

Taller #5
Econometría 06169
Grupo 5

Profesor: Julio César Alonso

Nota: Este taller debe ser entregado en papel y escrito en computador. No se revisarán trabajos escritos a mano.

1. En la pagina Web encontrará una base de datos que contiene la siguiente información para los Estados Unidos: el estado (State), 4 variables dummy que toman el valor de uno si el estado está ubicado en el este (E), en el norte (N), en el sur (S) o en el oeste (W), respectivamente y cero en caso contrario, el total de gasto público en el estado medido en millones de dólares (EXP), la ayuda federal en dólares (AID), el ingreso personal en dólares (INC) y la población de cada estado en miles de personas (POP).

a) Empleando esta base de datos, estime el siguiente modelo por el método de MCO:

$$EXP_i = \mathbf{b}_1 + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i + \mathbf{e}_i .$$

Presente sus resultados en una tabla y discuta brevemente el significado de cada uno de los coeficientes estimados y otros resultados que crea pertinente.

b) Discuta si existe algún tipo de problema de multicolinealidad en el modelo estimado en la parte a) (Si encuentra algún tipo de problema, no lo solucione),

c) Grafique los residuos estimados versus POP, AID y INC (haga dos gráficos por separado). Reporte su gráfico y discuta lo que observa. Recuerde que cada grafico necesita tener su respectivo título.

2. Empleando la misma información del punto anterior, responda las siguientes preguntas:

a) Efectúe el test de Goldfeld y Quandt. Escriba claramente los pasos seguidos y cual es su decisión. (Incluya el valor del estadístico de Goldfeld y Quandt y su significancia en la tabla que construyó anteriormente)

b) Efectúe el test de Breusch-Pagan. Escriba claramente los pasos seguidos y cual es su decisión. (Incluya el valor del estadístico de Breusch-Pagan y su significancia en la tabla que construyó anteriormente)

c) Efectúe el test de White. Escriba claramente los pasos seguidos y cual es su decisión. (Incluya el valor del estadístico de White y su significancia en la tabla que construyó anteriormente)

3. Empleando la misma información del punto anterior, responda las siguientes preguntas:

a) Calcule la matriz de varianzas y Covarianzas de los estimadores MCO por el método sugerido por White (1980). Y discuta la significancia de los coeficientes.

b) Pruebe la Hipótesis nula $\mathbf{b}_2 = \mathbf{b}_3$. Discuta su decisión.

4. Basado en sus resultados y en los gráficos, identifique la variable independiente que está provocando el problema de heteroscedasticidad (Explique brevemente). Ahora corrija el modelo empleando el método de mínimos cuadrados ponderados. Escriba el modelo corregido y muestre que este no posee problemas de heteroscedasticidad. Estime el modelo por el método de mínimos cuadrados ponderados. Reporte sus resultados en la misma tabla que reportó las estimaciones anteriores.

5. Empleando el modelo corregido presentado en el punto 4, escriba y estime un modelo que capture la siguiente hipótesis: “ En promedio, ceteris paribus, el gasto público estatal es diferente en cada una de las 4 regiones geográficas”. Muestre que su modelo si captura la hipótesis. Estímelo y escriba sus resultados en la misma tabla que ha empleado anteriormente para escribir sus resultados.

Taller #5
Respuestas Sugeridas
Econometría 06169
Grupo 5

Profesor: Julio César Alonso

Nota: Este taller debe ser entregado en papel y escrito en computador. No se revisarán trabajos escritos a mano.

1. En la pagina Web encontrará una base de datos que contiene la siguiente información para los Estados Unidos: el estado (State), 4 variables dummy que toman el valor de uno si el estado está ubicado en el este (E), en el norte (N), en el sur (S) o en el oeste (W), respectivamente y cero en caso contrario, el total de gasto público en el estado medido en millones de dólares (EXP), la ayuda federal en dólares (AID), el ingreso personal en dólares (INC) y la población de cada estado en miles de personas (POP).

a) Empleando esta base de datos, estime el siguiente modelo por el método de MCO:

$$EXP_i = b_1 + b_2 POP_i + b_3 AID_i + b_4 INC_i + e_i.$$

Presente sus resultados en una tabla y discuta brevemente el significado de cada uno de los coeficientes estimados y otros resultados que crea pertinente.

