Apuntes Econometría Teoria_T1_al_T6 - EL CORÁN

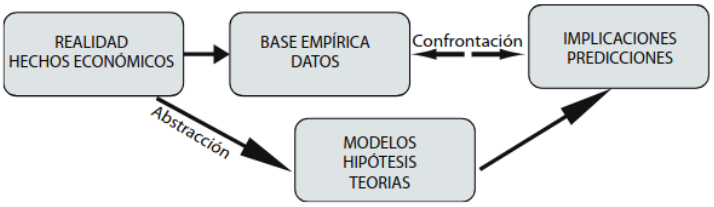
(Ejercicios propuestos y resueltos en clase en otro archivo)

Tema 1: Introducción a la econometría

¿Qué es el método econométrico?

El método econométrico está construido sobre un proceso de razonamiento lógico utilizando una información inicial y un conocimiento empírico. Compuesto y estructurado por:

Realidad: Estudio de los hechos económicos que surgen de la misma sociedad.
Datos: Recopilación de los datos obtenidos del conocimiento que el investigador tiene sobre la propia realidad.
Modelo: A partir de las bases de datos, el investigador formula hipótesis dando lugar a modelos, leyes y teorías.
Predicciones: Los modelos permiten al investigador predecir determinados comportamientos.



¿Qué es la econometría?

La Econometría es la ciencia que utiliza la Teoría Económica, las Matemáticas y la Inferencia Estadística de forma conjunta utilizando como herramienta la Informática.

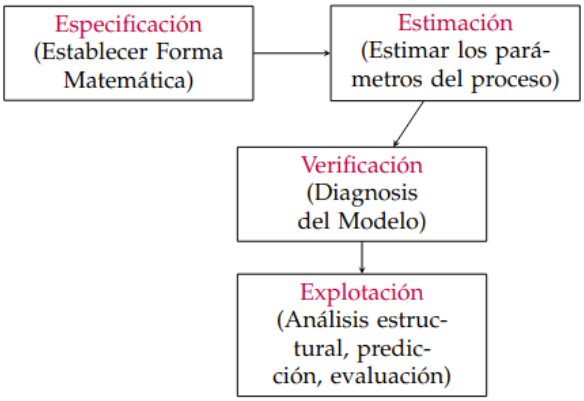
¿Qué son los Modelos Económicos y los Modelos Econométricos?

Los modelos económicos son representaciones algebraicas simplificadas de la realidad económica, usando ecuaciones que relacionan diferentes variables.

Los modelos econométricos son una aplicación práctica de estos modelos (económicos), en los que se describe matemáticamente la relación entre variables explicadas y explicativas, incluyendo un término aleatorio o perturbación.

Fases del Método Econométrico

1. **Especificación:** Se establece matemáticamente la relación entre las variables.
2. **Estimación:** Se calculan los parámetros del modelo. (Estimar los parámetros)
3. **Verificación:** Se comprueba el modelo para detectar errores y, si es necesario, reformularlo (es decir, si no cumple la verificación, volveremos al paso 1).
4. **Explotación:** El modelo, una vez validado, se usa para análisis, predicción y evaluación de políticas económicas.



(Este esquema se corresponde a las fases del método econométrico)

VARIABLES

En un modelo econométrico podemos diferenciar las variables en dos grandes grupos:

No observables: Son aquellas de las que NO disponemos de los datos muestrales. Y por ello, tienen algo llamado **perturbación (u_t).**
Observables: Son aquellas de las que disponemos los datos muestrales:

Las **variables endógenas corrientes** son las **variables explicadas** o regresandos, que son determinadas/explicadas por el modelo econométrico. Las **variables predeterminadas o explicativas,** también llamadas regresores, describen el comportamiento de las variables endógenas/explicadas. Estas se subdividen en:

Corriente: Momento actual // Retardada: Momento anterior

- **Endógenas retardadas**.
- **Exógenas corrientes**.
- **Exógenas retardadas**.

$$c_t = \alpha_0 + \alpha_1(1-\tau)y_t + \alpha_2 r_t + u_t,$$
$$i_t = \beta_0 + \beta_1(y_{t-1} - y_{t-2}) + \beta_2 r_{t-1} + v_t,$$
$$y_t \equiv c_t + i_t + g_t,$$

explicadas explicativa explicativa retardada

Donde:

C_t, i_t, Y_t: Son variables endógenas corrientes. (Las incógnitas, aquellas a la izquierda) // Con su variación con el subíndice-1 que en dicho caso serían endógenas retardadas.

r_t, g_t: Son variables exógenas corrientes. // Con su variación con el subíndice-1 que en dicho caso serían exógenas retardadas.

Todas aquellas con subíndices-1 son variables retardadas.

Además tenemos que saber que: **Los parámetros** (fijos y desconocidos) son los betas y las alphas del ejemplo anterior.

Las **relaciones** entre los distintos tipos de variables vienen especificadas mediante un sistema de ecuaciones.

FORMAS DE EXPRESAR UN MODELO ECONOMÉTRICO

Según el **NÚMERO DE ECUACIONES:**

- **Uniecuacionales:** Sistema descrito por una ecuación
  1. Uniecuacional simple $Y_t = \beta_0 + \beta_1 X_t + u_t.$
  2. Uniecuacional múltiple
     $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t.$
- **Multiecuacionales:** Formado por un sistema de ecuaciones donde puede haber o no relación entre las variables.

$$Y_{1t} = \gamma_{12} Y_{2t} + \beta_{11} X_{1t} + u_{1t}$$
$$Y_{2t} = \gamma_{21} Y_{1t} + \beta_{21} X_{1t} + \beta_{22} X_{2t} + u_{1t}$$

$$Y_{1t} = \beta_0 + \beta_1 X_{1t} + u_t$$
$$Y_{2t} = \alpha_0 + \alpha_1 X_{2t} + v_t$$

Según la **FORMA FUNCIONAL:**

Modelos lineales $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t.$

Modelos no lineales intrínsecamente linealizables
$Y_t = \beta_0 + \beta_1 e^{X_{1t}} + \beta_2 X_{2t} X_{3t} + u_t.$

Modelos no lineales intrínsecamente no linealizables $Y_t = \beta_0 + X_t^{\beta_1} + u_t.$

Según las **CARACTERÍSTICAS DE LA PERTURBACIÓN:**

- **Modelos clásicos:** Cuando la perturbación es "ruido blanco" es decir:
  1. $E[u_t] = 0 \; \forall t$
  2. $\text{var}[u_t] = \sigma^2 \; \forall t$
  3. $\text{cov}[u_t u_s] = 0 \; \forall t \neq s.$

  Cumplen dichas hipótesis.
- **Modelos generalizados:** Cuando las perturbaciones incumplen alguna de las DOS últimas hipótesis descritas en los modelos clásicos. Es decir, habrá un problema de **heterocedasticidad** (varianza no cte.) o **autocorrelación** (covarianza distinta de 0 para t distinto de s). Siendo t y s dos momentos diferentes.

Autocorrelación

NATURALEZA DE LA INFORMACIÓN

Las bases de datos en econometría se dividen en tres tipos:

1. **Sección Cruzada / Cortes transversal**: Datos de múltiples individuos observados en el mismo instante de tiempo. Ejemplo: Calificaciones de alumnos en un examen específico.
2. **Series Temporales**: Datos de un solo individuo observados en distintos momentos del tiempo. Ejemplo: Cotizaciones diarias de las acciones de un banco.
3. **Datos de Panel**: Datos de varios individuos observados en diferentes momentos del tiempo. Ejemplo: Cotizaciones diarias de todas las empresas de un índice bursátil.

EJEMPLOS:

Ejemplo de corte transversal:

N. de obs.	Calif. media/dist(Y)	Ratio estud-maest(X_1)	G/estud.(X_2)	% estud. ingls(X_3)
1	690.8	17.89	6.385	0.8
2	661.2	21.52	5.099	4.6
3	643.6	18.70	5.502	30.8
...	...	...	...	...
419	672.2	20.20	4.776	3.0
420	765.8	19.04	5.993	5.8

Ejemplo de series temporales:

Ao	Consumo (Y)	Renta (X)
1982	3081.5	4620.3
1983	3240.6	4803.7
1984	3407.6	5140.1
1985	3566.5	5323.5
1986	3708.7	5487.7
1987	3822.3	5649.5
1988	3972.7	5865.2
1989	4064.6	6062.0
1990	4132.2	6136.3
1991	4105.8	6079.4
1992	4219.8	6244.4
1993	4343.6	6389.6
1994	4486.0	6610.7
1995	4595.3	6742.1
1996	4714.1	6928.4

Ejemplo de datos de panel:

N. de obs.	Estado	Ao	Vta de cig (Y)	P. medio/paquete(X_1)	Imptos tot.(X_2)
1	Alabama	1985	116.5	1.022	0.333
2	Arkansas	1985	128.5	1.015	0.370
...	...	...	...	...	...
48	Wyoming	1985	129.4	0.935	0.240
49	Alabama	1986	117.2	1.080	0.334
...	...	...	...	...	...
96	Wyoming	1986	127.8	1.007	0.240
97	Alabama	1987	115.8	1.135	0.335
...	...	...	...	...	...
528	Wyoming	1995	127.8	1.007	0.240

Tema 2: Modelo Lineal (I)

Recordando del tema anterior:

Modelo Lineal Simple:

$$y_i \sim a + bx_i$$

Modelo Lineal Múltiple:
Permite estudiar el comportamiento de una variable endógena Y, a partir de k variables independientes $(X_2, X_3, \ldots, X_k)$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

Notación MLP

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

- el subíndice $i = 1, \ldots, n$, indica la i-ésima de las n observaciones de la muestra.
- El subíndice i será empleado cuando trabajemos con observaciones de sección cruzada o de corte transversal, en el caso de trabajar con información de caracter temporal se empleará el subíndice t.
- Y_i es la i-ésima observación de la variable dependiente.
- $X_{2i}, \ldots, X_{ki}$ son las i-ésimas observaciones de las k variables independientes.
- u_i es la perturbación de la i-ésima observación, encargada de recoger aquella parte de la variable Y_i que no es explicada por $X_{1i}, X_{2i}, \ldots, X_{ki}$.
- $\beta_1, \ldots, \beta_k$ son los parámetros del modelo. Además, β_1 es el denominado término constante.

