Final Econometría

February 20, 2026
Javier Ayala

Pregunta 1: 1b

Tenemos que

$$\Omega = \text{diag}(4, 9, 16, 25, 36) = PP', \qquad P = \text{diag}(2, 3, 4, 5, 6).$$

Sea $Var(u|X) = \Sigma = \Omega \otimes I_N$, definimos

$$T = (P^{-1} \otimes I_N) \quad \text{con } P^{-1} = (P^{-1})' \text{ por ser diagonal y } T \text{ es tal que:}$$

$$TT' = (P^{-1} \otimes I_N)(P^{-1} \otimes I_N)' = (P^{-1}P^{-1}) \otimes (I_N I_N) = \Omega^{-1} \otimes I_N = (\Omega \otimes I_N)^{-1} = \Sigma^{-1}$$

Partimos del modelo

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

y transformamos el modelo multiplicando por T a la izquierda:

$$\begin{aligned} TY &= TX\beta + Tu \\ \tilde{Y} &= \tilde{X}\beta + \tilde{u}. \end{aligned}$$

donde

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_{5N} \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{5N} \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ \vdots \\ u_{5N} \end{pmatrix}.$$

Ahora, si usamos OLS sobre el modelo transformado

$$\begin{aligned} \hat{\beta}_{GLS} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} \\ &= ((TX)'TX)^{-1}(TX)'TY \\ &= (X'T'TX)^{-1}X'T'TY \\ &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \end{aligned}$$

O sea, se obtiene $\hat{\beta}_{GLS}$ estimando por OLS el modelo transformado. Notemos también que es insesgado:

$$\begin{aligned} E(\hat{\beta}_{GLS}|X) &= E((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(X\beta + u)|X) \\ &= E(\beta|x) + (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}E(u|X) \\ &= \beta \end{aligned}$$

y con varianza

$$\begin{aligned}
Var(\hat{\beta}_{GLS}|X) &= Var((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(X\beta+u)|X) \\
&= Var(\beta|x) + (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Var(u|X)\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} \\
&= (X'\Sigma^{-1}X)^{-1}
\end{aligned}$$

que por el teorema de Gauss-Markov, es la cota inferior de los estimadores insesgados de β en modelos con heterocedasticidad. Ahora, en la práctica Σ es desconocida; por eso implementamos FGLS:

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

Para calcular $\hat{\Sigma}$, asumiremos que conocemos la estructura de $E(ee'|x)$, o sea, sabemos qué observaciones pertencen al grupo j, pero no sabemos los 5 valores que toma la varianza del error en cada grupo, es decir, estimamos las cinco varianzas del error: $\{\hat{\sigma}_j^2\}_{j=1}^5$ con $\{\frac{1}{N}\sum_{i=1}^N \hat{u}_{i,j}^2\}_{j=1}^5$. Y así definimos el primer estimador de Σ como:

$$\begin{aligned}
\hat{\Sigma}_1 &= \operatorname{diag}\left(\hat{\sigma}_1^2 \ldots, \hat{\sigma}_5^2\right) \otimes I_N \\
&= \operatorname{diag}\left(\frac{1}{N}\sum_{i=1}^N \hat{u}_{i,1}^2, \ldots, \frac{1}{N}\sum_{i=1}^N \hat{u}_{i,5}^2\right) \otimes I_N
\end{aligned}$$

Notemos que

$$\begin{aligned}
\hat{\beta}_{FGLS} &= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y \\
&= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}(X\beta+u) \\
&= \beta + (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}u \\
E(\hat{\beta}_{FGLS} \mid X) &= \beta + E\left[(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}u \,\middle|\, X\right].
\end{aligned}$$

Donde no podemos factorizar $\hat{\Sigma}$ pues depende de Y, y entonces está correlacionado con u. Su varianza es:

$$\begin{aligned}
\hat{\beta}_{FGLS} &= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y \\
&= \beta + (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}u \\
\operatorname{Var}(\hat{\beta}_{FGLS} \mid X) &= \operatorname{Var}\left((X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}u \,\middle|\, X\right).
\end{aligned}$$

Ahora, se tiene la siguiente observación: el estimador

$$\hat{\sigma}_j^2 = \frac{1}{N}\sum_{i\in j} \hat{u}_i^2$$

es un estimador consistente bajo $N \to \infty$, dado que $\hat{u} = MY$ depende de un estimador consistente de β. Sin embargo, en muestras finitas no es insesgado, ya que los residuos OLS están afectados por la estimación de los parámetros y, en particular, $E(\hat{u}_i^2 \mid X) \neq \sigma_j^2$.

