AYUDANTIA 3

~~Homo~~

Heteroskedasticity

→
- No relation with unbiased and consistency → tiene que ver con varianza de residuos ($E(\hat{\beta})$, $Var(\hat{\beta})$)
- No relation with fit (R^2) (MLR1 - MLR4)
- Related to simple size — asymptotical analysis

Standar errors (se) to estimate t & F distribution

DEF: constant variance MLR5

$Var(u \mid x_1, x_2, \dots, x_k) = \sigma^2 I_n$ — important to INFERENCE

② MELI (MLR1-MLR5) (best linear unbiased estimator)

OLS is no longer asymptotically eficient but unbiased.

① HYPOTHESIS under Gauss Markow assumptions

$$\sigma^2 = \begin{bmatrix} \sigma^2 & Cov(\varepsilon_1,\varepsilon_2) & \cdots & Cov(\varepsilon_i, \varepsilon_1) \\ Cov(\varepsilon_1,\varepsilon_2) & \sigma^2 & & \vdots \\ \vdots & & \ddots & \\ \vdots & & & \sigma_i^2 \end{bmatrix}$$

$$\sigma^2 = \begin{bmatrix} \sigma_2 & \circ & \circ \\ \circ & \sigma_2 & \circ \\ \circ & \circ & \sigma_2^2 \end{bmatrix}$$ constante.

① TEST

OPTION 1 - BREUSH PAGAN → heterocedasticidad lineal

$H_0: Var(u \mid x_1, x_2, \dots, x_k) = \sigma^2$ (test MLR5) ⟶ $\boxed{H_0: \delta_1 = \delta_2 = \dots = \delta_k = 0}$

$Var(u \mid X)$ — by MLR1 - 4

$E(u^2 \mid x) = \sigma^2 = E(u^2)$

Probaremos si u^2 está relacionado con alguna de las variables independientes

① Regresión lineal de residuos y regresores

$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k +$ errores

② Hipótesis conjunta

$$F = \frac{R^2_{\hat{u}^2}}{(1 - R^2_{\hat{u}^2})} \cdot \frac{(n-k-1)}{k}$$

OPTION 2 - BREUSH GODFREY → correlación serial

(a) Con exogeneidad estricta de variables

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

$\underbrace{H_0: \hat{\rho} = 0}_{MCO}$ $\qquad$ $\underbrace{H_1: \hat{\rho} \neq 0}_{MCO \text{ no } MELI}$

$\boxed{\cdot \text{ test } \hat{t}}$

En este caso me gustaría no rechazar Ho

(b) Ar(q) de orden superior

- $\boxed{u_t} = \rho_1 \boxed{u_{t-1}} + \rho_2 u_{t-2} + \dots +$
- $H_0: \boxed{\rho_1 = \rho_2 = \dots = 0}$
- test ML

$$(T - q) \cdot R^2_{\hat{u}_t} \xrightarrow{q} \boxed{\chi^2_q}$$

② SOLUTIONS

ROBUST INFERENCE

- Estimation in prescence of heteroskedasticity → corregir SE
- Asymptotical Approach ($n \to \infty$)

in prescence of multicollinearity we test RI

RLS

$y_i = \beta_0 + \beta_1 x_i + u_i$ ①

Heteroskedasticity $Var(u_i \mid x_i) = \sigma_i^2$ } σ_i^2 cambia cuando cambia i

Con homocedasticidad $\sigma_i^2 = \sigma^2$

$$Var(\hat{\beta}_1 \mid x_i) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2} = \frac{\sigma^2 \sum_{i=1}^{n}(x_i-\bar{x})^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2} = \frac{\sigma^2}{\underbrace{\sum_{i=1}^{n}(x_i-\bar{x})^2}_{SSR_x}}$$

Con heteroskedasticity

$$Var(\hat{\beta}_1 \mid x_i) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2}{\underbrace{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}_{SSR_x^2}}$$ ②

ROBUST STANDARD ERRORS (WHE SE)