Así que realmente estaremos trabajando con matrices:

$$\vec{y} = X\vec{\beta} + \vec{u}$$

$$\vec{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{21} & \ldots & X_{k1} \\ 1 & X_{22} & \ldots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \ldots & X_{kn} \end{pmatrix} \quad \vec{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad \vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

SUPUESTOS

Para el análisis estadístico del modelo, comenzaremos considerando las siguientes **hipótesis básicas:**

1. **Supuesto de Linealidad:** Existe una relación lineal entre la variable dependiente y las variables explicativas, lo que permite usar álgebra de matrices y vectores para simplificar el análisis. *Es decir, podemos escribir la relación muestral tipo así:*

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

con $i = 1, \ldots, n$ o equivalentemente $Y_i = \vec{X}_i^t \vec{\beta} + u_i$.

(*Aquí está la definición de Yi ;) *)

2. **Supuesto de Rango Completo por Columnas:** La matriz de observaciones debe tener rango completo por columnas, lo que significa que las variables explicativas no deben ser combinaciones lineales entre sí, evitando problemas de *multicolinealidad.* Es decir, las columnas "k" de la matriz X son linealmente independientes. Esto nos permite además calcular la inversa de la matriz de productos cruzados: $(X^tX)^{-1}$

3. **Supuesto de Exogeneidad:** Las variables explicativas no deben tener ninguna relación con el término de perturbación. Es decir, el valor esperado de la perturbación debe ser cero, lo que garantiza que las variables explicativas no predicen el término de error. Es decir, se cumple esto (*Recuerda la definición de Yi):

$$E\left[u_i \middle| (X_{1i}, X_{2i}, \ldots, X_{ki})\right] = 0 \qquad E\left[Y_i \middle| (X_{1i}, X_{2i}, \ldots, X_{ki})\right] = \overrightarrow{X}_i^t \overrightarrow{\beta}.$$

4. **Supuesto de Causalidad:** La relación entre las variables explicativas y la variable dependiente es unidireccional; las variables explicativas causan cambios en la variable dependiente, pero no al revés. (Esto ya se enunció en el T1, las de la derecha explican la de la izquierda, y esta última depende de la perturbación)

5. **Supuesto sobre la Perturbación:** El término de perturbación debe ser: (Ya visto en T1)
   - **Centrado:** Su valor esperado es cero. $E[u_i] = 0 \; \forall i = 1, \ldots, n.$
   - **Homocedástico:** Su varianza es constante para todas las observaciones.

     $\text{var}[u_i] = E[u_i^2] = \sigma^2$ Si no se cumple habrá *heterocedasticidad.*

   - **Incorrelado:** Las perturbaciones en diferentes observaciones no están correlacionadas. $[u_i u_j] = E[u_i u_j] = 0$

     Si no se cumple habrá *autocorrelación.*

Consecuentemente, se tiene que la matriz de varianzas-covarianzas del término perturbación es: $\text{var}[\overrightarrow{u}] = E\left[(\overrightarrow{u} - E[\overrightarrow{u}])(\overrightarrow{u} - E[\overrightarrow{u}])^t\right] = E[\overrightarrow{u}\,\overrightarrow{u}^t]$ *siendo este el procedimiento que utilizaremos para el cálculo de la varianza en clase (sabérselo).*

6. **Supuesto de Normalidad de la Perturbación:** Se asume que las *perturbaciones siguen una distribución normal con media cero y varianza constante,* lo que permite hacer inferencias sobre los parámetros del modelo.

$$u_i \sim N(0, \sigma^2) \quad \forall i = 1, \ldots, n \qquad \overrightarrow{u} \sim N\left(\overrightarrow{0}, \sigma^2 \mathrm{I}_n\right)$$

Con la función de densidad asociada a la distribución (extra):

$$f(\overrightarrow{u}|\sigma^2) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left[-\frac{1}{2\sigma^2}\overrightarrow{u}^t\overrightarrow{u}\right]$$

ESTIMACIÓN: MÍNIMOS CUADRADOS ORDINARIOS (MCO)

Definición: Es una técnica utilizada para estimar los parámetros en modelos de regresión lineal, buscando minimizar la diferencia entre los valores observados y los valores predichos por el modelo.

Para esto denotaremos dentro del **modelo lineal múltiple clásico:** $\vec{y} = X\vec{\beta} + \vec{u}$
Los siguiente valores:

Denotando por $\hat{\vec{y}}$ el valor ajustado de $\vec{y}$ por el modelo: $\hat{\vec{y}} = X\hat{\vec{\beta}}$

[Residuo] Diferencia existente entre el valor observado de la variable explicada y su estimación: $\vec{e} = \vec{y} - \hat{\vec{y}} = \vec{y} - X\hat{\vec{\beta}}$

Denotaremos a la función a minimizar por: $f(\hat{\vec{\beta}}) = \vec{e}^{\,t}\vec{e}$.

Una vez realizado las siguientes definiciones, comenzaremos con los pasos a seguir:

1. **Definimos la expresión a minimizar,** definiendo la suma de los cuadrados de los residuos como:

$$\begin{aligned} f\left(\hat{\vec{\beta}}\right) &= \left(\vec{y} - X\hat{\vec{\beta}}\right)^t \left(\vec{y} - X\hat{\vec{\beta}}\right) \\ &= \vec{y}^{\,t}\vec{y} - \vec{y}^{\,t}X\hat{\vec{\beta}} - \hat{\vec{\beta}}^{\,t}X^t\vec{y} + \hat{\vec{\beta}}^{\,t}X^tX\hat{\vec{\beta}} \\ &= \vec{y}^{\,t}\vec{y} - 2\hat{\vec{\beta}}^{\,t}X^t\vec{y} + \hat{\vec{\beta}}^{\,t}X^tX\hat{\vec{\beta}} \end{aligned}$$

2. **Derivamos con respecto al vector de beta estimada** e **igualamos a 0**:

$$\nabla f\left(\hat{\vec{\beta}}\right) = -2X^t\vec{y} + 2X^tX\hat{\vec{\beta}} = \vec{0}$$

3. **Despejamos el vector de parámetros** y **calculamos la inversa**:

$$-2X^t\vec{y} + 2X^tX\hat{\vec{\beta}} = \vec{0} \Leftrightarrow X^tX\hat{\vec{\beta}} = X^t\vec{y}$$

Considerando el supuesto de rango completo por columnas, sabemos que se verifica que $\rho(X^tX) = k$, condición necesaria para la existencia de la matriz $(X^tX)^{-1}$, luego:

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

4. Para que la solución se ha de cumplir:

$$Hess(f) = 2X^tX \succ 0 \quad (A \succ 0 \text{ sii } z^tAz > 0)$$

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

es el **Estimador Mínimo Cuadrático Ordinario (EMCO)** del vector de parámetros $\vec{\beta}$ del modelo de regresión lineal múltiple $\vec{y} = X\vec{\beta} + \vec{u}$.

Si el modelo tiene término independiente, este dentro de su matriz en la posición 0,0 tendrá el término n, en caso contrario no. (Si no se entiende consultar diapo. nº18 T2)

PROPIEDADES ALGEBRAICAS DE LOS MCO

- La suma de los residuos mínimo cuadráticos es cero: $\sum_{i=1}^{n} e_i = 0$.
- La suma de los valores observados coincide con la suma de los valores estimados: $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \widehat{Y}_i$.

Además debemos saber que:

- **SCT:** Cuánto varía la observada de la media. *(Suma de los cuadrados totales)*
- **SCR:** Cuánto varía la observada estimada. *(Suma de los cuadrados de los residuos)*
- **SCE:** Cuánto varía la estimada respecto a la media. *(Suma de cuadrados explicada)*
- **SCT = SCR + SCE**

$$SCT = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \overrightarrow{y}^t\overrightarrow{y} - n \cdot \overline{Y}^2$$

$$SCR = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \overrightarrow{y}^t\overrightarrow{y} - \overrightarrow{\beta}^t X^t \overrightarrow{y}$$

$$SCE = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 = \overrightarrow{\beta}^t X^t \overrightarrow{y} - n \cdot \overline{Y}^2$$

- Se cumple que $\overrightarrow{\widehat{y}}^t \overrightarrow{e} = 0$.

Considerando que **Sí** hay término independiente en el modelo ($X_{1i} = 1$), se tiene:

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

La condición necesaria de EMCO era:

$$-2X^t\overrightarrow{y} + 2X^tX\overrightarrow{\beta} = \overrightarrow{0}$$

Equivalentemente:

$$X^t\left(\overrightarrow{y} - X\overrightarrow{\beta}\right) = \overrightarrow{0}$$

Como $\overrightarrow{\widehat{y}} = X\overrightarrow{\beta}$, tenemos que: $X^t\left(\overrightarrow{y} - \overrightarrow{\widehat{y}}\right) = \overrightarrow{0}$, ó equivalentemente

$$X^t\overrightarrow{e} = \overrightarrow{0}.$$

- A continuación en las diapositivas del tema, daba el ejem Ecuaciones Normales (estarán en el archivo de ejercicios importancia)

BONDAD DEL AJUSTE

[Coeficiente de determinación ordinario / R^2] Porcentaje de variabilidad explicada por el modelo.

$$R^2 = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{Y}\right)^2}{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} = \frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} = \frac{SCE}{SCT}.$$

Si el modelo tiene término independiente: $R^2 = 1 - \frac{SCR}{SCT}$ y los valores del coeficiente estarán comprendidos entre 0 y 1.

Además de clase sabemos que: SCT=SCR+SCE y por lo tanto:

- Si SCR=0 -> SCT=SCE y R^2=1
- Si SCE=0 -> SCT=SCR y R^2=0

Suele decirse que: Series temporales(R^2 aprox.=0,85) y Sección cruzada(R^2 aprox.=0,6-0,7)

[Coeficiente de determinación corregido / _R^2] Porcentaje de variación de la variable explicada considerando el número de variables incluidas en el modelo, es decir, considerando el valor de k.

$$\overline{R}^2 = 1 - \frac{SCR/(n-k)}{SCT/(n-1)} = 1 - (1 - R^2)\cdot\frac{n-1}{n-k}.$$

Adviértase que obtener un R^2 o $\overline{R}^2$ cercano a 1 no indica que los resultados sean fiables, ya que, por ejemplo, puede ser que no se cumpla alguna de las hipótesis básicas y los resultados no ser válidos.
Es necesario señalar que ambos coeficientes no son capaces de detectar si alguna de las variables incluidas en el modelo son o no estadísticamente significativas.
Por tanto, estos indicadores han de ser considerados como una herramienta más a tener en cuenta dentro del análisis.