Una corrección que ajusta por los grados de libertad efectivos dentro del grupo j está dada por

$$\widehat{\sigma}_j^2 = \frac{\sum_{i\in j} \hat{u}_i^2}{\sum_{i\in j}(1 - h_{ii})},$$

donde h_{ii} son los elementos diagonales de la matriz de proyección $P_x = X(X'X)^{-1}X'$. El denominador $\sum_{i\in j}(1 - h_{ii})$ descuenta la influencia de la estimación de β sobre los residuos dentro del grupo, y constituye una corrección de grados de libertad efectiva que reduce el sesgo finito-muestral del estimador. Sin embargo, por practicidad, seguiremos usando el estimador $\hat{\sigma}_j^2 = \frac{1}{N}\sum_{i\in j} \hat{u}_i^2$. Veamos como se comporta en las simulaciones respecto a $\hat{\beta}_{GLS}$ cuando $5N = 5$

	GLS (Cholesky)		FGLS grupos	
	Media	Desvío	Media	Desvío
β_0	-3.030895	4.850936	-3.045046	5.631405
β_1	0.799237	0.166821	0.799445	0.192225

Table 1: Comparación entre GLS (Cholesky) y FGLS por grupos para $5N = 5$.

Vemos que el promedio de ambos estimadores es muy cercano, ya que ambos tienen la información de la estructura de la varianza del error, sin embargo, GLS es más eficiente que FGLS porque usa los verdaderos valores de las varianzas del error y es cota inferior por el teorema gauss markov, mientras que FGLS los estima con $\hat{\sigma}_j^2$, donde tampoco reescalamos por los grados de libertad, lo que contribuye al aumento de variabilidad de el estimador FGLS.

Pregunta 1: 7

Derivemos analíticamente cada resultado: Sea $T_n = T_n(W_n)$ un estadístico basado en la muestra W_n y sea $c \in \mathbb{R}$ un valor crítico.

Definimos el test como la función

$$\varphi_n((Y,X)_n) = \begin{cases} 0 & \text{si } |T_n((Y,X)_n)| \le c \quad \text{(accept } H_0\text{)}, \\ 1 & \text{si } |T_n((Y,X)_n)| > c \quad \text{(reject } H_0\text{)}. \end{cases}$$

El tamaño (nivel) del test con muestra n se define como

$$\alpha_n = P_{\theta_0}\big(\varphi_n((Y,X)_n) = 1\big) = P_{\theta_0}\big(|T_n((Y,X)_n)| > c\big).$$

Si bajo H_0 se cumple

$$T_n((Y,X)_n) \to^d \xi \sim G(\cdot)$$

fijamos un nivel asintótico α

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} P_{\theta_0}(|T_n((Y,X)_n)| > c) = P(|\xi| > c) = P(\xi > c) + P(\xi < -c) = 2(1 - G(c))$$

y elegimos c tal que

$$2(1 - G(c)) = \alpha$$
$$c = G^{-1}(1 - \alpha/2)$$

En cada simulación $s = 1, \dots, 5000$, generamos una muestra $(Y,X)_n^{(s)}$ y calculamos el estadístico tipo-t asociado a la hipótesis $H_0 : \beta_1 = 0.8$ como

$$T_n^{(s)}((Y,X)_n) = t_n^{(s)} = \frac{\hat{\beta}_1^{(s)}(n) - 0.8}{\widehat{\text{se}}\Big(\hat{\beta}_1^{(s)}(n)\Big)},$$

donde $\hat{\beta}_1^{(s)}(n)$ es la estimación FGLS con muestra de tamaño n de β_1 en la simulación s y $\widehat{\text{se}}(\hat{\beta}_1^{(s)})$ es su error estándar estimado. Por el Teorema del Límite Central tenemos que $c = \Phi(1 - \alpha/2)$, usaremos la distribución normal y no distribución de T, pues estamos usando un estimador robusto de la matriz de covarianzas. Dado el valor crítico c, en cada simulación definimos el indicador de rechazo