Tomando (1) o (2)

$$Var(\hat{\beta}_1 \mid x_i) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 \hat{u}_i^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}$$ R.L.S

o

$$Var(\hat{\beta}_j \mid x) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{\left[\sum_{i=1}^{n} \hat{r}_{ij}^2\right]^2}$$ R.L.M

$\hat{r}_{ij}^2$ = residuos iesimo de regresar x_i en cada x_j (correlación)

SCR_j^2 — ss de residuos → $SCT_j \cdot (1 - R_j^2)$

$se = \sqrt{Var(\beta_j \mid x)}$

SE ROBUST → $t =$ estimación - valor hipotético / error estandar robusto

solo válidos asintoticamente

Empirical regularities

1. If "normal" estimations are significant, probably in robust estimation to.
2. In general Robust SE > Normal SE ---> In homocedasticity, estimators are minimum (NO SIEMPRE)
3. Robust SE ≈ Normal SE in $\boxed{\text{homoceskdasticity}}$

- Los β no cambian, solo SE

SE are more valid ? Why compute normal SE.

→ If homocedasticity is valid, se ~ N, and t & F have exactly distributions independent of the size sample

→ Robust SE in high sample

→ si no es una aproximación.

AYUDANTIA N°3 - ECONOMETRÍA

Profesor: Juan Urquiza

Ayudante: Valentina Andrade (vandrade@uc.cl)

---

(1) $precio = \beta_0 + \beta_1 \cdot mt + \beta_2 \cdot antiguedad + \beta_3 \cdot piscina + \beta_4 \; chimenea$

TEMA I

Suponga que lo contratan para una pasantía en una corredora de propiedades. Su trabajo consiste en desarrollar un modelo que permita pronosticar el precio de una vivienda (y) en función de sus atributos (x). Usted dispone de una base de datos con 1,000 transacciones, donde se registra el precio de venta de la vivienda (y) (en miles de dólares), la superficie total (x_1) (en metros cuadrados), los años de antigüedad (x_2), y si tiene piscina (x_3) y/o chimenea (x_4), entre otros.

$\log y = x$

	Modelo			
	(1)	(2)	(3)	(4)
	MCO	MCO	MCO	MCO
V. Dependiente	precio	$\hat{u}^2$	log(precio)	log(precio)
x_1 mt2	β_1 0.796 se (0.034)	3.266 (0.989)	0.003 [0.001]	0.003 [0.001]
x_2 antigüedad	β_2 -0.294 (0.114)	5.617 (3.280)	-0.002 [0.001]	-0.001 [0.001]
x_3 piscina	β_3 5.930 (2.656)	-10.814 (76.704)	0.026 [0.012]	0.019 [0.005]
x_4 chimenea	β_4 0.934 (2.153)	4.544 (62.177)	0.005 [0.010]	0.007 [0.005]
x_5 → cercana	β_5			0.247 [0.005]
intercepto	β_0 32.950 (9.408)	198.16 (271.73)	4.617 [0.037]	4.511 [0.017]
N	1000	1000	1000	1000
R^2	0.361	0.014	0.359	0.861
SCR	1136471	9480396	194042	42018

Nota: para los modelos (1) y (2) se reportan entre paréntesis los errores estándar usuales, mientras que para los modelos (3) y (4) se reportan entre corchetes los errores estándar robustos. Además, tenga presente que $\hat{u}$ corresponde a los residuos del modelo (1).

a. Antes de utilizar su modelo para pronosticar, decide considerar la prueba de Breusch-Pagan. Se pide que implemente el contraste, y que luego se refiera a las implicancias del resultado obtenido. Considere un nivel de significancia del 5%.

¿Para qué utilizamos Breush Pagan?

→ Para testear si estamos en presencia de heterocedasticidad

¿Cómo se implementa?