CRITERIOS DE AKAIKE Y SCHWARZ (k parámetros, n observaciones)

Para comparar distintos modelos podríamos utilizar el coeficiente de determinación. Sin embargo, para poder llevar cabo dicha comparación será nes necesario que tengan la misma variable explicada.
Con el fin de buscar una solución al problema utilizaremos los criterios de selección de modelos de Akaike (AIC) y el bayesiano de Schwarz (BIC). Estos criterios se obtienen a partir de la suma de cuadrados de los residuos y de un factor que penaliza la inclusión de parámetros.
El criterio de información de Akaike:

$$AIC = \ln\left(\frac{SCR}{n}\right) + \frac{2k}{n},$$

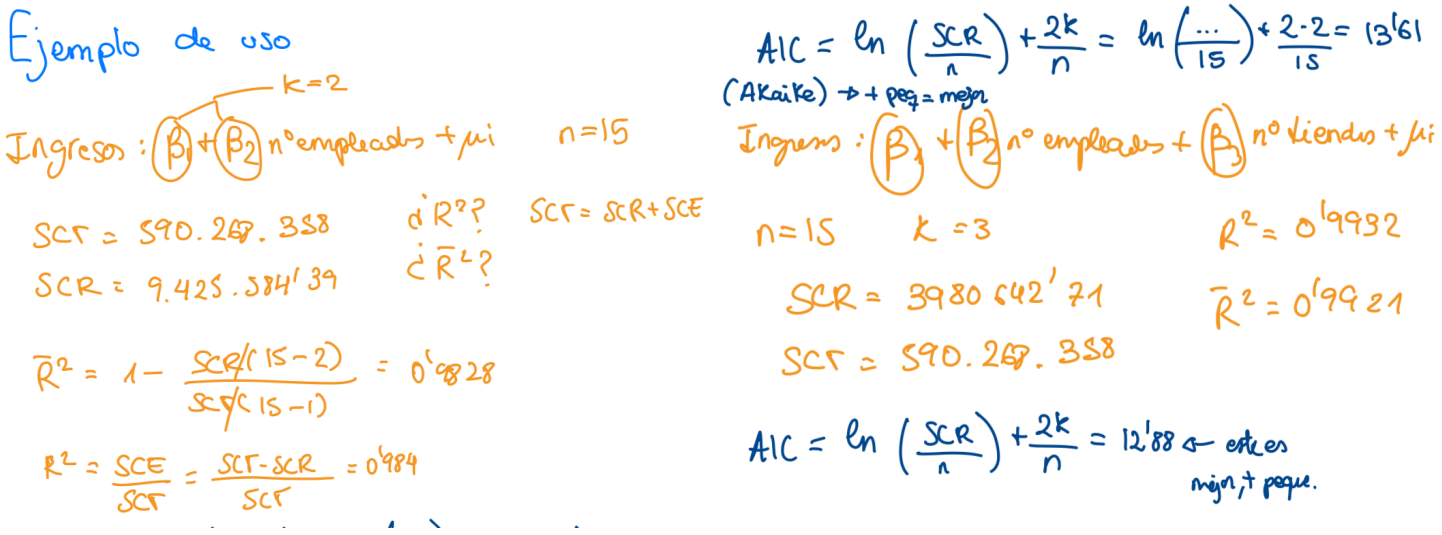
El criterio de información de Schwarz:

$$BIC = \ln\left(\frac{SCR}{n}\right) + \frac{k}{n} \cdot \ln(n).$$

Utilizando estos criterios se escogería aquel modelo con un menor valor de AIC o BIC.

26 / 32

Acompañado de un ejemplo que hicimos en clase:



El resto de cosas apuntadas son ejercicios que pondré en el archivo correspondiente.

Tema 3: Modelo Lineal (II)

Sabemos del tema anterior que:

$\vec{y} = x\vec{\beta} + \vec{u}$ $E(u)=0$ $Var(u)=\sigma^2$ $E(\vec{\beta})$ $Var(\vec{\beta})$

Donde u sería la variable aleatoria con tal media y varianza. En caso de que el modelo no tuviera u, sería la beta la variable aleatoria.

(Con la esperanza) -> No creo que caiga en el examen, es necesario saber que es.
DEMOSTRACIÓN DE QUE UN ESTIMADOR INSESGADO $\cdot\ \vec{\hat{\beta}}$ es un estimador **insesgado** de $\vec{\beta}$.

Debemos saber las siguientes afirmaciones:

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

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

Y sustituyendo …

$\vec{\hat{\beta}} = (x^tx)^{-1}x^t\vec{y}$, $\vec{y} = x\cdot\vec{\beta} + \vec{u}$ $\Rightarrow$ $\vec{\hat{\beta}} = (x^tx)^{-1}\cdot x^t(x\cdot\vec{\beta} + \vec{u}) = (x^tx)^{-1}\cdot x^tx\vec{\beta} + (x^tx)^{-1}x^t\vec{u}$

$E(\vec{\hat{\beta}}) = \vec{\beta} + E[(x^tx)^{-1}x^t\vec{u}) = \vec{\beta}$ ($I \to 0$)

Vemos que si es un estimador insesgado.

(Con la varianza) -> No creo que caiga en el examen, es necesario saber que es.
DEMOSTRACIÓN DE LA VARIANZA DE LOS EMCO

Sabemos de TC II que: $Var(x) = E(x - E(x))^2$ (ver por matrices)

Pues para esta demostración necesitaremos saber que*: $Var(\vec{\hat{\beta}}) = E[(\vec{\hat{\beta}} - E(\hat{\beta}))\cdot(\vec{\hat{\beta}} - E(\hat{\beta}))^t]$ ($E(\hat{\beta}) = \vec{\beta}$)

Y esas diferencias, gracias a la demostración anterior sabemos que: $\vec{\hat{\beta}} - \vec{\beta} = (x^tx)^{-1}x^t\vec{u}$

Ahora podremos desarrollar lo de la varianza*: $= E[(x^tx)^{-1}x^t\vec{u}\cdot\vec{u}^tx(x^tx)^{-1}] =$

$= E[(x^tx)^{-1}x^t\underbrace{E(\vec{u}\cdot\vec{u}^t)}_{\sigma^2 I}x(x^tx)^{-1}] =$ (solo de las vars. aleatorias) $= \sigma^2[(x^tx)^{-1}x^t x(x^tx)^{-1}] = \sigma^2(x^tx)^{-1}$ ($x(x^tx)^{-1} \to I$)

DEMOSTRACIÓN DE TEOREMA GAUSS-MARKOV (95% de que caiga en el examen)

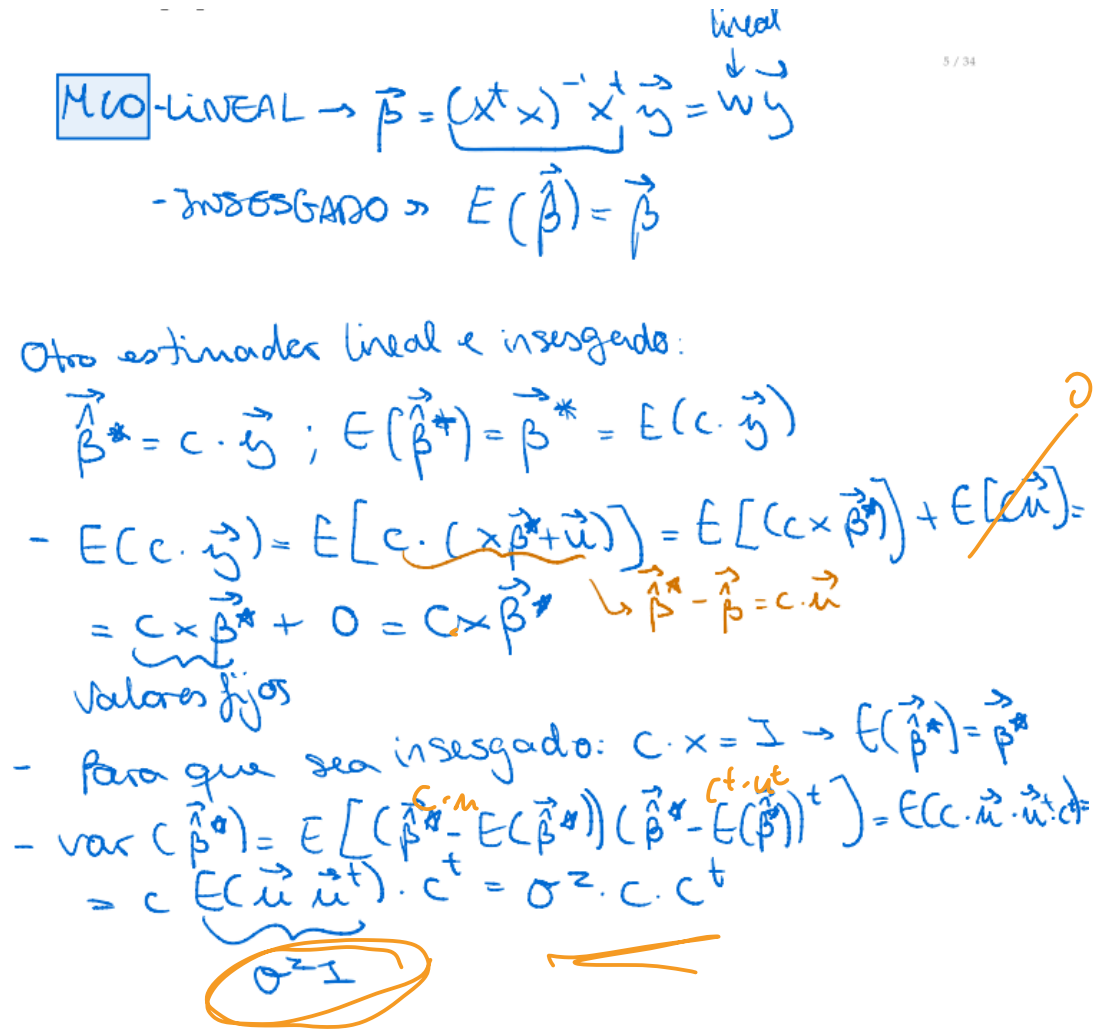
¿En qué consiste?

