este caso si tiene en cuenta la relación de las variables exógenas con la constante.*

3.5. Otras Medidas

3.5.1. Matriz de Correlaciones (R)

Ignora la constante y solo es capaz de detectar la multicolinealidad aproximada del tipo esencial dos a dos.

3.5.2. Determinante de la matriz de correlaciones *det(R)*

Sigue ignorando la relación con la constante, por lo que solo detecta la multicolinealidad aproximada del tipo esencial. En este caso valores cercanos a 0 indican un caso preocupante.

3.5.3. Coeficiente de variación de las variables explicativas

Detectar una escasa variabilidad de las observaciones, lo que puede ser una causa de la multicolinealidad aproximada del tipo no esencial.

3.6. Soluciones

- Mejora del diseño de la investigación, extrayendo información máxima de las variables observadas.
- Eliminación de variables redundantes o que sean sospechosas en cuanto a las causantes de la multicolinealidad.
- Aumentar el tamaño de la muestra.
- Utilizar la relación extramuestral entre las variables, es decir, utilizar información de otras fuentes.

4 Tema 5: Heteroscedasticidad

4.1. Concepto

En este caso la matriz de varianzas–covarianzas puede tomar la siguiente forma:

$$\mathrm{var}(u) = \sigma^2\Omega^{21} \tag{48}$$

Se dice que un modelo presenta Heteroscedasticidad cuando la varianza de la perturbación no permanece constante a lo largo del tiempo.

$$\sigma^2\omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \tag{49}$$

[21] donde ω es una matriz simétrica y definida positiva

4.2. Causas de la Heteroscedasticidad

La heteroscedasticidad ocurre cuando la varianza del término de error no es constante a lo largo de las observaciones en un modelo de regresión. A continuación, se enumeran algunas de las principales causas:

1. **Variación en la escala de la variable dependiente:** En datos de sección cruzada, la escala de la variable dependiente y el poder explicativo del modelo pueden variar entre las observaciones, provocando diferencias en la varianza del error.

2. **Agrupación de datos:** Si las observaciones se agrupan en categorías y se reemplazan los datos originales (que son homoscedásticos) por las medias aritméticas de cada categoría, estas pueden presentar heteroscedasticidad.

3. **Omisión de variables relevantes:** Si una variable relevante no se incluye en el modelo, el término de error puede depender de dicha variable omitida, lo que genera que su varianza no sea constante. Por ejemplo, en un modelo que relaciona el consumo familiar con la renta familiar, omitir factores como el tamaño del hogar puede introducir heteroscedasticidad.

4. **Estructura de los datos:** En series temporales o datos panel, factores como cambios estructurales, shocks económicos o eventos externos pueden provocar que la varianza del error varíe con el tiempo o entre grupos.

5. **Relación no lineal entre variables:** Cuando las variables explicativas están relacionadas de manera no lineal con la variable dependiente, la dispersión de los errores tiende a cambiar a lo largo del rango de las observaciones.

Estas causas reflejan que la heteroscedasticidad está asociada a la estructura del modelo, los datos y la especificación del mismo.

4.3. Consecuencias de la Heteroscedasticidad

La heteroscedasticidad tiene varias implicaciones importantes en los modelos de regresión estimados mediante el método de mínimos cuadrados ordinarios (MCO). A continuación, se detallan las consecuencias principales:

1. **Linealidad e insesgadez:** El estimador por MCO, dado por la expresión:

$$\hat{\beta}_{\text{MCO}} = (X^\top X)^{-1} X^\top y,$$

   sigue siendo un estimador lineal e insesgado. Esto significa que el valor esperado de $\hat{\beta}_{\text{MCO}}$ coincide con el verdadero parámetro poblacional β.

2. **Pérdida de eficiencia:** Aunque $\hat{\beta}_{\text{MCO}}$ es insesgado, su varianza ya no es mínima. La varianza del estimador se calcula como:

$$\text{Var}(\hat{\beta}_{\text{MCO}}) = (X^\top X)^{-1} X^\top \mathbb{E}[uu^\top] X (X^\top X)^{-1},$$

donde $\mathbb{E}[uu^\top] = \Omega$ es la matriz de varianzas-covarianzas del término de error. En presencia de heteroscedasticidad, esta varianza es distinta de la del modelo con perturbaciones esféricas:

$$\mathrm{Var}(\hat{\beta}_{MCO}) = \sigma^2(X^\top X)^{-1}.$$

Por lo tanto, los estimadores ya no son óptimos, lo que significa que no cumplen con el criterio de mínima varianza entre los estimadores lineales insesgados (BLUE, por sus siglas en inglés).

3. **Inferencia no confiable:** La heteroscedasticidad afecta la validez de los intervalos de confianza y las pruebas de hipótesis, ya que las estimaciones de las varianzas no son correctas. Esto puede llevar a conclusiones incorrectas sobre la significancia de los coeficientes del modelo.

En resumen, aunque los estimadores por MCO permanecen lineales e insesgados en presencia de heteroscedasticidad, pierden eficiencia y afectan la confiabilidad de la inferencia estadística.

4.4. Procedimientos de Detección

4.4.1. Métodos gráficos

En estos métodos se estudiarán los gráficos de los residuos y de dispersión.

- **Gráfico de residuos:** dispersión de los residuos, si en estos podemos observar grupos de observaciones con distinta varianza, podemos afirmar que el modelo es heteroscedástico.
- **Gráficos de dispersión:** Si la variabilidad de los residuos aumenta o disminuye conforme aumenta el valor de la variable dependiente, podemos pensar que la varianza depende de dicha variable, por lo que nos llevaría a pensar que el modelo es heteroscedástico.

4.4.2. Métodos analíticos

En estos se llevará a cabo diversos tests.

Cuando la muestra es pequeña y una variables es la causa de la heteroscedasticidad:

- **Test de Glesjer**

  Pasos

  1. Estimamos el modelo $y = X\beta + u$ y $e = y - \hat{y}$

2. Se ajusta usando el MCO el modelo en el que la varible endógena es $|e^t|$ y la exógena es X_i que es la que creemos que causa la heteroscedasticidad, que será denotada por z_t, es decir:

$$|e^t| = \delta_0 + \delta_1 z_t^h + \nu_t \tag{50}$$

Donde ν_t es ruido blanco[22] y z_t^h es la variable que creemos que causa la heteroscedasticidad[23].

3. Para cada h contrastar $H_0 : \delta_1 = 0$, si se rechaza la hipótesis nula, se puede afirmar que la variable z_t es la causante de la heteroscedasticidad.