1. Regresión lineal de residuos y regresores

$$\hat{u}^2 = \delta_0 + \delta_1 \cdot m_t + \delta_2\, antiguedad + \delta_3 \cdot piscina + \delta_4 \cdot chimenea + error$$

2. Hipótesis conjunta

① $H_0: \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$ $\quad$ H_1: Al menos 1 es distinto de cero

② Estadístico de contraste para hipótesis conjuntas

$$\hat{F} = \frac{R^2_{\hat{u}^2}}{(1 - R^2_{\hat{u}^2})} \cdot \frac{(n-k-1)}{k}$$

$n = 1000$
$k = 4.$

$$\hat{F} = \left(\frac{0.014}{1 - 0.014}\right) \cdot \frac{(1000-4-1)}{4} = 3.53$$

③ F teórico (¡ver tabla!)

$$F_{(q,\, n-k-1)} = F_{(4,\, 995)} \approx F_{(4,\, 120)} = 2.45.$$

$$\approx F_{(4,\, 1000000)} = 2.37.$$

$n = 1000$
$k = 4$
$q = 4$

④ Análisis

Si comparamos $\hat{F}_{(q,\, n-k-1)}$ con $F_{(q,\, n-k-1)}$ notaremos que $\hat{F} > F$, por lo que se rechaza la hipótesis nula, esto significa que al menos un regresor tiene un efecto sobre los residuos estimados.

¿Cuál es la implicancia de esto?

MCO ya no es MELI, ya que las pruebas de inferencia (t, F) ya no son válidas.

especificación = formalización.

b. A la luz de sus resultados, decide considerar no sólo el uso de errores estándar robustos sino también un [cambio en la especificación.] En particular, recuerda haber escuchado que el uso de logaritmos puede atenuar e incluso eliminar los problemas detectados, y por ende decide estimar el modelo considerando al logaritmo del precio de venta como variable dependiente – ver modelo (3). En este contexto, explique cómo se interpretan los coeficientes que acompañan a las variables explicativas.

¿Qué tipo de modelo es? $b_1 \cdot 100\%$ $\log Y - X$

- mt^2 : por cada mt^2, en promedio, el precio de la vivienda sube un 0.3%, ceteris paribus.
- Antigüedad: por cada año de antigüedad adicional, en promedio, el precio de la vivienda disminuye en 0.2%, ceteris paribus.
- Piscina : en promedio, la diferencia de precio de una vivienda con piscina respecto de una sin piscina es de 2.6% mayor, ceteris paribus.
- Chimenea : en promedio, la diferencia de precio de una vivienda con chimenea respecto de una sin chimenea es de 0.5% ceteri bus paribus.

TIPS

* Efectos tienen dirección y tamaño efecto
* No olvidar que los efectos son en promedio
* Los efectos son marginales
* Los efectos son parciales
* Tienen un nivel de significancia asociado

c. El modelo (4) incorpora a la variable binaria *cercana*, que toma el valor de 1 si la vivienda está ubicada a menos de 3 kilómetros de un centro de empleo y es 0 en caso contrario. Se pide que evalúe formalmente si existe un "premio" por cercanía a centros de empleo. Considere un nivel de significancia de 1%.

$\beta_s > 0$

① Hipótesis

$H_0: \beta_s = 0 \qquad H_1: \beta_s > 0$

② Test de contraste

$$t = \frac{\hat{\beta}_s - 0}{se(\hat{\beta}_s)} = \frac{0.247}{0.005} = 49.4.$$

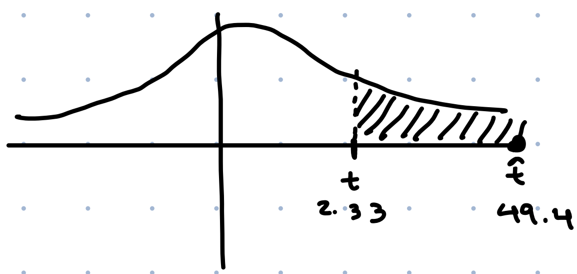
③ Test teórico

$t^{1-\frac{\alpha}{2}}$

$\alpha = 0.01$

$$t^{1-\alpha}_{n-k-1} = t^{1-0.01}_{1000-5-1} = t^{0.99}_{995} \approx t^{0.99}_{1000} = 2.33$$

④ Análisis



- Se rechaza Ho, puedo decir con un 99% de confianza que existe evidencia significativa sobre el efecto de la cercanía sobre el precio de la vivienda, ¡y esta es positiva!