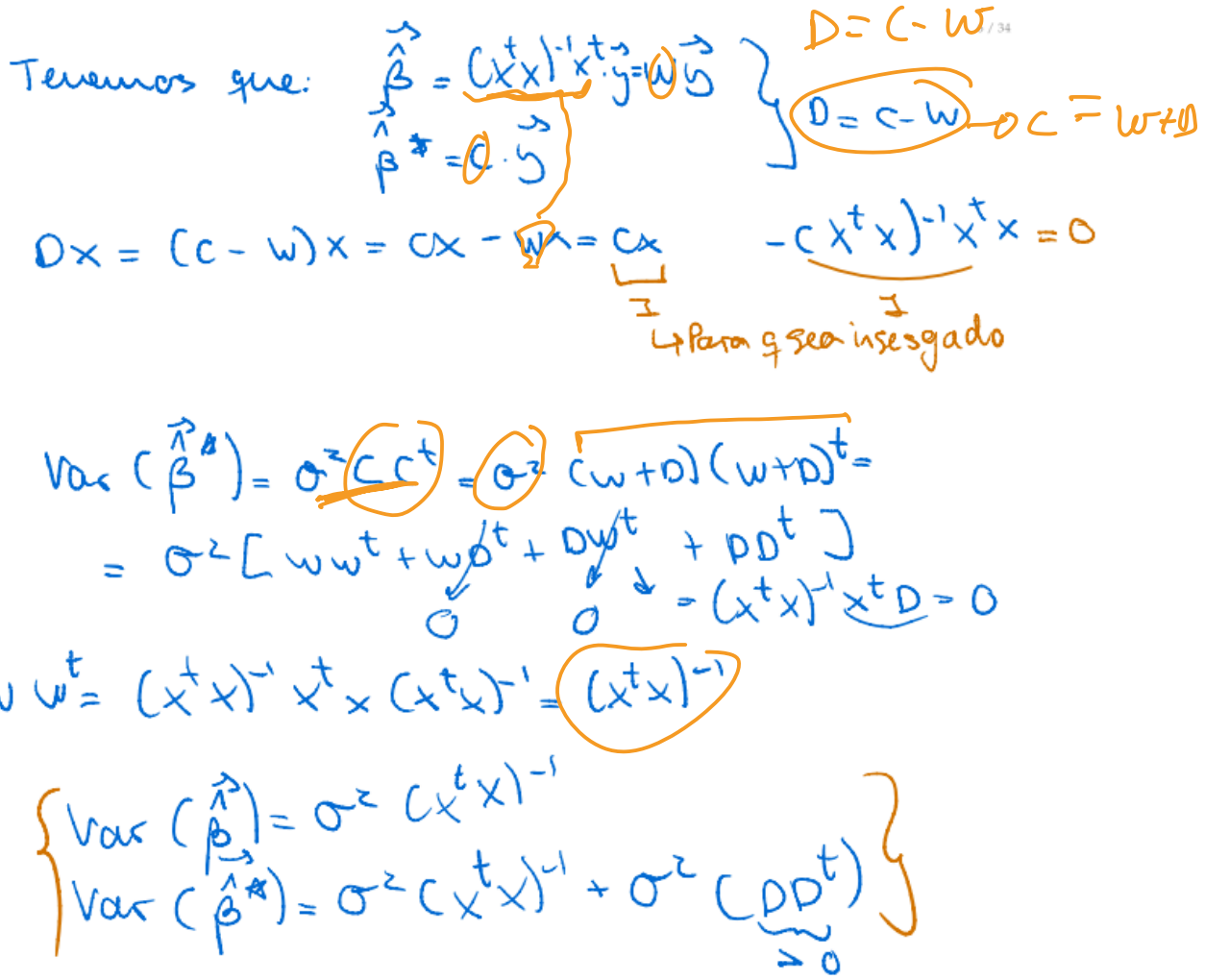
El Teorema de Gauss-Markov establece que, en un modelo de regresión lineal bajo ciertas condiciones, **los estimadores de mínimos cuadrados ordinarios** (OLS, por sus siglas en inglés) son **los estimadores lineales insesgados con varianza mínima.** Esto significa que, entre todos los estimadores lineales insesgados, los estimadores OLS **son los más eficientes** (es decir, tienen la menor varianza).

¿Para qué sirve?

Este teorema es fundamental en econometría y estadística, ya que garantiza que el método de mínimos cuadrados produce estimaciones óptimas bajo los supuestos adecuados. Se utiliza en la estimación de parámetros en modelos lineales para obtener predicciones precisas y fiables.

DEMOSTRACIÓN: Debemos saber ciertas afirmaciones en primer lugar (denotando W):





Finalizando con la conclusión de que la varianza del otro estimador beta será mayor a la varianza del beta original. (Es decir, es el más eficiente)

ESTIMACIÓN DE σ_u^2 (Lo que hay que saber)

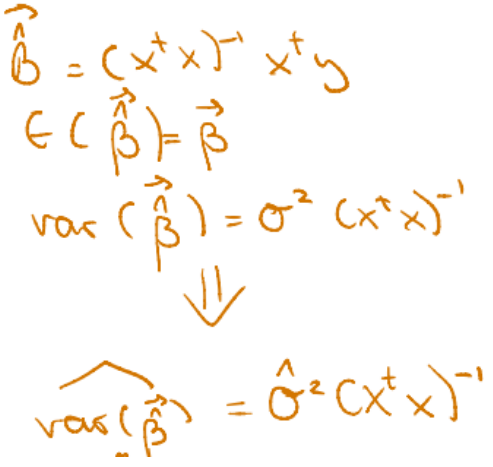
$$\hat{\sigma}_u^2 = \frac{\vec{e}^t\vec{e}}{n-k} = \frac{SCR}{n-k}$$

$$E\left[\vec{e}^t\vec{e}\right] = \sigma^2(n-k)$$

(Ignora la I dibujada en la igualdad)

Recuerda que SCR:

$$\vec{y}^t\vec{y} - \vec{\beta}^t X^t \vec{y}$$



(Aquí hicimos ejercicios, que estarán en el archivo correspondiente)

ESTIMACIÓN DE LA VARIANZA DE LOS EMCO (~~Saber usarlos, nos darán hoja formu.~~)

$$\text{var}\left(\overrightarrow{\hat{\beta}}\right) = \sigma^2 \cdot (X^t X)^{-1} \qquad \widehat{\sigma}^2 = \frac{\overrightarrow{e}^t \overrightarrow{e}}{n-k} = \frac{SCR}{n-k}, \qquad \widehat{\text{var}}\left(\overrightarrow{\hat{\beta}}\right) = \frac{SCR}{n-k} \cdot (X^t X)^{-1}$$

SUPUESTO DE NORMALIDAD E INFERENCIA
(Más fórmulas que no hacen falta saberse pero sí usar en ejercicios)

- Suponíamos que $\overrightarrow{u} \sim \mathcal{N}\left(\overrightarrow{0}, \sigma^2 I_n\right)$.
- Como $\overrightarrow{\hat{\beta}} = \overrightarrow{\beta} + (X^t X)^{-1} X^t \overrightarrow{u}$, y por T. Gauss-Markov:

$$\overrightarrow{\hat{\beta}} \sim \mathcal{N}\left(E\left[\overrightarrow{\hat{\beta}}\right], \text{var}\left(\overrightarrow{\hat{\beta}}\right)\right) = \mathcal{N}\left(\overrightarrow{\beta}, \sigma^2 \cdot (X^t X)^{-1}\right)$$

o equivalentemente:

$$\left(\overrightarrow{\hat{\beta}} - \overrightarrow{\beta}\right) \sim \mathcal{N}\left(\overrightarrow{0}, \sigma^2 \cdot (X^t X)^{-1}\right)$$

Concluimos con que sigue una chi-cuadrado de n-k grados de libertad:

$$\text{Concluimos que } \frac{1}{\sigma^2} \overrightarrow{u}^t M_X \overrightarrow{u} \sim \chi^2_{n-k} \Rightarrow \frac{(n-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$$

INTERVALOS DE CONFIANZA:

“Para la media con varianza desconocida” de TC2:

- IC para β_i:

$$\widehat{\beta}_i \pm t_{n-k,1-\frac{\alpha}{2}} \cdot \sqrt{\widehat{\text{var}}[\widehat{\beta}_i]}, \quad i = 1, \ldots, k. \qquad \text{var}\left(\overrightarrow{\hat{\beta}}\right) = \sigma^2 \cdot (X^t X)^{-1}$$

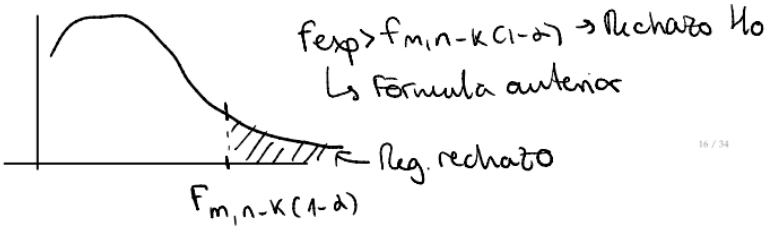
- IC para σ^2

$$\left[\frac{(n-k)\cdot\widehat{\sigma}^2}{\chi^2_{n-k,1-\frac{\alpha}{2}}}, \frac{(n-k)\cdot\widehat{\sigma}^2}{\chi^2_{n-k,\frac{\alpha}{2}}}\right], \qquad \widehat{\sigma}^2_u = \frac{\overrightarrow{e}^t \overrightarrow{e}}{n-k} = \frac{SCR}{n-k}$$

CONTRASTE DE HIPÓTESIS:

La hipótesis nula se rechazará si el valor del estadístico cae en la región de rechazo.

Es decir, si $F_{exp} > F_{m,n-k,1-\alpha}$ entonces se rechaza la hipótesis $H_0 : R\overrightarrow{\beta} = \overrightarrow{r}$.



PARA LOS EJERCICIOS
CASOS PARTICULARES (i): SIGNIFICACIÓN INDIVIDUAL

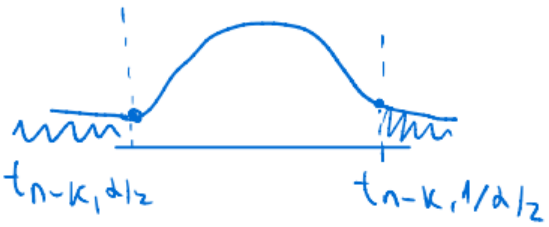
Si $m = 1$ y $r_i = 0\ \forall i$.