El modelo

$$EXP_i = b_1 + b_2 POP_i + b_3 AID_i + b_4 INC_i + e_i \quad (1)$$

fue estimado con EasyReg obteniendo los resultados reportados en la Tabla 1. Noten que:

$\hat{b}_1 = -46.80526$. Implica que de no existir ningún ingreso personal, ni población ni ayuda federal, en promedio el gasto público total en dicho estado sería de US\$-46.8 millones. Obviamente este coeficiente no tiene mucho sentido económico.

$\hat{b}_2 = -0.59662$, un aumento de mil personas en la población del estado provocará en promedio una disminución de US\$ 0,59 millones en el gasto público en dicho estado.

$\hat{b}_3 = 3.23823$, un aumento de un dólar en la ayuda federal a un estado provocará en promedio un aumento de US\$ 3,2 millones en el gasto público total en el estado.

$\hat{b}_4 = 0.00019$, un aumento de un dólar en el ingreso personal en el estado provocará en promedio un aumento en el gasto público en el estado de 190 dólares.

Otros resultados pertinentes, serían el R^2 es muy alto, pues tenemos que la regresión explica un 99.25% de la variabilidad de la variable dependiente. Noten que el F global no tiene sentido si estamos en presencia de Heteroscedasticidad. En este caso tenemos altas sospechas que exista Heteroscedasticidad, pues estamos estudiando datos de corte transversal, así analizar el F global no tiene sentido.

Tabla 1

	VARIABLE DEPENDIENTE: EXP_i			
	Estadísticos t entre paréntesis			
	Ecuación 1	Ecuación 2	Ecuación 3	Ecuación 4
	MCO	MCO	MCP	MCP
constante	-46.8053 (-0.56)	-46.8053 (-0.87)	22.7473 (0.99)	14.8091 (0.50)
POP_i	-0.5966 (-5.71) ***	-0.5966 (-5.34) ***	-0.6791 (-7.22) ***	-0.6915 (-6.39) ***
AID_i	3.2382 (13.64) ***	3.2382 (7.76) ***	2.3688 (10.58) ***	2.4288 (9.10) ***
INC_i	0.0002 (8.12) ***	0.0002 (7.21) ***	0.0002 (11.88) ***	0.0002 (11.18) ***
N_i	--	--	--	20.8827 (0.46)
S_i	--	--	--	-18.6325 (-0.39)
E_i	--	--	--	14.5018 (0.38)
R^2	0.99305	0.99305	0.89110	0.89271
Goldfeld-Quandt	61.76 ***	NA	NA	NA
Breusch-Pagan	44.31 ***	NA	NA	NA
Test de White	28.220 ***	NA	NA	NA
# de Obs.	50	50	50	50

(*) nivel de significancia: 10%

(**) nivel de significancia: 5%

(***) nivel de significancia: 1%

MCO: Mínimos Cuadrados Ordinarios

MCO: Mínimos Cuadrados Ordinarios

NA: No Aplica

- b) Discuta si existe algún tipo de problema de multicolinealidad en el modelo estimado en la parte a) (Si encuentra algún tipo de problema, no lo solucione),

Uno de los síntomas de multicolinealidad está presente, un R^2 alto. Noten que ni el F-global ni los t-calculados son buenos indicadores de la presencia de multicolinealidad, pues como estamos sospechando que exista heteroscedasticidad (por estar empleando datos de corte transversal), entonces la matriz de varianzas y covarianzas de los estimadores MCO estará sesgada y por tanto el F-global y los t-calculados no serán los correctos.

Exploremos ahora la matriz de correlaciones entre las variables independientes. En este

caso tenemos $R = \begin{bmatrix} 1 & 0.951467 & 0.9930415 \\ & 1 & 0.9626688 \\ & & 1 \end{bmatrix}$, con un determinante de 0.00099. Este

determinante es relativamente bajo, evidenciando un problema serio de multicolinealidad.