$$R_n^{(s)} = \mathbf{1}\{\, |t_n^{(s)}| > c \,\}.$$

de modo que $R_n^{(s)} = 1$ si se rechaza H_0 en la simulación s y $R_n^{(s)} = 0$ en caso contrario. Definiendo $\hat{\alpha}_n$ como sigue, y por ley de grandes numeros, tenemos:

$$\hat{\alpha}_n = \sum_{s=1}^{S} \frac{R_n^{(s)}}{S} \xrightarrow[S\to\infty]{p} E_{\theta_0}(R_n^{(s)}) = E_{\theta_0}(\mathbf{1}\{\, |t_n^{(s)}| > c \,\}) = P_{\theta_0}(|t_n^{(s)}| > c) = \alpha_n$$

Este $\hat{\alpha}_n$ es el que se reporta en la tabla para cada n. Ahora derivemos la potencia: dado el valor critico c, la función poder del test se define como

$$\pi_n(\theta) = P_\theta\big(\varphi_n((Y,X)_n) = 1\big) = P_\theta\big(|T_n((Y,X)_n)| > c\big).$$

para $\theta \neq \theta_0, \theta = 0, 0.4$, Entonces necesitamos generar nuevamente las 5000 simulaciones bajo los θ alternativos, el estadístico que usariamos sería

$$t_{\theta,n}^{(s)} = \frac{\hat{\beta}_1^{(s)}(n) - 0.8}{\widehat{\text{se}}\Big(\hat{\beta}_1^{(s)}(n)\Big)},$$

donde $\hat{\beta}_1^{(s)}(n)$ es la estimación FGLS con muestra de tamaño n de β_1 en la simulación s bajo el parámetro θ. Al igual que antes, dado el valor mismo valor crítico c de antes, en cada simulación definimos el indicador de rechazo

$$R_{\theta,n}^{(s)} = \mathbf{1}\{\,|t_{\theta,n}^{(s)}| > c\,\}.$$

Definiendo $\hat{\pi}(\theta)_n$ como sigue, y por ley de grandes numeros, tenemos:

$$\hat{\pi}(\theta)_n = \sum_{s=1}^{S} \frac{R_{\theta,n}^{(s)}}{S} \xrightarrow[S\to\infty]{p} E_\theta(R_{\theta,n}^{(s)}) = E_\theta(\mathbf{1}\{\,|t_n^{(s)}| > c\,\}) = P_\theta(|t_{\theta,n}^{(s)}| > c) = \pi_n(\theta)$$

Este $\hat{\pi}(\theta)_n$ es el que se reporta en la tabla pada cada n. En la tabla se reporta lo siguiente:

5N	Tamaño		Poder ($\beta_1 = 0$)		Poder ($\beta_1 = 0.4$)	
	1%	5%	1%	5%	1%	5%
5	28.64%	38.30%	96.96%	98.34%	80.82%	87.68%
10	16.28%	25.90%	99.96%	100.00%	95.66%	97.90%
30	4.38%	11.58%	100.00%	100.00%	100.00%	100.00%
100	1.78%	7.06%	100.00%	100.00%	100.00%	100.00%
200	1.40%	6.12%	100.00%	100.00%	100.00%	100.00%
500	1.04%	5.20%	100.00%	100.00%	100.00%	100.00%

Table 2: Tamaño y poder del test bajo FGLS

En conclusión, en muestas pequeñas, $\hat{\alpha}_n$ esta muy por encima del tamaño nominal de la prueba, lo que significa que, condional a que $H_0 : \theta = \theta_0$, el test está arrojando que se rechaza a H_0 muchas veces. Con muestras pequeñas, nuestra estimación de Σ con $V\hat{a}r(u)$ es muy mala, pero se necesita ese insumo para calcular $\hat{\beta}_{FGLS}$. A medida que aumenta n, esta estimación de la varianza mejora y el tamaño converge al tamaño asintótico α El poder particulamente es alto para $\beta_1 = 0$, pero no lo es tanto para $\beta_1 = 0.4$ en muestras pequeñas. Esto se debe a que 0.8 está mas cerca al 0.4 que de 0. Sin embargo, cuando se aumenta la muestra, la probabilidad de cometer error tipo II, que es que condicional a que H_0 sea falso se acepta, practicamente se anula.