- Hipótesis:

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

- Estadístico de contraste:

$$t_{exp} = \left| \frac{\widehat{\beta}_i}{\sqrt{\widehat{\text{var}}[\widehat{\beta}_i]}} \right|$$



- Conclusión: Se rechaza H_0 si $t_{exp} > t_{n-k,1-\frac{\alpha}{2}}$.

Es val. abs.

CASOS PARTICULARES (ii): ÚNICA RESTRICCIÓN LINEAL

$m = 1$ y $r_i \neq 0\ \forall i$.

- Hipótesis: $\begin{cases} H_0 : \beta_i = r_i \\ H_1 : \beta_i \neq r_i \end{cases}$
- Estadístico de contraste:

$$t_{exp} = \left| \frac{\widehat{\beta}_i - r_i}{\sqrt{\widehat{\text{var}}[\widehat{\beta}_i]}} \right|$$

- Conclusión: Se rechaza H_0 si $t_{exp} > t_{n-k,1-\frac{\alpha}{2}}$.

CASOS PARTICULARES (iii): ÚNICA RESTRICCIÓN LINEAL

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

Todos menos el término indep. (β_1)

- Hipótesis: $\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \text{ con } i = 2, 3, \ldots, k \end{cases}$
- Estadístico de contraste:

$$F_{exp} = \left(R\vec{\beta}\right)^t \cdot \frac{\left[R\left(X^tX\right)^{-1}R^t\right]^{-1}}{(k-1)\cdot\widehat{\sigma}^2} \cdot \left(R\vec{\beta}\right)$$

- Conclusión: Se rechaza H_0 si $F_{exp} > F_{k-1,n-k,1-\alpha}$.

ANOVA

Considerando que se verifica que SCT = SCR + SCE

Test sign. global: $\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = 0 \\ H_1 : \exists \beta_i \neq 0 \end{cases}$

$$F_{exp} = \frac{\frac{1}{k-1}SCE}{\frac{1}{n-k}SCR}$$

Regla decisión: $F_{exp} > f_{k-1,n-k}(1-\alpha)$
rechazo H_0.

TABLA ANOVA

Fuente de variación	Suma de Cuadrados	Grados de Libertad	Medias
Explicada	$SCE = \vec{\hat{\beta}}^t X^t \vec{y} - n\overline{Y}^2$	$k-1$	$\frac{SCE}{k-1}$
Residuos	$SCR = \vec{y}^t \vec{y} - \vec{\hat{\beta}}^t X^t \vec{y}$	$n-k$	$\frac{SCR}{n-k}$
Total	$SCT = \vec{y}^t \vec{y} - n\overline{Y}^2$	$n-1$	

Como

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

$$F_{exp} = \frac{\frac{1}{k-1}SCE}{\frac{1}{n-k}SCT(1-R^2)} = \frac{\frac{1}{k-1}R^2}{\frac{1}{n-k}(1-R^2)} = \frac{n-k}{k-1}\frac{R^2}{(1-R^2)}$$

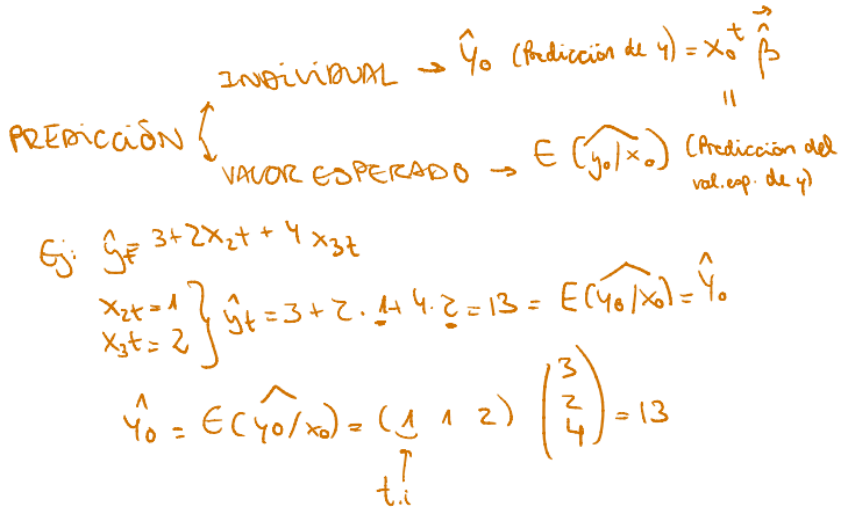
COTA R^2

Se rechaza H_0 cuando $F_{exp} > F_{k-1,n-k,1-\alpha}$, esto es si

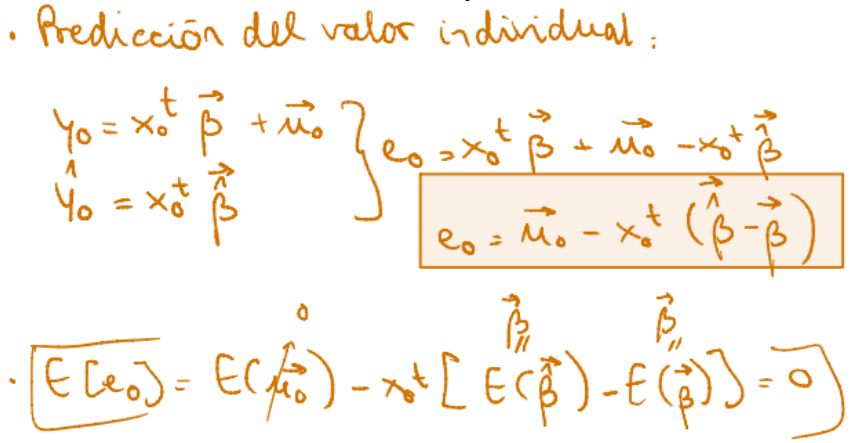
$$\frac{n-k}{k-1}\frac{R^2}{(1-R^2)} > F_{k-1,n-k,1-\alpha} \Rightarrow R^2 > \frac{(k-1)\cdot F_{k-1,n-k,1-\alpha}}{(n-k)+(k-1)\cdot F_{k-1,n-k,1-\alpha}}.$$

PREDICCIÓN PUNTUAL

Tiene que coincidir el individual y el esperado en la pred.puntual, veamos en un ejemplo:



PREDICCIÓN PUNTUAL para EL VALOR INDIVIDUAL (esperanza, no saber)



PREDICCIÓN PUNTUAL para EL VALOR INDIVIDUAL (varianza, esta tengo apuntado que si sabersela, pero no se)

$$Var(e_0) = Var\left[\vec{\mu}_0 - x_0^t(\vec{\hat{\beta}} - \vec{\beta})\right] =$$

$$= Var(\vec{\mu}_0) + Var(x_0^t(\vec{\hat{\beta}} - \vec{\beta})) - 2\,covar\left[\vec{\mu}_0\, x_0^t(\vec{\hat{\beta}} - \vec{\beta})\right] = \quad (\vec{\mu}_0 \text{ y } x_0^t(\vec{\hat{\beta}} - \vec{\beta}) \text{ son independientes, covar} = 0)$$

$$= \sigma^2 + E\left[x_0^t(\vec{\hat{\beta}} - \vec{\beta})(\vec{\hat{\beta}} - \vec{\beta})^t x_0\right] = \sigma^2 + x_0^t E\left[(\vec{\hat{\beta}} - \vec{\beta})(\vec{\hat{\beta}} - \vec{\beta})^t\right] x_0 =$$

$$= \sigma^2 + x_0^t\, \sigma^2 (x^t x)^{-1} x_0 = \sigma^2\left(1 + x_0^t (x^t x)^{-1} x_0\right)$$

$$e_0 \sim N\left(0, \sigma^2\left(1 + x_0^t (x^t x)^{-1} x_0\right)\right)$$

* $Var(x) = E(x - E[x])^2$
Ls Con matrices en vez de ^{2}, se hace por la trasp uesta

- Predicción del valor esperado:

$$\left.\begin{array}{l} E(y_0/x_0) = \vec{x}_0^t \vec{\beta} \\ E(y_0|x_0) = \vec{x}_0^t \vec{\beta} \end{array}\right\} \quad e_0^* = \vec{x}_0^t(\vec{\hat{\beta}} - \vec{\beta})$$

(Lo mejor bien dps para llegar a la matriz de vars-covars)

$$E(e_0^*) = E\left[-\vec{x}_0^t(\vec{\hat{\beta}} - \vec{\beta})\right] = -x_0^t\left(E(\vec{\hat{\beta}}) - E(\vec{\beta})\right) = 0$$

$$Var(e_0^*) = E\left(\vec{x}_0^t(\vec{\hat{\beta}} - \vec{\beta})(\vec{\hat{\beta}} - \vec{\beta})^t \vec{x}_0\right) =$$

$$= \vec{x}_0^t \underbrace{E\left[(\vec{\hat{\beta}} - \vec{\beta})(\vec{\hat{\beta}} - \vec{\beta})^t\right]}_{\sigma^2(x^t x)^{-1}} \vec{x}_0 = \sigma^2 x_0^t (x^t x)^{-1} x_0$$

Resumen:

Valor teórico	Valor estimado	Residuo	Distribución
y_0	$\hat{y}_0$	e_0	$N(0, \sigma^2(1 + x_0^t(x^t x)^{-1}x_0))$
$E(y_0/x_0)$	$E(\hat{y_0}\|x_0)$	e_0^*	$N(0, \sigma^2 \cdot x_0^t(x^t x)^{-1} x_0)$

$$\hat{y}_0 = E(\hat{y_0}/x_0) = x_0^t \vec{\hat{\beta}}$$

PREDICCIÓN POR INTERVALO

	Predicción puntual	Predicción por intervalos
V. individual	$\vec{x}_0^t \vec{\hat{\beta}}$	$\vec{x}_0^t \vec{\hat{\beta}} \pm t_{n-k(1-\alpha/2)} \hat{\sigma} \sqrt{1 + x_0^t (x^t x)^{-1} x_0}$
V. esperado	$\vec{x}_0^t \vec{\hat{\beta}}$	$\vec{x}_0^t \vec{\hat{\beta}} \pm t_{n-k(1-\alpha/2)} \hat{\sigma} \sqrt{x_0^t (x^t x)^{-1} x_0}$

Recordatorio de distribuciones (TC2) + Ej 11/12 importante (en archivo de ejercicios)

DISTRIBUCIONES χ^2 Y t-STUDENT

- La distribución χ^2 de Pearson con n grados de libertad, χ^2_n, se construye como la suma de los cuadrados de n variables aleatorias independientes e idénticamente distribuidas según una $\mathcal{N}(0,1)$:

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

- La distribución t-Student con n grados de libertad, t_n, se construye como el cociente entre una variable aleatoria $\mathcal{N}(0,1)$ y la raíz cuadrada de una χ^2-cuadrado de n grados de libertad dividida entre sus grados de libertad, siendo ambas distribuciones independientes:

$$t_n = \frac{X}{\sqrt{\frac{Y}{n}}}, \quad X \sim N(0,1),\ Y \sim \chi^2_n.$$

Propiedades:

- χ^2 NO es simétrica, t-Student SÍ es simétrica.
- Para $n > 30$, la distribución t-Student se puede aproximar a una distribución Normal.