Finalmente, noten que no hay necesidad de estudiar la matriz de varianzas y covarianzas de los coeficientes estimados, pues como esperamos que exista un problema de heteroscedasticidad, esta matriz será sesgada y por tanto la información que derivemos de ella no es confiable.

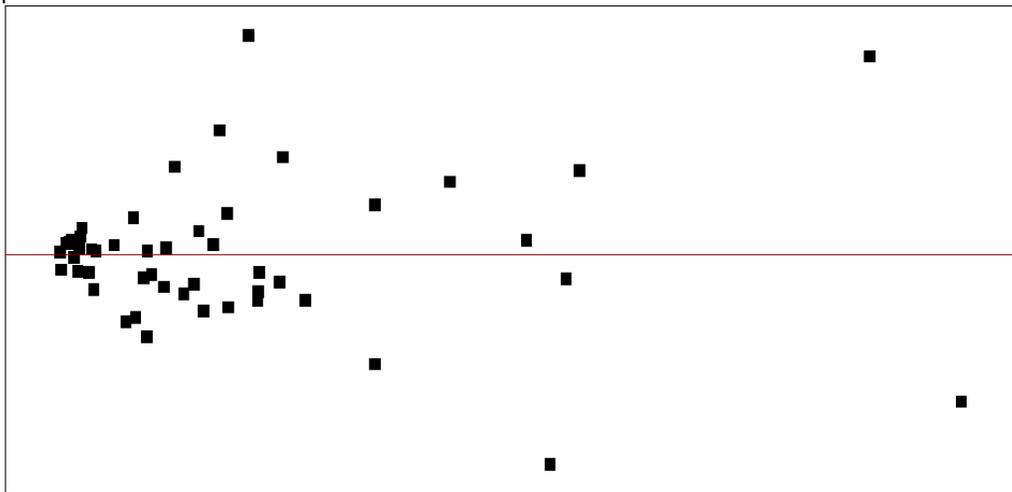
- c) Grafique los residuos estimados versus POP, AID y INC (haga dos gráficos por separado). Reporte su gráfico y discuta lo que observa. Recuerde que cada gráfico necesita tener su respectivo título.

A continuación se reportar los gráficos de los residuos estimados versus la población, la ayuda del gobierno federal y el ingreso personal.

Gráfico 1 Residuos Estimados versus la Población.

Y=DLS Residual of EXP (t=1->50)

Y



X=POP (t=1->50)

Gráfico 2 Residuos Estimados versus ayuda del gobierno federal

Y=OLS Residual of EXP (t=1->50)