TEMA II

El régimen monetario que se aplica en Chile requiere para su implementación de la proyección de la tasa de inflación a distintos horizontes. En esta pregunta revisaremos modelos simples para pronosticar dicha tasa a partir de datos anuales para el período 1965 a 2019. $n = 56$

Se pide entonces que conteste las siguientes preguntas sobre la base de los resultados presentados a continuación:

$\Omega\ \sigma^2 = \left[\begin{array}{c} \text{correla } \phi \text{ serial} \\ \text{Heterocedas...} \end{array} \right]$

	Modelo				
	(1)	(2)	(3)	(4)	(5)
	MCO	MCO	MCO	MCO	MCO
V. Dependiente	inf_t	$\hat{u}_t$	Δinf_t	$\hat{\varepsilon}_t$	Δinf_t
$desemp_t$	0.50 (0.26)		-0.52 (0.25)	0.02 (0.18)	-0.52 [0.29]
$\hat{u}_{t-1}$		ρ 0.57 (0.11)			
$\hat{\varepsilon}_{t-1}$				-0.03 (0.12)	
intercepto	1.05 (1.55)	-0.11 (0.32)	2.83 (1.23)	0.04 (1.05)	2.83 [1.33]
N	56	55	55	54	55
R^2	0.06	0.35	0.10	0.12	0.10
SCR	476.82	294.72	282.06	197.83	282.06

donde inf_t es la tasa de inflación anual, $desemp_t$ es la tasa de desempleo, $\hat{u}_t$ son los residuos del modelo (1) y $\hat{\varepsilon}_t$ son los residuos del modelo (3).

a. A partir de los residuos del modelo (1), se pide que implemente la prueba t de correlación serial de orden 1 y que explique cómo se interpreta el resultado de dicho contraste. Considere un nivel de significancia del 1%.

Para detectar autocorrelación de los errores utilizamos Breush - Godfrey

→ ¿Qué hacemos con esta prueba?

① Hipótesis

$H_0: \rho = 0 \qquad H_1: \rho \neq 0$

② Test de contraste

$$\hat{t} = \frac{\hat{\rho} - 0}{se(\hat{\rho})} = \frac{0.57}{0.11} = 5.18$$

③ Test teórico $t^{1-\frac{\alpha}{2}}$

$\alpha = 0.01$
$n = 56$
$K = 1$

$$t^{1-\frac{0.01}{2}}_{56-1-1} = t^{0.995}_{54} \approx t^{0.995}_{60} = 2.66$$

④ Análisis

Se rechaza Ho $\longrightarrow$ en presencia de correlación serial de los residuos.

b. El modelo (3) corresponde a una especificación alternativa, donde la variable dependiente ahora es el cambio en la tasa de inflación. Compruebe que este modelo no sufre de problemas de correlación serial mediante la versión *ML* de la prueba de Breusch-Godfrey para correlación serial de orden 1 al 99%. Recuerde especificar claramente las hipótesis nula y alternativa.

① Hipótesis

$$H_0 : \underbrace{\hat{\rho} = 0}_{q=1} \qquad H_1 : \hat{\rho} \neq 0$$

② Test de contraste $q = 1$

$$ML = (T-q) \cdot R^2_{\hat{\varepsilon}}$$
$$= (55-1) \cdot 0.12 = 6.48 \quad \hat{\chi}^2(1)$$

③ Compararemos con $\chi^2(q)$

$$\chi^2(q) = \chi^2(1) = 6.63 \quad \chi^2(1)$$

④ Análisis

No se rechaza Ho con un 99% de confianza

$\Downarrow$

Con 1% de error, no hay correlación serial de los residuos.