32 /

DISTRIBUCIÓN F-SNEDECOR

- La distribución F de Snedecor con n y m grados de libertad, que denotaremos por $F_{n,m}$, se construye a partir del cociente de dos variables aleatorias independientes y distribuidas según chi-cuadrados con n y m grados de libertad, respectivamente, divididas entre sus correspondientes grados de libertad. Por tanto, la distribución F de Snedecor responde a la siguiente estructura:

$$F_{n,m} = \frac{\frac{X}{n}}{\frac{Y}{m}}, X \sim \chi^2_n,\ Y \sim \chi^2_m.$$

Propiedades:

- La distribución $F_{n,m}$ NO es simétrica.
- $F_{m,n,1-\alpha} = \frac{1}{F_{n,m,\alpha}}$.
- $t_n^2 = F_{1,n}$.

33 /

Notación matricial (nada)

Tema 4: Multicolinealidad

CONCEPTO:

La multicolinealidad ocurre cuando hay relaciones lineales entre las variables independientes en un modelo (2 o más variables).

Puede ser perfecta (exacta) o aproximada.

Dentro de la aproximada, puede clasificarse como errática (inesperada) o sistemática (predecible), y esencial (afecta el modelo) o no esencial (no afecta significativamente).

MULTICOLINEALIDAD EXACTA

Concepto: Hace referencia a que existe una relación lineal exacta entre 2 o más var. indep.

Causas: Esta multicolinealidad se traduce en el incumplimiento de la hipótesis básica del modelo uniecuacional múltiple: la matriz X no es de rango completo por columnas, rg(X)<k

Consecuencias: No permite que se invierta la matriz X^t*X por lo que ya no hay una solución única, sino infinitas.

MULTICOLINEALIDAD APROXIMADA

Concepto: Hace referencia a que existe una relación lineal aprox. entre 2 o más var. indep.

Causas:

1. Relación causal entre las variables explicativas del modelo
2. Escasa variabilidad en las observaciones de las variables independientes
3. Reducido tamaño de la muestra

TIPOS DE MULTICOLINEALIDAD APROXIMADA según sus causas:

1. **Spanos y McGuirk:**
   - **Multi. Sistemática:** Por alta correlación lineal entre variables exógenas.
   - **Multi. Errática:** Por problemas numéricos o mal manejo de datos.
2. **Marquandt y Snee:**
   - **Multi. No esencial:** Relación lineal de las variables exógenas con la constante (se corrige centrando las variables).
   - **Multi. Esencial:** Relación lineal entre las variables exógenas, excluyendo la constante. (Endógena=Explicada -> Exógenas=Explicativas/Independientes)

Luego, se podrían distinguir los siguientes cuatro casos:

Multicolinealidad	Sistemática	Errática
No esencial	1	2
Esencial	3	4

Consecuencias: No se cumplirá la hipótesis básica de que la matriz X es completa por columnas (rg(X)=k) por lo que se podrá invertir X^t*X y obtener estimadores por MCO, pero el determinante será próximo a 0, por lo que la inversa de lo anterior será alto y por tanto:

- Varianzas altas en los estimadores.
- Los contrastes de significación individuales no rechazan la hipótesis nula, pero las conjuntas sí.
- Los coeficientes estimados son inestables ante cambios pequeños.
- R^2 y R^2 corregido elevado (coeficiente de determinación elevado)

Ejemplo: La edad y la experiencia suelen tener una alta relación, por lo tanto será difícil separar el efecto de cada una sobre la variable dependiente y que se produzca multicolinealidad debido a la relación entre ambas variables.

EFECTOS NOCIVOS SOBRE EL ANÁLISIS ESTADÍSTICO DEL MODELO

(Ejercicios y ejemplos que estarán en el archivo de ejercicios)

FACTOR DE INFLACIÓN DE LA VARIANZA (FIV) ó VIF en inglés

Es una de las medidas más usadas para detectar el grado de multicolinealidad existente:

$$FIV(i) = \frac{1}{1 - R_i^2}, \quad i = 2, \ldots, p,$$

Para la esencial

Si esta medida es superior a 10, se supone que el grado de multicolinealidad presente es preocupante. Recordemos que el VIF no tiene en cuenta la relación de las variables exógenas del modelo, por tanto, no detecta la multicolinealidad no esencial.

NÚMERO DE CONDICIÓN (NC)

$$NC = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}},$$

Si esta medida es superior a 20 se supone que el grado de multicolinealidad es moderado, y si es superior a 30 es preocupante. Este si tiene en cuenta la relación de las variables exógenas del modelo.

Para la no esencial

OTRAS MEDIDAS

1. **Matriz de correlaciones simples (R):**
   - Analiza las relaciones entre pares de variables independientes (dos a dos).
   - Ignora la constante y no puede detectar multicolinealidad más compleja.
2. **Determinante de la matriz de correlaciones (det(R)):**
   - Considera estructuras más complejas que las relaciones dos a dos.

- También ignora la constante.
- Valores cercanos a cero indican una multicolinealidad (del tipo esencial) grave.

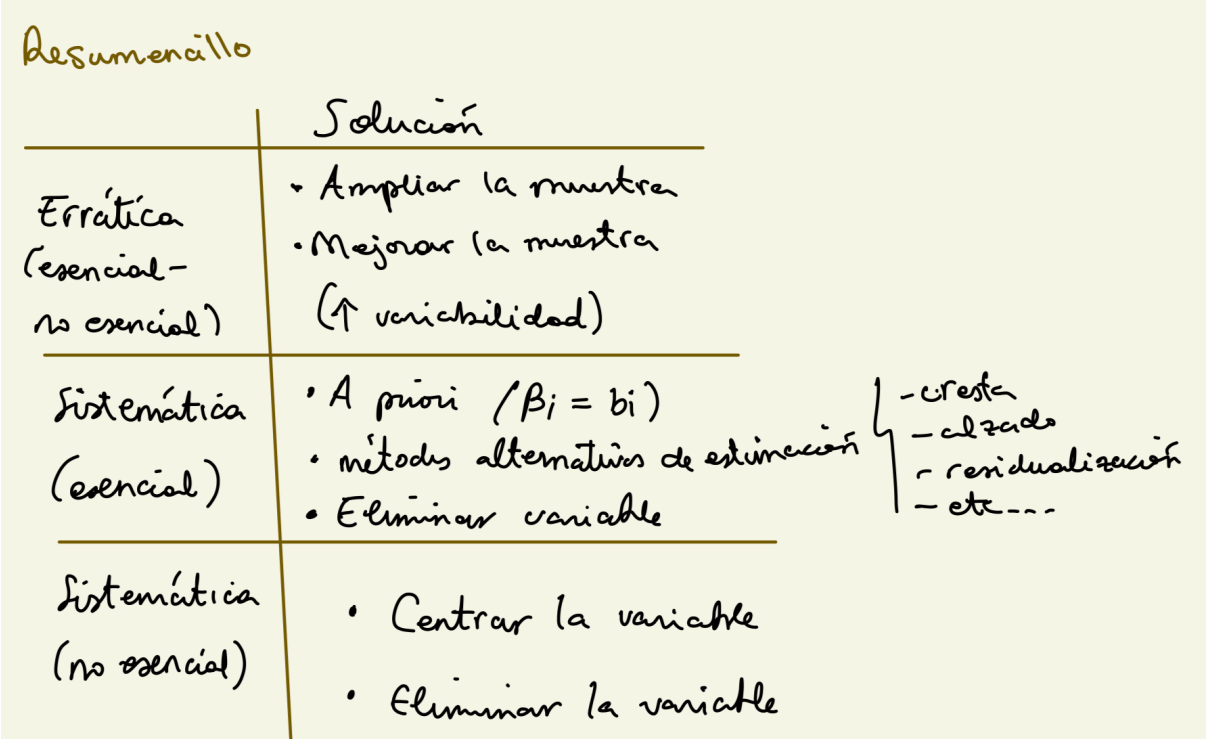
3. **Coeficiente de variación de las variables explicativas (CV(Xi)):**
   - Identifica baja variabilidad en las variables independientes.
   - Esto puede causar multicolinealidad no esencial.

SOLUCIONES Para abordar el problema de **multicolinealidad**:

1. **Mejorar el diseño muestral:** Maximizar la información obtenida de las variables observadas.
2. **Eliminar variables problemáticas:** Quitar aquellas que sean sospechosas de causar multicolinealidad.
3. **Aumentar el tamaño de la muestra:** Especialmente útil cuando se tienen pocas observaciones.
4. **Usar información extramuestral:** Incorporar información a priori y estimar el modelo mediante métodos como mínimos cuadrados restringidos.

Además, algunos autores sugieren tratar el problema con métodos numéricos, como:

- **Estimación cresta.**
- **Estimación alzada. (desde la perspectiva geométrica)**
- **Residualización.**



Tema 5: Heterocedasticidad

CONCEPTO: Se dice que un modelo de regresión presenta heteroscedasticidad cuando la varianza del término de perturbación no permanece constante a lo largo del tiempo. (Varianza no cte.)

CAUSAS:

1. Datos de sección cruzada:
   - La escala de la variable dependiente y el poder explicativo del modelo pueden variar entre observaciones.
2. Uso de datos agrupados:
   - Al agrupar observaciones en categorías y trabajar con sus medias en lugar de los datos originales, se introduce heterocedasticidad, aunque los datos originales sean homoscedásticos.
3. Omisión de una variable relevante:
   - Si una variable importante no está incluida en el modelo, la varianza del error puede depender de esta variable omitida, causando heterocedasticidad.