- **Test de Goldfeld-Quandt** Los pasos a seguir para la realización de este test son los siguientes[24]:
  - Ordenar la información muestral en función de los valores crecientes de este regresor [25] de *menor a mayor.*
  - Omitimos los m valores de las variables centrales.
  - Ajustar mediante el MCO los subgrupos obtenidos en el paso anterior. La estimación del modelo del primer subgrupo nos permitirá obtener la SCR_1 y de la misma forma, la estimación del modelo del segundo subgrupo nos permitirá obtener la SCR_2.
  - Bajo la ausencia de heteroscedasticidad ($\sigma_1^2 = \sigma_2^2 = sigma_3^2$)

$$F_{exp} = \frac{SCR_2}{SCR_1} > F_{\frac{n-m}{2}-k, \frac{n-m}{2}-k, 1-\alpha} \tag{51}$$

En este caso se rechará la hipótesis de homocedasticidad, por lo que podemos afirmar que hay heteroscedasticidad.

Cuando la muestra es grande y no se sabe que variables causan la heteroscedasticidad:

- **Test de Breusch-Pagan y White**[26]

Pasos

1. Estimos el modelo $y = X\beta + u$ y el vector correspondiente de mínimos cuadrados ordinarios, $e = y - \hat{y}$.
2. Breusch-Pagan:

$$e^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + \nu \tag{52}$$

[22] Ruido blanco en el contexto de las ecuaciones de econometría es un término que se añade a la ecuación para representar el error en la medición de las variables.

[23] donde h puede llegar a tomar valores de $h = \pm 1, \pm 2, \pm\frac{1}{2} \ldots, \pm n$

[24] Para ver ejemplos de estos tests, se puede mirar la diapositivas del tema 5

[25] Se entiende por regresor a la variable que se sospecha que es la causante de la heteroscedasticidad.

[26] La hipótesis nula de todos estos contrastes es que el modelo es homoscedásticos, por lo que si se rechaza la hipótesis nula, se puede afirmar que el modelo es heteroscedástico.

3. White:

$$e^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + \delta_{k+1} x_1^2 + \delta_{k+2} x_2^2 + \ldots + \delta_{k+n} x_k^2 + \nu \quad (53)$$

4. Se contrasta $H_0 : \delta_1 = \delta_2 = \ldots = \delta_k = 0$ (homocedasticidad)
5. Si se rechaza la hipótesis nula, se puede afirmar que el modelo es heteroscedástico.

4.5. Estimación bajo Heterocedasticidad

Supongamos que tenemos el modelo $y = X\beta + u$, donde sabemos que existe heteroscedasticidad, es decir:

$$E[uu^t] = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \to (\sigma^2 = \sigma_1^2) = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2^2}{\sigma_1^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_n^2}{\sigma_1^2} \end{pmatrix} = \quad (54)$$

$$\sigma^2\Omega, \Omega = diag(w_1, w_2, \ldots, w_n) \text{ y } w_i = \frac{\sigma_i^2}{\sigma_1^2}$$

$$\to Var(u_i) = E[u_i^2] = \sigma^2 w_i, \forall i.$$

4.6. Mínimos Cuadrados Generalizados (MCG)

El método de Mínimos Cuadrados Generalizados (MCG) se aplica cuando la matriz de varianzas y covarianzas de los errores, Ω, no es escalar. Dado que $\Omega \succ 0$ y es simétrica, por la **Descomposición de Cholesky**, existe una matriz P (con $P = Q^{-1}$), no singular, tal que:

$$\Omega^{-1} = P^\top P$$

Esto permite transformar el modelo original de tal manera que Ω se convierta en la matriz identidad. De esta forma, se puede aplicar el método de Mínimos Cuadrados Ordinarios (MCO) al modelo transformado. Las expresiones resultantes son:

$$\hat{\beta}_{MCG} = [(X^*)^\top X^*]^{-1}[(X^*)^\top y^*] = [X^\top P^\top P X]^{-1}[X^\top P^\top P y] = [X^\top \Omega^{-1} X]^{-1}[X^\top \Omega^{-1} y]$$

La varianza de los estimadores se calcula como:

$$Var(\hat{\beta}_{MCG}) = \sigma^2[(X^*)^\top X^*]^{-1} = \sigma^2[X^\top P^\top P X]^{-1} = \sigma^2[X^\top \Omega^{-1} X]^{-1}$$

Donde:

- $X^* = PX$ y $y^* = Py$ son las versiones transformadas de las matrices de diseño y el vector de observaciones, respectivamente.

- Ω es la matriz de varianzas y covarianzas de los errores.
- σ^2 es la varianza de los errores.

El uso de estas transformaciones permite trabajar con errores heterocedásticos o autocorrelacionados de forma más eficiente.

4.7. Estimador de Mínimos Cuadrados Ponderados (MCP)

En el caso particular de la heterocedasticidad, la matriz de varianzas y covarianzas de los errores, Ω, se presenta como una matriz diagonal cuyos elementos corresponden a las varianzas de los errores. Su forma es:

$$\Omega = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix}$$

Por lo tanto, su inversa, Ω^{-1}, se calcula fácilmente como:

$$\Omega^{-1} = \begin{pmatrix} \frac{1}{w_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{w_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{w_n} \end{pmatrix}$$

De acuerdo con la descomposición de Cholesky, existe una matriz P tal que:

$$\Omega^{-1} = P^\top P$$

En este caso particular, la matriz P se define como:

$$P = \begin{pmatrix} \frac{1}{\sqrt{w_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{w_n}} \end{pmatrix}$$

Esta descomposición permite transformar el modelo original y aplicar el método de Mínimos Cuadrados Ponderados (MCP). La transformación se realiza multiplicando ambos lados del modelo por la matriz P, de manera que:

$$Py = PX\beta + Pu$$

Donde las nuevas variables transformadas son:

- $y^* = Py$, el vector de observaciones transformado.
- $X^* = PX$, la matriz de diseño transformada.
- $u^* = Pu$, el nuevo término de error.

Esto permite estimar β aplicando el método de Mínimos Cuadrados Ordinarios al modelo transformado.