Pregunta 2: 1h

Consideramos el modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \qquad \varepsilon_i = \sqrt{v_i}\, u_i, \quad i = 1, \dots, n,$$

con $E(u_i \mid X) = 0$.

El problema es contrastar:

$$H_0: \ \mathrm{Var}(\varepsilon_i \mid X) = \sigma^2 \quad \text{(homocedasticidad)}$$

contra

$$H_1: \ \mathrm{Var}(\varepsilon_i \mid X) = \sigma_i^2 \quad \text{(heterocedasticidad general).}$$

Bajo H_0, se tiene $E(\varepsilon_i^2 \mid X) = \sigma^2$. El test de White se basa en aproximar la función condicional $E(\varepsilon_i^2 \mid X)$ mediante una expansión cuadrática en los regresores:

$$E(\varepsilon_i^2 \mid X) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{1i}^2 + \gamma_4 x_{2i}^2 + \gamma_5 x_{1i} x_{2i}.$$

Entonces, bajo la hipótesis nula:

$$H_0: \ \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0.$$

Sea $\hat{\varepsilon}_i$ el residuo MCO del modelo original y considere la regresión auxiliar:

$$\hat{\varepsilon}_i^2 = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{1i}^2 + \gamma_4 x_{2i}^2 + \gamma_5 x_{1i} x_{2i} + v_i.$$

El estadístico para el test es $LM = nR^2$, donde R^2 es el de la regresión auxiliar. Bajo H_0 y condiciones regulares,

$$LM \xrightarrow{d} \chi_q^2, \qquad q = 5.$$

La regla de decisión al nivel α es:

$$\text{Rechazar } H_0 \quad \text{si} \quad LM > \chi^2_{q,1-\alpha}.$$

El codigo muestra los siguientes resultados:

n	Tamaño (D0)	Poder (D1)	Poder (D2)
20	0.48 / 4.60 / 9.04	0.60 / 6.00 / 11.38	0.76 / 6.92 / 12.26
60	0.90 / 4.86 / 9.94	5.14 / 14.66 / 23.86	3.44 / 11.20 / 18.06
100	1.08 / 4.70 / 9.78	11.04 / 26.66 / 39.46	5.06 / 15.10 / 22.88
200	0.96 / 4.88 / 9.42	33.44 / 58.92 / 72.06	11.24 / 26.82 / 38.78
400	1.02 / 5.34 / 9.52	79.92 / 93.86 / 97.26	28.48 / 52.40 / 64.68
600	1.24 / 4.90 / 10.28	97.52 / 99.62 / 99.86	47.88 / 71.14 / 80.78

Table 3: Test de White. Cada celda muestra resultados al 1% / 5% / 10%.

El tamaño empírico se aproxima bien al nivel nominal para todos los n, incluyendo $n = 20$, lo que indica que la aproximación χ^2 funciona correctamente para el tamaño asintótico elegido. En contraste, la potencia es muy baja en muestras pequeñas: con

$n = 20$, la tasa de rechazo bajo H_1 ($\approx$ 6%) es prácticamente igual al tamaño ($\approx$ 5%), indicando que el test no distingue entre H_0 y H_1. Esto se debe a que con $n = 20$ y $q = 5$ regresores en la auxiliar, el R^2 tiene alta variabilidad y $W = nR^2$ no alcanza los valores críticos. A medida que n crece, $R^2 \xrightarrow{p} R^2_\infty > 0$ bajo H_1, entonces $W = nR^2 \to \infty$ y la potencia converge a 1 (consistencia). Bajo el diseño 2 (t_5), la potencia es sustancialmente menor que bajo el diseño 1 (normal) para todos los n. Esto se debe a que la mayor curtosis de la t_5 infla $\text{Var}(\varepsilon^2)$, reduciendo la precisión de la regresión auxiliar y por lo tanto el R^2. Si bien el test es consistente bajo ambos diseños, la convergencia es más lenta bajo no-normalidad y más aun, todavia hay una probabilidad significativa de cometer error Tipo 2 aun con $n = 600$ en el diseño 2, por ejemplo, con 5% de tamaño, la probabilidad de cometer error tipo II es de casi 28%