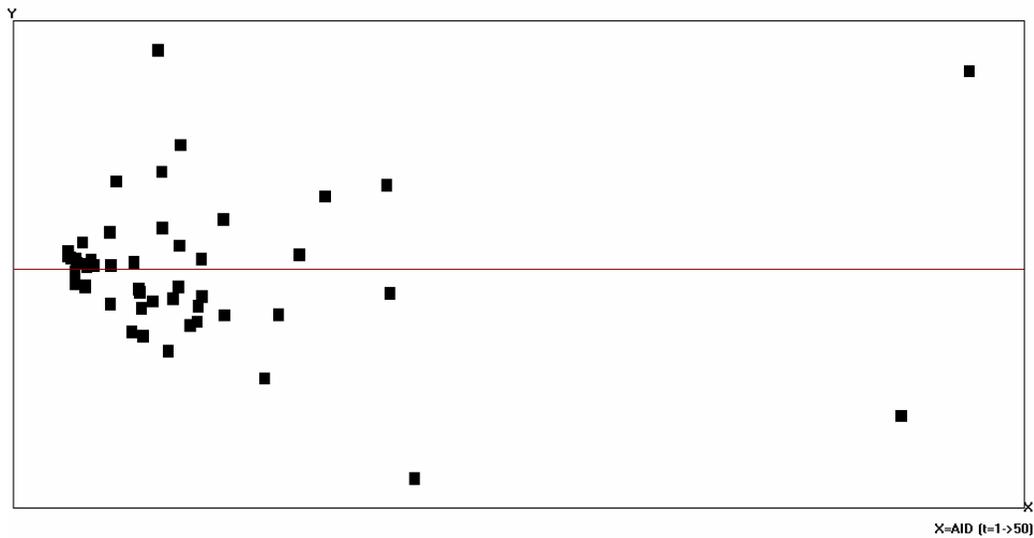
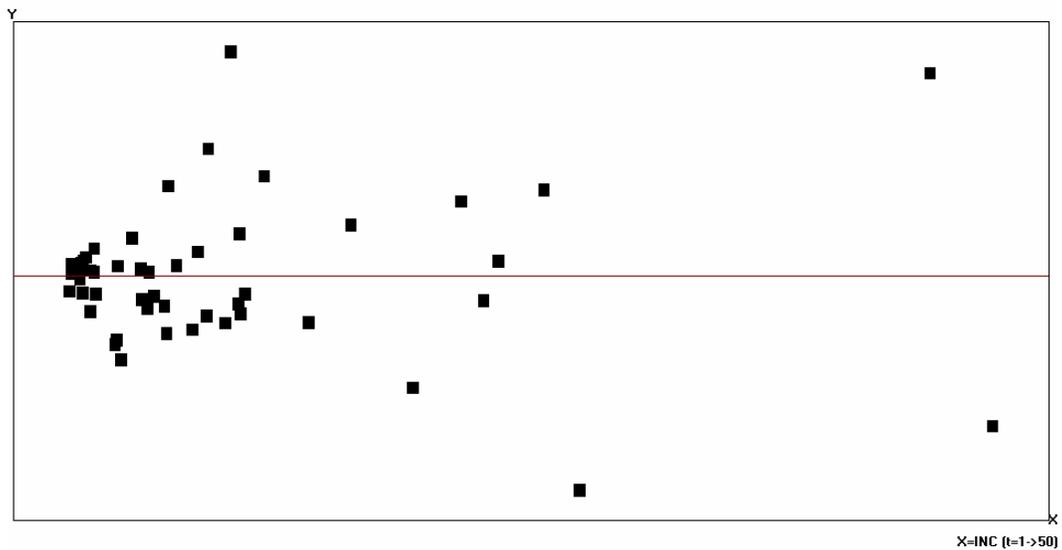


Gráfico 3 Residuos Estimados versus el ingreso personal.

Y=OLS Residual of EXP (t=1->50)



Noten que los tres gráficos muestran que los errores tienden a tener mas variabilidad cuando la respectiva variable aumenta. Los tres gráficos son bastante parecidos. Pero existe una diferencia muy sutil en el gráfico de la población. Este gráfico puede mostrar un poco de mas dispersión que los otros 3 gráficos. Lo que si es claro es que existe un problema de heteroscedasticidad.

2. Empleando la misma información del punto anterior, responda las siguientes preguntas:

- a) Efectúe el test de Goldfeld y Quandt. Escriba claramente los pasos seguidos y cual es su decisión. (Incluya el valor del estadístico de Goldfeld y Quandt y su significancia en la tabla que construyó anteriormente)

Recuerde que este test implica ordenar los datos de mayor a menor de acuerdo a la variable que se crea que afecte más a la varianza del error. En este caso no es claro cual de las variables afecta más a la variabilidad del error. Pero como vimos anteriormente, existe una fuerte correlación (mayor a 0.9 entre los posibles pares de las tres variables independientes del modelo. Así uno esperaría, para nuestro caso, que el resultado del test de Goldfeld y Quandt sea más o menos el mismo sin importar cual variable sea empleada para ordenar los datos. Así empleemos la población como la variable que más afecta la varianza del error.