4.8. Mínimos Cuadrados Ponderados (MCP)

El método de Mínimos Cuadrados Ponderados (MCP) se aplica al transformar el modelo original utilizando la matriz P definida previamente. Esto genera nuevas variables transformadas, que son:

$$y^* = Py, \quad X^* = PX, \quad u^* = Pu$$

Las expresiones de estas transformaciones son:

$$y^* = \begin{pmatrix} \frac{y_1}{\sqrt{w_1}} \\ \vdots \\ \frac{y_n}{\sqrt{w_n}} \end{pmatrix}, \quad X^* = \begin{pmatrix} \frac{1}{\sqrt{w_1}} & \frac{x_{21}}{\sqrt{w_1}} & \cdots & \frac{x_{k1}}{\sqrt{w_1}} \\ \frac{1}{\sqrt{w_2}} & \frac{x_{22}}{\sqrt{w_2}} & \cdots & \frac{x_{k2}}{\sqrt{w_2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{w_n}} & \frac{x_{2n}}{\sqrt{w_n}} & \cdots & \frac{x_{kn}}{\sqrt{w_n}} \end{pmatrix}, \quad u^* = \begin{pmatrix} \frac{u_1}{\sqrt{w_1}} \\ \vdots \\ \frac{u_n}{\sqrt{w_n}} \end{pmatrix}$$

Tras esta transformación, las propiedades del nuevo término de error u^* se definen como:

- Esperanza:
$$\mathbb{E}[u_i^*] = \mathbb{E}\left[\frac{u_i}{\sqrt{w_i}}\right] = \frac{1}{\sqrt{w_i}}\mathbb{E}[u_i] = 0$$

- Varianza:
$$\mathrm{Var}(u_i^*) = \mathrm{Var}\left(\frac{u_i}{\sqrt{w_i}}\right) = \frac{1}{w_i}\mathrm{Var}(u_i) = \frac{1}{w_i}\sigma^2 w_i = \sigma^2$$
Esto muestra que los errores transformados son homocedásticos.

- Covarianza:
$$\mathbb{E}[u_i^* u_j^*] = \mathbb{E}\left[\frac{u_i}{\sqrt{w_i}}\frac{u_j}{\sqrt{w_j}}\right] = \frac{1}{\sqrt{w_i w_j}}\mathbb{E}[u_i u_j] = 0$$
Esto indica que los errores transformados son incorrelacionados.

Dado que los errores transformados son homocedásticos e incorrelacionados, se puede aplicar el método de Mínimos Cuadrados Ordinarios (MCO) al modelo transformado:

$$y^* = X^*\beta + u^*$$

Esto permite estimar β de manera eficiente incluso en presencia de heterocedasticidad en el modelo original.

5 Tema 6: Autocorrelación

5.1. Naturaleza

$$E[u_i, u_j] = \mathrm{Cov}[u_t, u_j] = 0 \quad \forall\, i \neq j \quad \in 1, \dots, n \text{ (Incorrelación)}$$

$$\text{Si} \quad \mathrm{Cov}[u_t, u_{t-k}] \neq 0 \quad \forall k > 0 \rightarrow \text{se dice que hay autocorrelación}$$

Entonces tenemos que $u_t = \rho u_{t-1} + \epsilon_t$, donde ϵ_t es ruido blanco gaussiano.

5.2. Causas de la Autocorrelación

La autocorrelación en un modelo econométrico puede surgir por diversas razones, entre las cuales se destacan:

- **Error en la especificación del modelo:** Esto ocurre cuando se omiten variables exógenas relevantes en la construcción del modelo. Dichas variables pueden ser esenciales para explicar el comportamiento de la variable endógena, y su omisión provoca dependencia serial en los términos de error.
- **Existencia de ciclos y tendencias:** Si la variable endógena presenta ciclos o tendencias que no están adecuadamente explicados por las variables exógenas, estos patrones residuales se reflejan en el término de perturbación, generando autocorrelación.
- **Existencia de relaciones no lineales:** Cuando la verdadera relación entre las variables no es lineal y se utiliza un modelo lineal para estimarla, las discrepancias pueden generar dependencia serial en los errores.
- **Existencia de relaciones dinámicas:** Si el modelo no incorpora adecuadamente los efectos dinámicos entre las variables (como rezagos de la variable dependiente o de las exógenas), el término de perturbación puede presentar autocorrelación.

La detección y corrección de la autocorrelación son esenciales para garantizar la validez de las inferencias en el modelo econométrico.

5.3. Consecuencias y Detección de la Autocorrelación

5.3.1. Consecuencias de la Autocorrelación

La autocorrelación en el término de error tiene importantes implicaciones para los estimadores y la validez del modelo econométrico. Entre sus consecuencias se destacan:

- Los estimadores obtenidos por el método de Mínimos Cuadrados Ordinarios (MCO) son **insesgados** pero **no óptimos**, debido a que no minimizan la varianza entre todos los estimadores lineales insesgados.
- El estimador por Mínimos Cuadrados Generalizados (MCG), definido como:

  $$\hat{\beta}_{MCG} = [X^\top \Omega^{-1} X]^{-1} [X^\top \Omega^{-1} y]$$

  es **insesgado y óptimo**. Su matriz de varianzas-covarianzas viene dada por:

  $$\mathrm{Var}(\hat{\beta}_{MCG}) = \sigma^2 [X^\top \Omega^{-1} X]^{-1}$$

5.3.2. Detección de la Autocorrelación

La autocorrelación se puede detectar de manera gráfica y analítica, según los métodos siguientes:

Gráficamente

- **Gráfico temporal de los residuos:** Si los residuos obtenidos por MCO están incorrelados, deben distribuirse de manera aleatoria alrededor de cero. Sin embargo, si están correlacionados:
  - Observaremos **rachas de residuos** por debajo y por encima de la media (autocorrelación positiva).
  - Observaremos **alternancia en el signo** de los residuos (autocorrelación negativa).
- **Gráficos de dispersión:** El gráfico de dispersión de los residuos e_t frente a algún retardo suyo, e_{t-k} (normalmente se considera $k = 1$), puede revelar:
  - Una **tendencia creciente**, lo que sugiere autocorrelación positiva.
  - Una **tendencia decreciente**, lo que indica autocorrelación negativa.

Analíticamente Existen diversos contrastes estadísticos para la detección de autocorrelación:

- **Contraste de Durbin-Watson:** Evalúa la autocorrelación de primer orden en los residuos.
- **Estadístico H de Durbin:** Diseñado para detectar autocorrelación en modelos con variables dependientes rezagadas.
- **Contraste de Ljung-Box:** Prueba la autocorrelación en retardos múltiples.

5.4. Contraste de Durbin-Watson

El contraste de Durbin-Watson es adecuado para detectar autocorrelación en los términos de error, bajo el supuesto de que el término de perturbación sigue un proceso autorregresivo de orden uno:

$$u_t = \rho u_{t-1} + \epsilon_t$$

donde ϵ_t es ruido blanco[27].

El contraste evalúa si el coeficiente ρ es significativo, mediante las siguientes hipótesis:

[27] En este contexto, el término ”ruido blanco” se refiere a un proceso estocástico con media cero, varianza constante, y ausencia de autocorrelación. Es decir, los errores no tienen ninguna estructura predecible.

$$\begin{cases} H_0 : \rho = 0 & \text{(Incorrelación)} \\ H_1 : \rho > 0 & \text{(Correlación positiva)} \end{cases} \quad \text{o} \quad \begin{cases} H_0 : \rho = 0 & \text{(Incorrelación)} \\ H_1 : \rho < 0 & \text{(Correlación negativa)} \end{cases}$$

El estadístico de contraste de Durbin-Watson se define como:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\sum_{t=2}^{n}(e_t^2 - 2e_t e_{t-1} + e_{t-1}^2)}{\sum_{t=1}^{n} e_t^2}$$

donde e_t son los residuos obtenidos por MCO: $e_t = y_t - \hat{y}_t = y_t - x^\top \hat{\beta}$.