Pregunta 2: 2h

El modelo de regresión es:

$$Y = X\beta + e$$

donde

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \qquad e = \begin{pmatrix} \sqrt{\nu_1}\, u_1 \\ \vdots \\ \sqrt{\nu_n}\, u_n \end{pmatrix}.$$

Y potencialmente tenemos heterocedasticidad:

$$E(e|X) = E(\sqrt{\nu}\, u|X) = E(\sqrt{\nu}|X)E(u|X) = 0$$

$$\begin{aligned} \text{Var}(e|X) = E(ee'|X) &= diag\left(E(v_1(x)u_1^2|X), ..., E(v_n(x)u_n^2|X)\right) \\ &= diag\left(v_1(x)E(u_1^2|X), ..., v_n(x)E(u_n^2|X)\right) \end{aligned}$$

Esta expresión va a depender en que diseños nos encontremos

- **Diseño 0:** $u_i \sim \mathcal{N}(0,1)$ y $\nu_i = 1$ (normalidad y homocedasticidad)
- **Diseño 1:** $u_i \sim \mathcal{N}(0,1)$ y $\nu_i = e^{0.25x_{1i}+0.25x_{2i}}$ (normalidad y heterocedasticidad)
- **Diseño 2:** $u_i \sim t_5$ y $\nu_i = e^{0.25x_{1i}+0.25x_{2i}}$ (no normalidad y heterocedasticidad)

Las expresiones para la varianza quedarían como:

$$\begin{aligned} Var^{D0}(e|X) &= diag(1, ..., 1) \\ Var^{D1}(e|X) &= diag(e^{0.25x_{11}+0.25x_{21}}, ..., e^{0.25x_{1n}+0.25x_{2n}}) \\ Var^{D2}(e|X) &= diag(\frac{5}{5-2}e^{0.25x_{11}+0.25x_{21}}, ..., \frac{5}{5-2}e^{0.25x_{1n}+0.25x_{2n}}) \\ &= \frac{5}{3}diag(e^{0.25x_{11}+0.25x_{21}}, ..., e^{0.25x_{1n}+0.25x_{2n}}) \end{aligned}$$

la estimación OLS nos da

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

y su varianza es

$$\begin{aligned}
Var^{Dh}(\hat{\beta}_{OLS}|X) &= Var((X'X)^{-1}X'y|X) \\
&= Var((X'X)^{-1}X'(X\beta + e)|X) \\
&= Var((X'X)^{-1}X'X\beta|X) + Var((X'X)^{-1}X'e|X) \\
&= Var(\beta|X) + (X'X)^{-1}X'Var(e|X)X(X'X)^{-1} \\
&= (X'X)^{-1}X'\Omega^{Dh}X(X'X)^{-1}
\end{aligned}$$

donde $\Omega^{Dh} = Var^{Dh}(e|X)$. donde $h = 0, 1, 2$ es la especificación del diseño. Notemos que Ω^{Dh} no depende de la simulación, ya que la fuente de aleatoriedad está en los diseños, es decir, en u. Dado un diseño h, nosotros estimaremos $Var^{Dh}(\hat{\beta}_{OLS}|X)$ en cada simulación s con:

$$\hat{Var}^{Dh,(s)}(\hat{\beta}_{OLS}|X) = (X'X)^{-1}X'\hat{\Omega}^{Dh,(s)}X(X'X)^{-1}$$

donde $\hat{\Omega}^{Dh,(s)} = diag((\hat{u}_1^{Dh,(s)})^2, ..., (\hat{u}_n^{Dh,(s)})^2)$ con $\hat{u}^{Dh,(s)} = Y^{Dh,(s)} - X\hat{\beta}_{OLS}^{Dh,(s)}$ siendo los residuos de la regresion OLS en la simulación s bajo el disen1o f. Si calculemos el sesgo simple