CONSECUENCIAS:

Puesto que en el método de estimación por MCO no influye la matriz de varianzas-covarianzas de la perturbación aleatoria es claro que el estimador por MCO será igualmente:

$$\widehat{\beta}_{MCO} = (X^t X)^{-1} X^t y.$$

$\widehat{\beta}_{MCO}$ es lineal e insesgado. Sin embargo, ya no se tiene asegurado que la varianza sea mínima:

$$\begin{aligned} Var\left(\widehat{\beta}_{MCO}\right) &= (X^t X)^{-1} X^t \mathbb{E}[u \cdot u^t] X (X^t X)^{-1} \\ &= \sigma^2 (X^t X)^{-1} X^t \Omega X (X^t X)^{-1}, \end{aligned}$$

distinta a la del modelo con perturbaciones esféricas: $\sigma^2 (X^t X)^{-1}$.

Por tanto, la consecuencia de la presencia de heteroscedasticidad en un modelo lineal es que los estimadores obtenidos, aunque serán lineales e insesgados, no serán óptimos.

7 / 26

PROCEDIMIENTOS DE DETECCIÓN de HETEROCEDASTICIDAD, 2 enfoques:

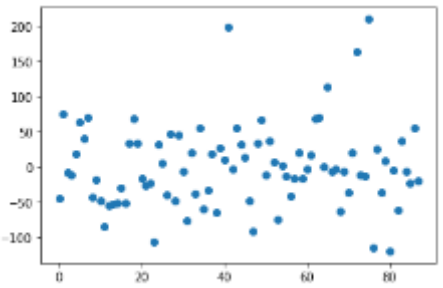
1. **Métodos gráficos:**
   - Analizar gráficos de residuos y de dispersión para identificar visualmente posibles variables que causen heterocedasticidad.
2. **Métodos analíticos:**
   - Realizar tests específicos, como:

- Glejser y Goldfeld-Quandt: Para muestras pequeñas, cuando se sospecha que una variable específica causa el problema.
- Breusch-Pagan y White: Para muestras grandes, cuando no se conoce la causa exacta de la heterocedasticidad.

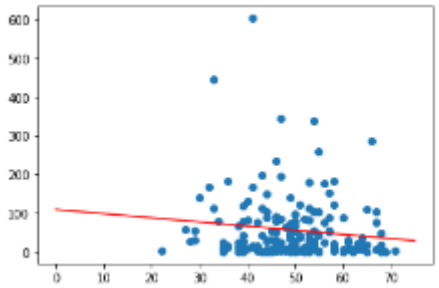
- En todos estos casos, la hipótesis nula es que el modelo es homocedástico (varianza constante).

MÉTODOS GRÁFICOS

GRÁFICOS DE LOS RESIDUOS: Si en dichos gráficos observamos grupos de observaciones con distinta varianza, podemos pensar en la presencia de heteroscedasticidad.



GRÁFICOS DE DISPERSIÓN: Si la variabilidad de los residuos aumenta o disminuye conforme aumenta el valor de la variable independiente, entonces podemos pensar que la varianza de la perturbación aleatoria depende de dicha variable y, por tanto, habría presencia de heteroscedasticidad.



muestras < 30.

CONTRASTE DE GOLDFELD QUANDT
(Ejemplo en el archivo de ejercicios y presentaciones)

Para muestras pequeñas, y si sospechamos que σ_i^2 está relacionada positivamente con la variable X_i:

1. Ordenar la información muestral, según los valores crecientes de este regresor, de menor a mayor.
2. Omitir los m valores muestrales centrales, (por ejemplo $m = \frac{n}{3}$).
3. Ajustar, mediante MCO, cada uno de los subgrupos obtenidos en el paso 2, una vez eliminados los valores centrales. Cada uno de estos subgrupos estarán formados por $\frac{n-m}{2}$ observaciones. La estimación del modelo del primer subgrupo nos permitirá obtener la SCR_1 y la del subgrupo 2, la SCR_2.
4. Bajo ausencia de heteroscedasticidad ($\sigma_1^2 = \sigma_2^2 = \sigma^2$):

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

se rechaza la hipótesis de homocedasticidad (hay heteroscedasticidad)

```
sms.het_goldfeldquandt(mco.resid, mco.model.exog)
```

CONTRASTE DE BREUSCH-PAGAN Y WHITE (Solo lo hicimos por comando)

Para número elevado de observaciones:

1. Estimamos el modelo $y = X\beta + u$ y el vector correspondiente a los residuos de mínimos cuadrados ordinarios, $e = y - \widehat{y}$.
2. Breusch-Pagan:
$$e^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_p X_p + v$$
3. White:
$$e^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_p X_p + + \delta_{p+1} x_1^2 + \delta_{p+2} x_1 x_2 + \cdots + \delta_{p+p^2} x_p^2 + v$$
4. Se contrasta $H_0 : \delta_1 = \delta_2 = \cdots = 0$ (HOMOCEDASTICIDAD)
5. Si se rechaza H_0: Existe heterocedasticidad!

```
sms.het_breuschpagan(mco.resid, mco.model.exog)
sms.het_white(mco.resid, mco.model.exog)
```

CONTRASTE DE GLESJER (ejemplo en diapos, pero no es importante)

Si sospechamos de que la heteroscedasticidad está ligada a X_i:

1. Estimamos el modelo $y = X\beta + u$ y $e = y - \widehat{y}$.
2. Se ajusta por MCO el modelo en el que la variable endógena es $|e_t|$ y la variable exógena es la variable X_i, la cual pensamos que es la que provoca la heteroscedasticidad, que será denotada por z_t, es decir
$$|e_t| = \delta_0 + \delta_1 z_t^h + v_t$$
donde v_t es ruido blanco. La regresión se realiza para distintos valores de $h = \pm 1, \pm 2, \pm \frac{1}{2} \ldots$.
3. Para cada h, contrastar $H_0 : \delta_1 = 0$ (t-test). Si se rechaza H_0, entonces existe heteroscedasticidad en el modelo.

ESTIMACIÓN DE MODELOS CON HETEROCEDASTICIDAD
Si conocemos *omega* Ω = Aplicación directa de MCG
Si no conocemos MCGF

MÍNIMOS CUADRADOS GENERALIZADOS (MCG)

Como $\Omega \succ 0$ y simétrica, por Descomposición de Cholesky, existirá una matriz $P(= Q^{-1})$, no singular, tal que,
$$\Omega^{-1} = P^t P$$

De manera que se podría transformar el modelo original para que la matriz Ω se convierta en la matriz identidad y se pudiera aplicar MCO sobre el modelo transformado aplicando estas expresiones:

$$\begin{aligned} \widehat{\beta}_{MCG} &= \left[(X^*)^t X^*\right]^{-1} \left[(X^*)^t y^*\right] = \left[X^t P^t P X\right]^{-1} \left[X^t P^t P y\right] = \\ &= \left[X^t \Omega^{-1} X\right]^{-1} \left[X^t \Omega^{-1} y\right] \end{aligned}$$

$$\text{var}\left[\widehat{\beta}_{MCG}\right] = \sigma^2 \left[(X^*)^t X^*\right]^{-1} = \sigma^2 \left[X^t P^t P X\right]^{-1} = \sigma^2 \left[X^t \Omega^{-1} X\right]^{-1}$$

ESTIMADOR MÍNIMOS CUADRADOS PONDERADOS

En el caso particular de la existencia de heterocedastidad, ¿Cómo se descompone $\Omega^{-1} = P^t P$?.

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

Considerando:

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

MÍNIMOS CUADRADOS PONDERADOS

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

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

- $\mathbb{E}[u_i^*] = \mathbb{E}[\frac{u_i}{\sqrt{w_i}}] = \frac{1}{\sqrt{w_i}} \mathbb{E}[u_i] = 0.$
- $Var(u_i^*) = Var(\frac{u_i}{\sqrt{w_i}}) = \frac{1}{w_i} Var(u_i) = \frac{1}{w_i} \sigma^2 w_i = \sigma^2$: HOMOCEDASTICIDAD!!!
- $\mathbb{E}[u_i^* u_j^*] = \mathbb{E}[\frac{u_i}{\sqrt{w_i}} \frac{u_j}{\sqrt{w_j}}] = \frac{1}{\sqrt{w_i w_j}} \mathbb{E}[u_i u_j] = 0$: INCORRELACIÓN!!

Podemos aplicar MCO para estimar $y^* = X^* \beta + u^*$.

(En presentaciones hay un ejemplo del test de Glesjer al final, con algo que hizo en la pizarra pero tampoco tiene mucha importancia, solo por revisar)

Tema 6: Autocorrelación

¿De dónde viene la autocorrelación, cual es su naturaleza?

$$y = X\beta + u$$

con $\mathbb{E}[u] = 0$ y $\text{var}[u] = \mathbb{E}[uu^t] = \sigma^2 I_n$, lo que implica:

- $\mathbb{E}[u_t] = 0, \forall t \in 1, \ldots, n$
- $\mathbb{E}[u_t^2] = \sigma^2, \forall t \in 1, \ldots, n$ (Homocesdasticidad)
- $\mathbb{E}[u_i, u_j] = \text{Cov}[u_i, u_j] = 0, \forall i \neq j \in 1, \ldots, n$ (Incorrelación).

Si $\text{Cov}[u_t, u_{t-k}] \neq 0 \; \forall k > 0 \rightarrow$ **Autocorrelación**.

CAUSAS

1. **Error de especificación:** Omitir variables exógenas (→ explicativas) relevantes para explicar la variable endógena. (→ explicado)
2. **Ciclos y tendencias:** Si los ciclos de la variable endógena no son explicados adecuadamente por las variables exógenas, surge autocorrelación en los errores (en la perturbación).
3. Existencia de relaciones no lineales
4. Existencia de relaciones dinámicas

CONSECUENCIAS

Los estimadores MCO son insesgados pero **no óptimos.**
El estimador por Mínimos Cuadrados Generalizados (MCG)

$$\widehat{\beta}_{MCG} = \left[X^t\Omega^{-1}X\right]^{-1}\left[X^t\Omega^{-1}y\right]$$

cuya matriz de varianzas-covarianzas viene dada por

$$\text{var}\left[\widehat{\beta}_{MCG}\right] = \sigma^2\left[X^t\Omega^{-1}X\right]^{-1}$$

Sí es **óptimo**.