Si el tamaño de muestra es lo suficientemente grande, se puede aproximar que $\sum_{t=2}^{n} e_t^2 \simeq \sum_{t=2}^{n} e_{t-1}^2$, dado que estas sumatorias difieren únicamente en la primera observación muestral. Entonces, el estadístico se puede expresar como:

$$DW = 2\left(\frac{\sum_{t=2}^{n} e_t^2}{\sum_{t=1}^{n} e_t^2} - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right) \simeq 2(1 - \hat{\rho})$$

donde $\hat{\rho}$ es la estimación del parámetro ρ. Según el valor de $\hat{\rho}$, podemos interpretar:

- $DW \simeq 0$ cuando $\hat{\rho} \simeq 1$ (correlación positiva).
- $DW \simeq 2$ cuando $\hat{\rho} \simeq 0$ (residuos incorrelacionados).
- $DW \simeq 4$ cuando $\hat{\rho} \simeq -1$ (correlación negativa).

El estadístico se puede calcular directamente utilizando el paquete `statsmodels` en Python:

```
from statsmodels.stats.stattools import durbin_watson
durbin_watson(mco.resid)
```

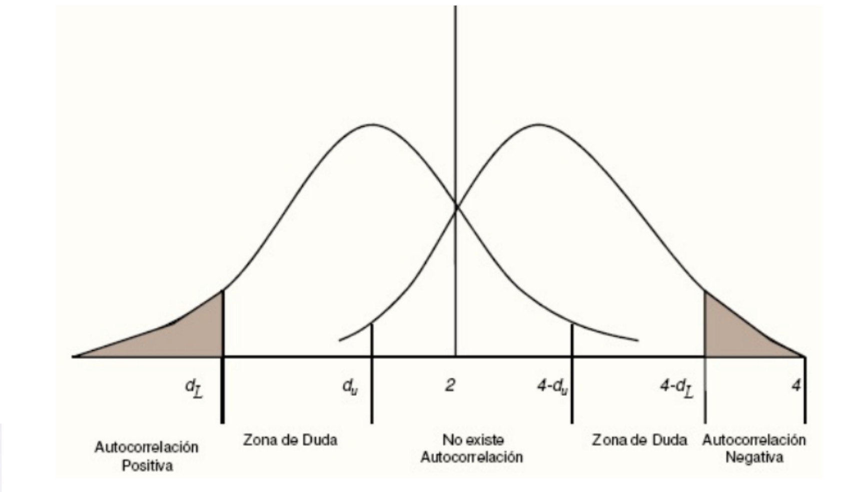

Figura 1: Interpretación del estadístico de Durbin-Watson.

5.4.1. Reglas de Decisión con el Contraste de Durbin-Watson

Las tablas estadísticas para el contraste de Durbin-Watson proporcionan dos cotas: una cota inferior (d_L) y una cota superior (d_U). Las reglas de decisión para evaluar la autocorrelación son las siguientes:

- Si $0 < DW < d_L$, se rechaza H_0 y se concluye que $\rho > 0$, es decir, existe **autocorrelación positiva**.
- Si $d_L < DW < d_U$, el test es **no concluyente**.
- Si $d_U < DW < 4 - d_U$, se mantiene H_0 y se concluye que $\rho = 0$, es decir, los residuos están **incorrelacionados**.
- Si $4 - d_U < DW < 4 - d_L$, el test es **no concluyente**.
- Si $4 - d_L < DW < 4$, se rechaza H_0 y se concluye que $\rho < 0$, es decir, existe **autocorrelación negativa**.

Estas reglas permiten interpretar el valor del estadístico DW en función de las cotas obtenidas para el tamaño de muestra y el número de regresores del modelo.

5.5. Contraste H de Durbin

El contraste de Durbin-Watson no es válido si en la regresión se incluye algún retardo de la variable dependiente entre las variables explicativas. En este caso, se utiliza el contraste h de Durbin, especialmente en modelos como:

$$Y_t = \alpha Y_{t-1} + \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t,$$

donde $u_t = \rho u_{t-1} + \nu_t$.

Las hipótesis a contrastar son:

$$\begin{cases} H_0 : \rho = 0 & \text{(incorrelación)} \\ H_1 : \rho \neq 0 & \text{(correlación)} \end{cases}$$

El estadístico utilizado es:

$$h = \frac{\rho \cdot \sqrt{n}}{\sqrt{1 - n \cdot \widehat{\mathrm{Var}}(\widehat{\alpha})}} \sim N(0, 1),$$

donde se rechaza la hipótesis nula si $|h| > Z_{1-\alpha/2}$.

5.6. Contraste de Ljung-Box

El contraste de Ljung-Box es utilizado para verificar la independencia de los residuos, bajo la hipótesis de que las primeras m autocorrelaciones son iguales a cero. Las hipótesis planteadas son:

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0 & \text{(incorrelación)} \\ H_1 : \exists i \in \{1, 2, \ldots, m\} \text{ tal que } \rho_i \neq 0 & \text{(correlación)} \end{cases}$$

El estadístico de contraste se define como:

$$QLB = n(n+2) \sum_{s=1}^{m} \frac{r(s)^2}{n-s},$$

donde $r(s)$ es el coeficiente de autocorrelación muestral de orden s:

$$r(s) = \frac{\sum_{t=s+1}^{n} e_t e_{t-s}}{\sum_{t=1}^{n} e_t^2}.$$

La hipótesis nula se rechaza si:

$$QLB > \chi^2_{m-1}(1-\alpha).$$

Si las observaciones son independientes, los coeficientes $r(s)$ serán cercanos a cero, y no se rechazará la hipótesis nula.

El contraste puede ser implementado en Python utilizando el siguiente comando:

```
from statsmodels.stats.diagnostic import acorr_ljungbox
acorr_ljungbox(mco.resid, lags=3)
```

5.7. Estimación bajo Autocorrelación

Consideremos el modelo:

$$y = X\beta + u,$$

donde:

- $\mathbb{E}[u] = 0$,
- $\mathbb{E}[u_i u_j] \neq 0$ para $i \neq j$ con $i, j \in \{1, \ldots, n\}$,
- $\mathrm{Var}[u] = \sigma^2 \Omega$.