$$\begin{aligned}
B^{Dh,(s)} &= diag(\hat{Var}^{Dh,(s)}(\hat{\beta}_{OLS}|X) - Var(\hat{\beta}_{OLS}|X)) \\
&= diag\left((X'X)^{-1}X'\hat{\Omega}^{Dh,(s)}X(X'X)^{-1} - (X'X)^{-1}X'\Omega^{Dh}X(X'X)^{-1}\right) \\
&= diag\left((X'X)^{-1}X'(\hat{\Omega}^{Dh,(s)} - \Omega^{Dh})X(X'X)^{-1}\right)
\end{aligned}$$

y el sesgo relativo se define, para $k = 0, 1, 2$, como:

$$b_k = \frac{1}{5000}\frac{\sum_{s=1}^{5000} B_{kk}^{Dh,(s)}}{Var^{Dh}(\hat{\beta}_{OLS}|X)_{kk}}$$

donde $B_{kk}^{Dh,(s)}$ y $Var^{Dh}(\hat{\beta}_{OLS}|X)_{kk}$ son los k-ésimos elementos de la diagonales de $B^{Dh,(s)}$ y $Var^{Dh}(\hat{\beta}_{OLS}|X)$. El código reporta lo siguiente:

	Diseño 1				**Diseño 2**			
n	b_0	b_1	b_2	$\sum\lvert b_j\rvert$	b_0	b_1	b_2	$\sum\lvert b_j\rvert$
20	-0.1592	-0.2099	-0.2123	0.5813	-0.1546	-0.2071	-0.2226	0.5843
60	-0.0511	-0.0677	-0.0644	0.1831	-0.0614	-0.0730	-0.0790	0.2133
100	-0.0317	-0.0430	-0.0407	0.1154	-0.0364	-0.0443	-0.0466	0.1273
200	-0.0160	-0.0204	-0.0204	0.0568	-0.0182	-0.0204	-0.0250	0.0636
400	-0.0070	-0.0093	-0.0088	0.0251	-0.0094	-0.0093	-0.0140	0.0327
600	-0.0048	-0.0055	-0.0067	0.0171	-0.0049	-0.0036	-0.0088	0.0173

Table 4: Sesgo relativo del estimador de varianza de White (residuos OLS).

La Tabla X reporta el sesgo relativo del estimador de varianza de White cuando se utilizan los residuos OLS para construir $\hat{\Omega}^{Dh,(s)}$. Los sesgos b_0, b_1 y b_2 son negativos en todos los casos, para ambos diseños y todos los n. Esto indica que el estimador de White subestima sistemáticamente la varianza verdadera de $\hat{\beta}_{OLS}$. La intuición es que los residuos OLS $\hat{u}_i$ no son los errores verdaderos ε_i, sino que resultan de proyectar Y sobre el espacio columna de X. Esta proyección "achica" los residuos respecto de los errores verdaderos, de modo que $\hat{u}_i^2$ tiende a ser menor que ε_i^2, y en consecuencia $\hat{\Omega}$ subestima a Ω^{Dh}. En muestras pequeñas este efecto es pronunciado porque la proyección consume una fracción importante de los grados de libertad disponibles: con $n = 20$ y $k = 3$ parámetros, el sesgo total supera el 58% en ambos diseños.

A medida que n crece, la fracción de grados de libertad consumida por la proyección se vuelve despreciable y los residuos se aproximan cada vez más a los errores verdaderos, por lo que el sesgo decrece. Esto se observa claramente en la tabla: el sesgo total pasa de 0.58 en $n = 20$ a 0.017 en $n = 600$ para el Diseño 1. El Diseño 2 (t_5) presenta un sesgo levemente mayor que el Diseño 1 en muestras pequeñas, ya que la mayor variabilidad de la t_5 hace que los residuos individuales sean más volátiles y, por lo tanto, peores aproximaciones de los errores verdaderos. Sin embargo, esta diferencia se vuelve insignificante para n grande.