Estimador MCO → No óptimo

Estimador MCG → óptimo

DETECCIÓN DE LA AUTOCORRELACIÓN

- **Gráficamente**
  - **Gráfico temporal de los residuos obtenidos por MCO:** Si los residuos están incorrelados deben distribuirse de forma aleatoria alrededor de 0.
  - **Gráfico de dispersión de los mismo frente a algún retardo suyo:** Los residuos estarán autocorrelacionados cuando el gráfico presenta una tendencia creciente o decreciente.
- **Analíticamente:**
  - **Contraste de Durbin-Watson.**
  - **Estadístico H de Durbin.**
  - **Contraste de Ljung-Box**

MÉTODOS GRÁFICOS

Para detectar autocorrelación en un modelo se pueden utilizar gráficos:

1. **Gráfico temporal de los residuos:**
   - Residuos incorrelados: Se distribuyen aleatoriamente alrededor de cero.
   - Residuos correlados:
     - Correlación Positiva: Rachas de residuos por encima y por debajo de la media.
     - Correlación Negativa: Alternancia en el signo de los residuos.
2. **Gráfico de dispersión:**
   - Positiva: Tendencia creciente en el gráfico.
   - Negativa: Tendencia decreciente.

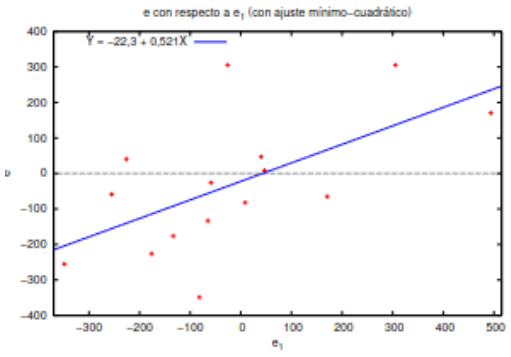


Figure. Gráfico de dispersión

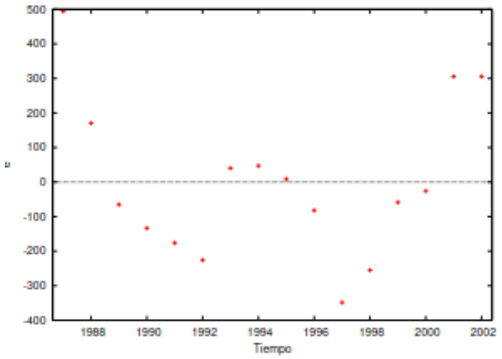


Figure. Gráfico temporal de los residuos

CONTRASTE DE DURBIN-WATSON

El contraste de Durbin-Watson resulta adecuado si suponemos que la autocorrelación que presenta el término de perturbación pudiera ser descrito mediante un proceso autorregresivo de orden uno:

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

donde ε_t es ruido blanco.
Estudiamos si el coeficiente ρ es significativo o no mediante los contrastes:

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

Estadístico de **Contraste de Durbin-Watson**:

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

siendo e_t los residuos MCO: $e_t = y_t - \widehat{y}_t = y_t - x\widehat{\beta}$.

Si la muestra seleccionada es suficientemente grande, se puede considerar que $\sum_{t=2}^{n} e_t^2 \simeq \sum_{t=2}^{n} e_{t-1}^2$, ya que ambas sumatorias difieren únicamente en la primera de las observaciones muestrales:

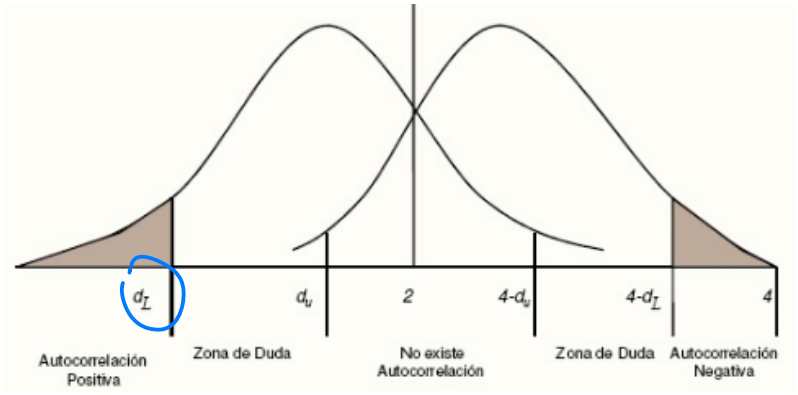
$$DW = \frac{2\sum_{t=2}^{n} e_t^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} = 2\left(\frac{\sum_{t=2}^{n} e_t^2}{\sum_{t=1}^{n} e_t^2} - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right) \simeq 2(1 - \widehat{\rho})$$

donde $\widehat{\rho}$ es la estimación del parámetro ρ.

Como $|\rho| < 1$:

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

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



H DE DURBIN

El contraste de Durbin-Watson es válido cuando la autocorrelación de los errores es autorregresiva de orden 1 y cuando la regresión no incluye entre las variables explicativas algún retardo de la variable dependiente.

En este segundo caso se recurre al contraste h de Durbin. Esto es, en modelos del tipo

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

donde $u_t = \rho u_{t-1} + v_t$, para contrastar

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

utilizaremos el estadístico $h = \rho \cdot \sqrt{\frac{n}{1 - n \cdot \widehat{Var}(\widehat{\alpha})}} \sim N(0,1)$, de forma que se rechazará la hipótesis nula si $|h| > Z_{1-\frac{\alpha}{2}}$.

H de Durbin solo me dice si hay o no autocorrelación, no me da el signo. Aunque podemos intuir por la estimación de "ro" -> p (su signo)

CONTRASTE DE LJUNG-BOX (En presentaciones ejemplo de aplicación directa con fórmula)

Si los residuos son independientes, sus primeras m autocorrelaciones son cero, para cualquier valor de m. El test de Ljung-Box contrasta la hipótesis nula de que las primeras m autocorrelaciones, ρ_m, son cero. Esto es:

$$\left.\begin{array}{l} H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0 \\ H_1 : \exists\, i \in \{1, 2, \ldots, m\} \text{ tal que } \rho_i \neq 0 \end{array}\right\}$$

Se rechaza la hipótesis nula (de incorrelación) si $Q_{LB} = n(n+2)\sum_{s=1}^{m} \frac{r(s)^2}{n-s} > \chi^2_{m-1}(1-\alpha)$ donde

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

es el coeficiente de autocorrelación muestral de orden k.

Si las observaciones son independientes (incorrelación), los coeficientes $r(s)$ serán próximos a cero, por lo que no se rechazaría la hipótesis nula.

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

ESTIMACIÓN BAJO AUTOCORRELACIÓN (MCG)

$$y = X\beta + u$$

con:

- $\mathbb{E}[u] = 0$,
- $\mathbb{E}[u_i u_j] \neq 0$ ($i \neq j$ con $i, j \in \{1, \ldots, n\}$) y
- $\text{var}[u] = \sigma^2 \Omega$.

(no se verifican los supuestos de MCO)
Por el método de Mínimos Cuadrados Generalizado: necesitamos calcular P tal que $\Omega^{-1} = P^t P$.
Suponemos que u_t viene determinada por un proceso autorregresivo de orden 1, $AR(1)$: $u_t = \rho u_{t-1} + \varepsilon_t$ (ε_t un ruido blanco) y ρ el coeficiente de autocorrelación con $-1 < \rho < 1$.

$$u_t \cdot u_{t-1} = (\rho u_{t-1} + \varepsilon_t) \cdot u_{t-1} = \rho u_{t-1}^2 + \varepsilon_t \cdot u_{t-1} \Rightarrow \mathbb{E}[u_t u_{t-1}] = \rho\sigma^2$$

De la misma forma:

$$\mathbb{E}[u_t u_{t-2}] = \mathbb{E}[(\rho u_{t-1} + \varepsilon_t) u_{t-2}] = \rho\mathbb{E}[u_{t-1}u_{t-2}] = \rho(\rho\sigma^2) = \rho^2\sigma^2$$
$$\mathbb{E}[u_t u_{t-3}] = \mathbb{E}[(\rho u_{t-1} + \varepsilon_t) u_{t-3}] = \rho\mathbb{E}[u_{t-1}u_{t-3}] = \rho(\rho^2\sigma^2) = \rho^3\sigma^2$$

$$\mathbb{E}[u_t u_{t-k}] = \rho^k\sigma^2$$

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

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

Suponemos;

$$y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t$$

y

$$y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + \cdots + \beta_k X_{kt-1} + u_{t-1}$$

Luego:

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

Llamando $y_t^* = y_t - \rho y_{t-1}$ y $X_{it}^* = X_{it} - \rho X_{it-1}$:

$$y_t^* = \beta_1 + \beta_2 X_{1t}^* + \cdots + \beta_k X_{kt}^* + \varepsilon_t$$

donde ahora ε cumple las hipótesis para aplicar MCO.
Nota: Se perdería la primera observación, para ello Prais-Wistem proponen $y_1^* = \sqrt{1-\rho^2}y_1$, $X_{i1}^* = \sqrt{1-\rho^2}X_{i1}$...

1º

2º

MODIFICACIÓN PRAIS-WINSTEN

1. Estimando por MCO y se obtienen los residuos e_t.
2. Se estima $u_t = \rho u_{t-1} + \varepsilon_t$ el valor de ρ empleando los e_t:

$$\widehat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2}$$

3. Se estima por MCO $y^* = X^*\beta + u^*$: e_t^*.
4. Estimamos $\widehat{\rho}$ de la regresión $e_t^* = \widehat{\rho} e_{t-1}^* + \varepsilon_t^*$: $\widehat{\widehat{\rho}} = \frac{\sum_{t=2}^n e_t^* e_{t-1}^*}{\sum_{t=2}^n (e_t^*)^2}$.
5. Repetimos... La estimación de ρ es el valor que estabiliza la secuencia $\widehat{\rho}$, $\widehat{\widehat{\rho}}$, $\widehat{\widehat{\widehat{\rho}}}$, ... (hasta que la diferencia sea $< 10^{-3}$).

Proceso iterativo de Cochrane-Orcutt: Obviamos la primera observación en las transformaciones.

```
sm.GLSAR(y, sm.add_constant(x), rho=rho).iterative_fit(maxiter
```

Esto es mucho más fácil en la práctica que en la teoría (que no hay que saberse), de hecho es simplemente aplicar la fórmula y ya está. (Ejemplo en las diapositivas últimas del tema 6)