Estos supuestos implican que no se cumplen las condiciones de los Mínimos Cuadrados Ordinarios (MCO). Por lo tanto, utilizamos el método de Mínimos Cuadrados Generalizados (MCG), para lo cual necesitamos calcular P tal que $\Omega^{-1} = P^\top P$.

5.7.1. Modelo Autorregresivo de Orden 1

Suponemos que u_t sigue un proceso autorregresivo de orden 1, AR(1):

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

donde ε_t es ruido blanco[28], ρ es el coeficiente de autocorrelación con $-1 < \rho < 1$. A partir de esta especificación, se derivan las siguientes relaciones:

$$\begin{aligned}
\mathbb{E}[u_t u_{t-1}] &= \rho\sigma^2, \\
\mathbb{E}[u_t u_{t-2}] &= \rho^2\sigma^2, \\
\mathbb{E}[u_t u_{t-3}] &= \rho^3\sigma^2, \\
\mathbb{E}[u_t u_{t-k}] &= \rho^k\sigma^2.
\end{aligned}$$

La matriz de covarianza de u, Ω, toma la forma:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}.$$

Su inversa es:

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & -\rho \\ 0 & 0 & 0 & -\rho & 1 \end{pmatrix}.$$

5.7.2. Transformación del Modelo

Partiendo del modelo original:

$$\begin{aligned}
y_t &= \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t, \\
y_{t-1} &= \beta_1 + \beta_2 X_{2,t-1} + \cdots + \beta_k X_{k,t-1} + u_{t-1},
\end{aligned}$$

obtenemos la transformación:

$$y_t - \rho y_{t-1} = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2,t-1}) + \cdots + \beta_k(X_{kt} - \rho X_{k,t-1}) + (u_t - \rho u_{t-1}).$$

Definiendo $y_t^* = y_t - \rho y_{t-1}$ y $X_{it}^* = X_{it} - \rho X_{i,t-1}$, el modelo queda:

$$y_t^* = \beta_1 + \beta_2 X_{2t}^* + \cdots + \beta_k X_{kt}^* + \varepsilon_t,$$

donde ε_t cumple los supuestos para aplicar MCO.

[28]El término ruido blanco se refiere a un proceso estocástico con media cero, varianza constante y ausencia de correlación serial.

5.7.3. Primer Observación: Método Prais-Winsten

Para no perder la primera observación, Prais-Winsten propone:

$$y_1^* = \sqrt{1-\rho^2}y_1, \quad X_{i1}^* = \sqrt{1-\rho^2}X_{i1}.$$

5.8. Modificación Prais-Winsten

El método de Prais-Winsten permite ajustar un modelo con autocorrelación mediante un procedimiento iterativo que combina transformaciones y reestimaciones. El proceso se describe a continuación:

1. Estimación inicial por Mínimos Cuadrados Ordinarios (MCO), obteniendo los residuos e_t.

2. Estimación inicial de ρ asumiendo el modelo $u_t = \rho u_{t-1} + \varepsilon_t$, utilizando los residuos e_t:
$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2}.$$

3. Transformación del modelo a:
$$y_t^* = y_t - \hat{\rho}y_{t-1}, \quad X_{it}^* = X_{it} - \hat{\rho}X_{i,t-1}.$$
Se estima nuevamente por MCO el modelo transformado:
$$y^* = X^*\beta + u^*,$$
obteniendo nuevos residuos e_t^*.

4. Estimación actualizada de ρ a partir de la regresión $e_t^* = \hat{\rho}e_{t-1}^* + \varepsilon_t^*$:
$$\hat{\hat{\rho}} = \frac{\sum_{t=2}^n e_t^* e_{t-1}^*}{\sum_{t=2}^n (e_t^*)^2}.$$

5. Repetición del proceso: el procedimiento se itera hasta que la estimación de ρ se estabilice, es decir, la diferencia entre iteraciones consecutivas de $\hat{\rho}$ sea menor a 10^{-3}.

5.8.1. Proceso Iterativo Cochrane-Orcutt

El proceso iterativo de Cochrane-Orcutt es similar al de Prais-Winsten, pero omite la primera observación en las transformaciones. Esto simplifica las estimaciones pero puede reducir la eficiencia en muestras pequeñas.

5.8.2. Implementación en Python

La implementación del procedimiento iterativo de Prais-Winsten puede realizarse utilizando la clase `GLSAR` de la biblioteca `statsmodels`:

```
import statsmodels.api as sm

# Ajuste del modelo con iteraciones
model = sm.GLSAR(y, sm.add_constant(x), rho=rho)
results = model.iterative_fit(maxiter=100)
```

6 Recursos de Prácticas Relacionados con Python

Para ello pincha aquí.

Recursos de Prácticas en Econometría con Python

Nota: A continuación, se van a ir explicando lo que se considere más relevante de las prácticas de Econometría con Python de cara al examen.

6.1. Introducción

```
import pandas as pd #librería para manejo de datos


data= pd.read_csv("https://rtgodwin.com/data/houseprice.csv") #Lee base de datos de web...

data
```

Listing 1: Importación de librerías

```
    import statsmodels.api as sm
    # Modelo Regresión:  modeldata=stock.values
    X=data[["Age", "Rooms"]]
    y=data["Price"]


    results = sm.OLS(y, sm.add_constant(X)).fit()

    print(results.summary())
```