Finalmente, cuando se reemplaza $\hat{u}_i^2$ por los errores verdaderos ε_i^2, el sesgo cae a valores cercanos a cero para todo n, confirmando que el problema no está en el estimador de White en sí, sino exclusivamente en el uso de residuos en lugar de errores verdaderos. En términos prácticos, para $n \geq 200$ el sesgo total se ubica por debajo del 6%, nivel en que el estimador de White resulta una corrección confiable para la inferencia en presencia de heterocedasticidad.

Pregunta 2: 2g

Usando los errores verdaderos, dado un diseño h, nosotros estimaremos $Var^{Dh}(\hat{\beta}_{OLS}|X)$ en cada simulación s con:

$$\hat{Var}^{Dh,(s)}(\hat{\beta}_{OLS}|X) = (X'X)^{-1}X'\hat{\Omega}^{Dh,(s)}X(X'X)^{-1}$$

donde $\hat{\Omega}^{Dh,(s)} = diag((e_1^{Dh,(s)})^2, ..., (e_n^{Dh,(s)})^2)$ donde, para $i = 1, ..., n$, en una simulación tenemos

$$e^{Dh,(s)} = \sqrt{v_i^{Dh}}\, u_i^{Dh,(s)},$$

$$(u_i^{Dh,(s)}, v_i^{Dh}) = \begin{cases} \left(u_i^{(s)} \sim \mathcal{N}(0,1),\ v_i = 1\right), & \text{si } h = 0, \\ \left(u_i^{(s)} \sim \mathcal{N}(0,1),\ v_i = \exp(0.25\, x_{1i} + 0.25\, x_{2i})\right), & \text{si } h = 1, \\ \left(u_i^{(s)} \sim t_5,\ v_i = \exp(0.25\, x_{1i} + 0.25\, x_{2i})\right), & \text{si } h = 2. \end{cases}$$

y definimos b_k como lo hicimos antes. EL código reporta la siguiente tabla:

El sesgo relativo cuando se reemplazan los residuos OLS por los errores verdaderos ε_i^2 en la construcción de $\hat{\Omega}$. Los resultados son contundentes: el sesgo total cae a valores cercanos a cero en todos los n y ambos diseños, sin mostrar una tendencia sistemática con el tamaño de muestra. En el Diseño 1, los sesgos individuales oscilan entre -0.002 y 0.014 y no presentan un signo sistemático, lo que indica que son puramente ruido de Monte Carlo. En el Diseño 2 los valores son levemente mayores en magnitud (sesgo total

n	Diseño 1				Diseño 2			
	b_0	b_1	b_2	$\sum \lvert b_j \rvert$	b_0	b_1	b_2	$\sum \lvert b_j \rvert$
20	0.0018	-0.0008	0.0144	0.0169	-0.0018	-0.0109	-0.0158	0.0286
60	0.0006	0.0011	0.0069	0.0086	-0.0097	-0.0067	-0.0086	0.0250
100	-0.0006	-0.0015	0.0031	0.0052	-0.0050	-0.0039	-0.0036	0.0125
200	-0.0003	0.0004	0.0014	0.0021	-0.0025	-0.0006	-0.0042	0.0073
400	0.0008	0.0008	0.0019	0.0034	-0.0017	0.0005	-0.0035	0.0057
600	0.0004	0.0013	0.0005	0.0023	0.0003	0.0030	-0.0019	0.0052

Table 5: Sesgo relativo del estimador de White utilizando errores verdaderos.

de 0.029 para $n = 20$), lo cual es esperable dado que la t_5 tiene colas más pesadas que la normal, aumentando la variabilidad de ε_i^2 como estimador de σ_i^2 y generando mayor ruido de simulación. Sin embargo, incluso en este caso los sesgos son de una magnitud completamente distinta a los observados con residuos OLS.

Comparando con la Tabla 4 de la pregunta (2.2h), la reducción del sesgo total es drástica: de 0.58 a 0.017 en el Diseño 1 y de 0.58 a 0.029 en el Diseño 2, para $n = 20$. Esto confirma que el sesgo observado en los puntos (a)–(f) no es una deficiencia del estimador de White en sí, sino que se origina exclusivamente en el uso de residuos en lugar de errores verdaderos. Dado que en la práctica los errores verdaderos son inobservables, este ejercicio sirve como experimento de control que aísla la fuente del problema de muestra finita.