c. A pesar del resultado anterior, decide considerar una estimación por MCO con errores robustos de Newey-West – ver modelo (5). Construya un intervalo de confianza bilateral al 95% para β_1, y utilícelo para evaluar la hipótesis nula que $\beta_1 = -1$.

$$\left[\hat{\beta}_1 \pm se(\hat{\beta}_1) \cdot t_{\alpha/2}\right] = \left[-0.52 \pm 0.29 \times 2.00\right]$$

$$= \left[-1.1 \,,\, 0.06\right]$$

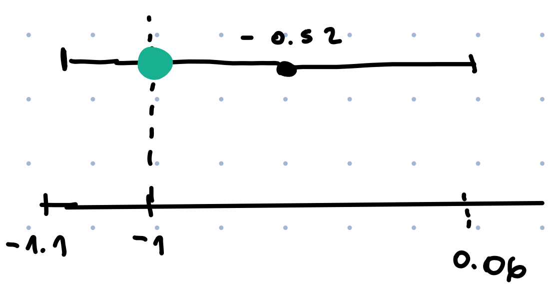
Obtengo de la tabla

$$t^{1-\frac{\alpha}{2}}_{n-k-1} = t^{0.975}_{54-1-1} = 2.00$$

$$-0.52 + 2 \cdot 0.29 = 0.06$$

$$-0.52 - 2 \cdot 0.29 = -1.1$$

¿Se puede rechazar Ho? Usemos un gráfico para tomar intuición



Ho: $\beta_1 = -1$ $\Longrightarrow$ No puedo rechazar Ho.

☆

TEMA III (TAREA 3)

Los macroeconomistas suelen estudiar la relación entre salarios reales y productividad. En esta pregunta revisaremos los datos para la economía estadounidense entre los años 1947 y 1987.

Se pide entonces que conteste a las siguientes preguntas sobre la base de los resultados presentados a continuación:

	Modelo		
	(1)	(2)	(3)
	MCO	MCO	MCO
V. Dependiente	$log(wage_t)$	$\Delta log(wage_t)$	$\hat{u}_t$
$log(outphr_t)$	1.64 (0.09)		
$\Delta log(outphr_t)$		0.73 (0.17)	-0.06 (0.15)
$\Delta log(outphr_{t-1})$		0.46 (0.17)	-0.01 (0.16)
$\hat{u}_{t-1}$			0.36 (0.21)
intercepto	-5.33 (0.37)	-0.01 (0.01)	0.01 (0.01)
N	41	39	38
R^2	0.97	0.49	0.09
SCR	0.031	0.009	0.006

donde $wage_t$ es el salario promedio por hora en el período t, $outphr_t$ es la producción por hora en el período t, y $\hat{u}_t$ son los residuos del modelo (2).

```
Covariance matrix of coefficients

         e(V) | dlog_out~r  dlog_out~1       _cons
--------------+------------------------------------
  dlog_outphr |  .02796329
 dlog_outph~1 | -.00706215   .02742754
        _cons | -.00037817  -.00039393   .00002065
```

a. En base al modelo (1), explique cómo se interpreta β_1 y luego evalúe la hipótesis nula que $\beta_1 = 1$ frente a la alternativa que $\beta_1 > 1$. Considere un nivel de significancia del 1%.

b. Debido a la alta persistencia que exhiben los salarios, un colega le recomienda estimar la regresión en primeras diferencias, y además le sugiere incluir al primer rezago de la tasa de crecimiento de la productividad – es decir, $\Delta log(outphr_{t-1})$ – para permitir efectos rezagados. Compruebe que este modelo (2) no presenta problemas de correlación serial a partir de una prueba de Breusch-Godfrey al 5% de significancia. Recuerde especificar claramente las hipótesis nula y alternativa.

c. En base al modelo (2), evalúe la hipótesis nula que $\beta_1 + \beta_2 = 1$ frente a la alternativa que $\beta_1 + \beta_2 > 1$. Considere un nivel de significancia del 1%.