Listing 2: Modelo de Regresión Lineal

6.1.1. Salida

```
    OLS Regression Results
==============================================================================
Dep. Variable:                  Price   R-squared:                       0.303
Model:                            OLS   Adj. R-squared:                  0.303
Method:                 Least Squares   F-statistic:                     375.5
Date:                Sat, 04 Jan 2025   Prob (F-statistic):         4.04e-136
Time:                        17:03:06   Log-Likelihood:                -22006.
No. Observations:                1728   AIC:                         4.402e+04
Df Residuals:                    1725   BIC:                         4.404e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       7.036e+04   6767.987     10.395      0.000    5.71e+04    8.36e+04
Age         -492.3251     67.957     -7.245      0.000    -625.613    -359.037
Rooms       2.206e+04    856.930     25.746      0.000    2.04e+04    2.37e+04
==============================================================================
Omnibus:                      569.117   Durbin-Watson:                   1.572
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2391.072
Skew:                           1.537   Prob(JB):                         0.00
Kurtosis:                       7.874   Cond. No.                         140.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

6.1.2. Interpretación de Resultados de la Regresión OLS

En este análisis, se ajustó un modelo de regresión lineal ordinaria (OLS) para predecir el precio (*Price*) como variable dependiente, utilizando la edad (*Age*) y el número de habitaciones (*Rooms*) como variables explicativas. A continuación, se interpretan los principales resultados:

- **R-cuadrado:** El valor del R-cuadrado es 0.303, lo que indica que el modelo explica aproximadamente el 30.3 % de la variabilidad observada en el precio. Aunque no es un ajuste muy alto, sugiere que las variables incluidas tienen cierta capacidad predictiva.

- **Coeficientes:**
  - El intercepto (*const*) tiene un coeficiente estimado de $7{,}036 \times 10^4$, lo que implica que, en promedio, cuando las variables explicativas son cero, el precio base es de 70,360 unidades monetarias.
  - Para la variable *Age*, el coeficiente es de $-492{,}33$, lo que sugiere que, manteniendo las demás variables constantes, por cada año adicional de anti-

güedad, el precio disminuye en promedio 492.33 unidades monetarias. Este efecto es estadísticamente significativo ($p < 0{,}001$).

- Para la variable *Rooms*, el coeficiente es de $2{,}206 \times 10^4$, lo que indica que, manteniendo las demás variables constantes, cada habitación adicional incrementa el precio en promedio 22,060 unidades monetarias. Este efecto también es estadísticamente significativo ($p < 0{,}001$).

- **Estadísticos de bondad de ajuste y pruebas de diagnóstico:**
  - El valor del estadístico de Durbin-Watson es 1.572, lo que sugiere cierta autocorrelación positiva en los residuos.
  - Las pruebas de normalidad (Omnibus y Jarque-Bera) indican que los residuos no siguen una distribución normal ($p < 0{,}001$). Esto podría afectar la validez de algunas inferencias estadísticas.

- **Notas:** Los errores estándar asumen que la matriz de covarianzas de los errores está correctamente especificada. La condición de número del modelo es 140, lo que no sugiere problemas graves de multicolinealidad.

En resumen, el modelo muestra que la antigüedad afecta negativamente el precio, mientras que el número de habitaciones tiene un impacto positivo y significativo. Sin embargo, la baja capacidad explicativa del modelo y las desviaciones de normalidad en los residuos sugieren que podrían explorarse variables adicionales o transformaciones para mejorar el ajuste.

```
# Definimos el número de muestras a generar
n = 100

# Importamos el paquete numpy para trabajar con arrays y funciones matemáticas
import numpy as np

# Generamos un array de n valores aleatorios con una distribución normal
# Media (loc) = 0, Desviación estándar (scale) = 10
X = np.random.normal(0, 10, n)

# Generamos otra variable Y basada en X, sumándole un ruido aleatorio
# El ruido se extrae de una distribución normal con media 0 y desviación estándar 1
Y = X + np.random.normal(0, 1, n)
```

Listing 3: Generación de datos sintéticos con numpy

```
    # Importamos el módulo de visualización matplotlib.pylab
    import matplotlib.pylab as plt

    # Graficamos los puntos (X, Y) como un diagrama de dispersión con tamaño de marcador 1
    plt.scatter(X, Y, s=1)

    # Mostramos el gráfico
    plt.show()

    # Ajustamos un modelo de regresión OLS (Mínimos Cuadrados Ordinarios) a los datos
    results = sm.OLS(Y, sm.add_constant(X)).fit()

    # Mostramos un resumen detallado de los resultados del modelo
    results.summary()

    # Obtenemos los parámetros estimados: el intercepto (cte) y la pendiente (beta)
    cte = results.params[0]
    beta = results.params[1]

    # Graficamos la recta de regresión estimada
    plt.plot([-20, 20], [cte + beta * (-20), cte + beta * 20], color='r')

    # Sobreponemos el gráfico de dispersión original
    plt.scatter(X, Y, s=1)

    # Mostramos el gráfico con la línea de regresión
    plt.show()

    # Importamos herramientas de análisis de estadísticas en statsmodels
    import statsmodels.stats.api as sms
    from statsmodels.compat import lzip
```

```python

        # Cargamos un conjunto de datos de ejemplo sobre crímenes estatales
            desde statsmodels
        data = sm.datasets.statecrime.load_pandas()

        # Accedemos a los datos (endógenos y exógenos)
        data.data
        data.exog

        # Definimos y ajustamos un modelo OLS a los datos
        modeloMCO = sm.OLS(data.endog, sm.add_constant(data.exog))
        resultado = modeloMCO.fit()

        # Mostramos un resumen de los resultados del modelo
        print(resultado.summary())
```

Listing 4: Ajuste y visualización de un modelo OLS con statsmodels

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.989
Model:	OLS	**Adj. R-squared:**	0.988
Method:	Least Squares	**F-statistic:**	8459.
Date:	Sat, 04 Jan 2025	**Prob (F-statistic):**	6.21e-
Time:	17:13:10	**Log-Likelihood:**	-145.1
No. Observations:	100	**AIC:**	294.3
Df Residuals:	98	**BIC:**	299.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-0.0007	0.105	-0.007	0.994	-0.209	0.208
x1	0.9923	0.011	91.975	0.000	0.971	1.014

Omnibus:	5.961	**Durbin-Watson:**	2.203
Prob(Omnibus):	0.051	**Jarque-Bera (JB):**	5.575
Skew:	0.571	**Prob(JB):**	0.0616
Kurtosis:	3.178	**Cond. No.**	9.78

Figura 2: Modelo OLS.

6.1.3. Interpretación de Resultados de la Regresión OLS

Se ajustó un modelo de regresión lineal ordinaria (*Ordinary Least Squares*, OLS) para explicar la variable dependiente y en función de una única variable independiente x_1. A continuación, se presenta el análisis detallado de los resultados:

- **Bondad de ajuste:**

- **R-cuadrado:** El valor de R^2 es 0.989, lo que indica que el modelo explica el 98.9 % de la variabilidad observada en la variable y.
- **R-cuadrado ajustado:** El valor ajustado es 0.988, lo que refleja una ligera penalización por el número de parámetros en el modelo.

- **Coeficientes estimados:**
  - El intercepto (*const*) tiene un valor estimado de $-0{,}0007$, pero no es estadísticamente significativo ($p = 0{,}994$). Esto implica que el modelo no detecta un impacto relevante de la constante en el valor de y.
  - El coeficiente de la variable x_1 es 0.9923, lo que significa que, en promedio, por cada unidad adicional en x_1, y aumenta en 0.9923 unidades. Este coeficiente es altamente significativo ($p < 0{,}001$), con un intervalo de confianza al 95 % que va de 0.971 a 1.014.

- **Estadísticas del modelo:**
  - **Estadístico F:** Con un valor de 8459 y una probabilidad asociada (p) de $6{,}21 \times 10^{-97}$, el modelo tiene una capacidad predictiva altamente significativa.
  - **Log-verosimilitud:** El valor de -145.16 es consistente con el ajuste observado.
  - **AIC y BIC:** Los valores del Criterio de Información de Akaike (AIC = 294.3) y el Criterio de Información Bayesiano (BIC = 299.5) son útiles para comparar este modelo con otros posibles modelos.

- **Diagnósticos de los residuos:**
  - **Pruebas de normalidad:** La prueba de Omnibus (valor $p = 0{,}051$) y la prueba de Jarque-Bera (valor $p = 0{,}0616$) no rechazan la hipótesis de que los residuos siguen una distribución normal al nivel de significancia del 5 %.
  - **Durbin-Watson:** El valor de 2.203 sugiere que no hay autocorrelación significativa en los residuos.
  - **Condición de número:** Con un valor de 9.78, el modelo no muestra problemas graves de multicolinealidad.

En conclusión, el modelo muestra un ajuste excelente, con un coeficiente significativo para x_1 que predice el 98.9 % de la variabilidad en y. Los diagnósticos no sugieren problemas relevantes en los residuos ni en la especificación del modelo.

6.1.4. Tipo de Covarianza: nonrobust

En los resultados del modelo de regresión, la línea `Covariance Type: nonrobust` indica el tipo de matriz de covarianza utilizado para calcular los errores estándar de los coeficientes estimados. A continuación, se detalla su significado:

- **Nonrobust:**

- Este tipo de matriz de covarianza supone que los errores (*residuos*) del modelo son **homocedásticos** (tienen varianza constante) y **no están correlacionados**.
- Bajo esta suposición, los errores estándar, los valores t, y los valores p son válidos siempre que se cumplan las condiciones clásicas del modelo de regresión lineal (*Classical Linear Regression Model, CLRM*).

- **Alternativas robustas:** Si las condiciones del CLRM no se cumplen (por ejemplo, si hay heterocedasticidad o autocorrelación en los errores), es posible usar matrices de covarianza ajustadas:
  - **HC0, HC1, HC2, HC3:** Correcciones robustas a la heterocedasticidad, que permiten obtener errores estándar válidos incluso si los residuos tienen varianza no constante.
  - **HAC (Heteroskedasticity and Autocorrelation Consistent):** Útil cuando los errores presentan tanto heterocedasticidad como autocorrelación.

Uso práctico: En *statsmodels*, se puede especificar un tipo de covarianza robusta al ajustar el modelo, utilizando el argumento `cov_type` en el método `fit`. Por ejemplo:

```
results = sm.OLS(Y, sm.add_constant(X)).fit(cov_type='HC3')
```

Listing 5: Ajuste de un modelo OLS con covarianza robusta

En este caso, se ajustan los errores estándar para ser robustos a la heterocedasticidad.

Conclusión: El uso de `nonrobust` asume que los residuos cumplen con las condiciones clásicas de homocedasticidad y no autocorrelación. Si estas suposiciones no se verifican, es recomendable optar por una matriz de covarianza robusta para garantizar la validez de las inferencias estadísticas.

6.2. Autocorrelación

```
    # Seleccionamos las columnas de interés en el DataFrame y eliminamos
      filas con valores nulos
    datos = datos[['pm2.5', 'TEMP', 'PRES', 'Iws']].dropna()
```

Listing 6: Selección y limpieza de datos en un DataFrame

```
    # Importamos la función durbin_watson desde statsmodels
    from statsmodels.stats.stattools import durbin_watson

    # Calculamos la estadística de Durbin-Watson sobre los residuos del
      modelo ajustado
    dw = durbin_watson(mco.resid)

    # Imprimimos el resultado de la estadística de Durbin-Watson
    print("Durbin-Watson statistic:", dw)
```

Listing 7: Cálculo de la estadística de Durbin-Watson para residuos

```python
# Importamos las librerías necesarias
import numpy as np
import statsmodels.api as sm
from statsmodels.stats.stattools import durbin_watson

# Datos simulados: serie autoregresiva
np.random.seed(0)  # Fijamos la semilla para reproducibilidad
n = 100  # Número de observaciones
y_original = datos["pm2.5"]  # Variable original de interés
y_retardada = np.roll(y_original, 1)  # Variable desplazada una posición hacia atrás
y_retardada[0] = 0  # Establecemos el primer valor como 0 para evitar problemas

# Ajustamos un modelo OLS con la variable retardada
modelo = sm.OLS(y_original, sm.add_constant(y_retardada)).fit()

# Extraemos los parámetros del modelo
beta = modelo.params["x1"]  # Coeficiente estimado para la variable retardada
var_beta = modelo.bse["x1"] ** 2  # Varianza estimada del coeficiente

# Calculamos la estadística h de Durbin
h = (1 - dw / 2) * np.sqrt(n / (1 - var_beta))

# Mostramos el valor de h
print("h-Durbin:", h)

# Verificamos la significancia de h
if np.abs(h) > 1.96:  # Intervalo de confianza del 95%
    print("Autocorrelación detectada.")
else:
    print("Autocorrelación no significativa.")
```

Listing 8: Detección de autocorrelación con la estadística h de Durbin

```python
# Importamos la función acorr_ljungbox
from statsmodels.stats.diagnostic import acorr_ljungbox

# Aplicamos la prueba de Ljung-Box a los residuos del modelo
# Evaluamos hasta 5 retardos (lags)
ljungbox_result = acorr_ljungbox(mco.resid, lags=5)

# Mostramos el resultado
print(ljungbox_result)
```

Listing 9: Prueba de Ljung-Box para detectar autocorrelación

```python
# Calcula el coeficiente rho basado en el estadístico de Durbin-Watson (DW)
rho = 1 - dw / 2  # dw = 2(1-rho) => rho = 1 - DW/2
print(rho)  # Imprime el valor calculado de rho

# Ajusta un modelo de regresión generalizada de mínimos cuadrados con autocorrelación (GLSAR)
```

```python
    mco_autocorr = sm.GLSAR(y, sm.add_constant(X), rho=rho)

    # Realiza un ajuste iterativo del modelo hasta un máximo de 100
        iteraciones o hasta alcanzar la tolerancia
    res = mco_autocorr.iterative_fit(maxiter=100, rtol=10**(-10))

    # Muestra el número de iteraciones realizadas y si el modelo ha
        convergido
    print('Iteraciones = %d -->  Converge: %s' % (res.iter, res.converged))

    # Imprime el valor final de rho utilizado en el modelo
    print('Rho =  ', mco_autocorr.rho)

    # Imprime un resumen detallado de los resultados del modelo ajustado
    print(res.summary())
```

Listing 10: Implementación de autocorrelación usando GLSAR en Python.

```python
    # Define una función lambda para calcular el Coeficiente de Variación (
        CV)
    cv = lambda x: np.std(x, ddof=1) / np.mean(x) * 100

    # Aplica la función CV a cada columna del DataFrame 'datos'
    datos.apply(cv)

    # Importa el módulo para realizar pruebas estadísticas de Statsmodels
    import statsmodels.stats.diagnostic as diag

    # Realiza la prueba de Kolmogorov-Smirnov para verificar la normalidad
        de los residuos del modelo 'mco'
    diag.kstest_normal(mco.resid)
```

Listing 11: Cálculo del Coeficiente de Variación (CV) y prueba de normalidad en los residuos.

6.3. Práctica 1

```python
    # Calcula el conteo de valores únicos en la columna "Fuel.Type" del
        DataFrame 'datos'
    datos["Fuel.Type"].value_counts()
```

Listing 12: Conteo de frecuencias para la columna "Fuel.Type"del DataFrame.

6.4. Test Heterocedasticidad

```python
    # Importa el módulo de diagnóstico estadístico de Statsmodels
    import statsmodels.stats.diagnostic as sm_diagnostic

    # GOLDFELD-QUANDT (Prueba para heterocedasticidad en muestras pequeñas)
    # Caso: Hipótesis alternativa de varianza creciente
    GQ = sm_diagnostic.het_goldfeldquandt(y, sm.add_constant(datos["Age"]),
         alternative="increasing")
```

```
    print("Goldfeld Quandt (increasing): ", GQ)

    # Caso: Hipótesis alternativa de varianza decreciente
    GQ = sm_diagnostic.het_goldfeldquandt(y, sm.add_constant(datos["Age"]),
        alternative="decreasing")
    print("Goldfeld Quandt (decreasing): ", GQ)

    # Caso: Hipótesis alternativa bilateral
    GQ = sm_diagnostic.het_goldfeldquandt(y, sm.add_constant(datos["Age"]),
        alternative="two-sided")
    print("Goldfeld Quandt (two-sided): ", GQ)

    # BREUSH-PAGAN (Prueba para detectar heterocedasticidad)
    BP = sm_diagnostic.het_breuschpagan(mco.resid, mco.model.exog)
    print("Breush Pagan: ", BP)

    # WHITE (Prueba para detectar heterocedasticidad generalizada)
    W = sm_diagnostic.het_white(mco.resid, mco.model.exog)
    print("White: ", W)

    # GLEJSER (Prueba para detectar heterocedasticidad con transformación
       de datos)
    import numpy as np
    z = np.array(datos["Land.Value"].values, dtype=float)
    for h in [-2, -1, -0.5, 0.5, 1, 2]:
        # Ajusta el modelo |e| = delta_0 + delta_1 z^h + eps
        mcoaux = sm.OLS(abs(mco.resid), sm.add_constant(z**h)).fit()
        pval = mcoaux.pvalues["x1"]
        print("h: ", h, "-> pvalt: ", pval, "R2: ", mcoaux.rsquared)

    # MÍNIMOS CUADRADOS PONDERADOS (Weighted Least Squares, WLS)
    # Ajusta un modelo ponderando por la raíz cuadrada inversa de z
    mcp = sm.WLS(y, sm.add_constant(X), weights=1. / np.sqrt(z)).fit()
    print(mcp.summary())
```

Listing 13: Pruebas de heterocedasticidad y ajuste de mínimos cuadrados ponderados.

6.5. Autocorrelación

```
    # IMPORTANTE: El Número de Condición evalúa la colinealidad en el diseñ
       o del modelo.
    # Si es alto (típicamente >30 en su forma no raíz cuadrada), indica
       posible multicolinealidad.
    import numpy as np

    # Calcula la raíz cuadrada del Número de Condición del modelo 'mco'
    CN = np.sqrt(mco.condition_number)  # Número de Condición
    print(CN)  # Imprime el Número de Condición para interpretar la
       estabilidad numérica

    # FACTOR DE INFLACIÓN DE LA VARIANZA (VIF)
    # El VIF evalúa la multicolinealidad para cada variable independiente.
    # Valores altos (típicamente >10) indican problemas de
       multicolinealidad.
```

```
    import statsmodels.stats.outliers_influence as oi

    # Calcula el VIF para cada variable en el conjunto de predictores 'X'
    vifs = [oi.variance_inflation_factor(X.values, i) for i in range(X.
        shape[1])]
    vifs  # Devuelve los VIFs calculados para interpretar colinealidad

    # MATRIZ DE CORRELACIONES
    # Evalúa la relación lineal entre las variables independientes.
    # Correlaciones altas (>0.8 o <-0.8) pueden sugerir multicolinealidad.
    corr_matrix = np.corrcoef(X.T)  # Calcula la matriz de correlaciones de
         los predictores
    print(corr_matrix)  # Imprime la matriz para inspeccionar las
        correlaciones

    # VISUALIZACIÓN DE LA MATRIZ DE CORRELACIONES
    # Utiliza gráficos para identificar patrones visuales de
        multicolinealidad.
    import statsmodels.graphics.api as smg
    import matplotlib.pylab as plt

    # Genera un gráfico de correlaciones con nombres de las variables
    smg.plot_corr(corr_matrix, xnames=['Lot.Size', 'Age', 'Land.Value', '
        Bedrooms'])
    plt.show()
```

Listing 14: Cálculo del Número de Condición, Factor de Inflación de la Varianza (VIF) y Matriz de Correlaciones.

6.6. Normalidad

```
        # Prueba de normalidad de Kolmogorov-Smirnov para los residuos
        # Diagnostica si los residuos del modelo siguen una distribución
            normal.
        # Una p-valor bajo (<0.05) indica que los residuos no son normales.
        import statsmodels.stats.api as sms
        sms.diagnostic.kstest_normal(mco.resid)

        # Suma de cuadrados residual (SSR)
        # Representa la variación no explicada por el modelo.
        # Un valor bajo de SSR indica que el modelo ajusta bien los datos.
        mco.ssr

        # Estimación de la varianza residual (sigma_gorro)
        # Calcula la varianza de los errores basada en SSR, el número de
            observaciones (nobs),
        # y el número de parámetros del modelo (params).
        sigmagorro = mco.ssr / (mco.nobs - len(mco.params) - 1)
        sigmagorro

        # Suma de cuadrados explicada (ESS)
        # Representa la variación explicada por el modelo. Un valor alto
            indica que el modelo
        # explica gran parte de la variabilidad de los datos.
```

```
        mco.ess

        # Predicción del modelo para nuevos datos
        # Estima el valor de la variable dependiente para un conjunto específico de predictores.
        # Ejemplo: [1, 0.009, 0.42, 50000, 90.6, 2] representa los valores de las variables independientes.
        mco.predict([1, 0.009, 0.42, 50000, 90.6, 2])
```

Listing 15: Pruebas estadísticas, estimación de varianza residual, suma de cuadrados y predicción del modelo.

6.7. Anotaciones Extras

Recursos de Prácticas en Econometría con Python

Nota: Los que no se hayan mencionado es porque se sobreentiende el código proporcionado anteriormente (Enlace Github).