(Ustedes podían hacer este test ordenando los datos de acuerdo a cualquiera de las tres variables explicatorias. De hecho un análisis completo debería hacer el test ordenando los datos de acuerdo a las tres variables, y calculando así tres diferentes estadísticos. Esto se reporta al final)

Los pasos para efectuar este test son, en este caso, los siguientes:

Ordenar los datos de menor a mayor de acuerdo a la variable independiente POP_i

Omita los $d=8$ datos de la mitad ($d < \frac{1}{5}50 = 10$)

Corra dos regresiones, una con los datos asociados a los valores bajos de I_i (esta será la regresión 1 que se correrá con los datos $i = 1, 2, \dots, 21$) y otra regresión con los datos asociados con los valores altos de I_i (esta será la regresión 2 que se correrá con los datos $i = 30, 31, \dots, 50$)

Encuentre $SSE_1 = 88308.84$ y $SSE_2 = 5453827.093231$

$$\text{Calcule } F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{5453827.093231}{88308.841294} = 61.76$$

$$\text{Compare } F_{GQ} \text{ con } F_{(n-d-2k, n-d-2k)} = F_{(34, 34), 0.01} = 2.258.$$

Claramente, podemos rechazar la hipótesis nula de no heteroscedasticidad. Si ordenamos los datos de acuerdo a la variable AID_i el estadístico de Goldfeld y Quandt es

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{3993465.447147}{159005.291025} = 25.12. \text{ Si ordenamos los datos de acuerdo a la}$$

variable INC_i el estadístico de Goldfeld y Quandt es

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{5889660.299909}{104204.947542} = 56.52. \text{ Como se puede observar y lo esperábamos, en}$$

los tres casos rechazamos la hipótesis nula de la existencia de homoscedasticidad.

b) Efectúe el test de Breusch-Pagan. Escriba claramente los pasos seguidos y cual es su decisión. (Incluya el valor del estadístico de Breusch-Pagan y su significancia en la tabla que construyó anteriormente)

Noten que si ustedes consideran la siguiente H_A :
 $s_i^2 = f(\mathbf{g} + \mathbf{d}_1 POP_i + \mathbf{d}_2 AID_i + \mathbf{d}_3 INC_i)$, entonces el EasyReg le entrega automáticamente el estadístico de Breusch-Pagan relevante. Es decir BP= 44.307384, el cual es mucho mas grande que los valores críticos para la distribución Chi-cuadrado con 3 grados de libertad. Por tanto hay suficiente evidencia para rechazar la hipótesis nula de unos errores homoscedásticos.

A continuación encontrarán los pasos si se considera la siguiente hipótesis alterna $H_0 : s_i^2 = f(\mathbf{g} + \mathbf{d} POP_i)$.

Corra el modelo original $EXP_i = \mathbf{b}_1 + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i + \mathbf{e}_i$ y encuentre la serie de los residuos.

$$\text{Calcule } \hat{S}^2 = \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{n} = \frac{150985.503368}{50} = 3019,71006736.$$

Estime la siguiente regresión auxiliar: $\frac{\hat{\mathbf{e}}_i^2}{\hat{S}^2} = \mathbf{g} + \mathbf{d} POP_i + \mathbf{m}_i$.

Construya los SSR = 223.409320- 151.686550=71.72277 de la última regresión.

$$\text{Calcule } BP = \frac{SSR}{2} = \frac{71.72277}{2} = 35.861385.$$

Compare BP con $c_{1,01}^2 = 6.63$

Claramente, El estadístico de Breusch-Pagan es mayor que el valor crítico. Por tanto existe suficiente evidencia para rechazar la hipótesis nula de no heteroscedasticidad.

Nota: Era posible que ustedes emplearán diferentes hipótesis alternas. Lo importante es que ustedes argumentarán bien la razón por la cual emplean esa hipótesis alterna.

c) Efectúe el test de White. Escriba claramente los pasos seguidos y cual es su decisión. (Incluya el valor del estadístico de White y su significancia en la tabla que construyó anteriormente)

En este caso el test de White implica los siguientes pasos:

Corra el modelo original y encuentre la serie de los residuos.

Ahora corra la siguiente regresión auxiliar:

$$\begin{aligned}\hat{\mathbf{e}}_i^2 &= \mathbf{g} + \sum_{m=1}^k \sum_{j=1}^k \mathbf{d}_s X_{mi} X_{ji} + \mathbf{m}_i \\ &= \mathbf{g} + \mathbf{d}_1 POP_i + \mathbf{d}_2 AID_i + \mathbf{d}_3 INC_i + \mathbf{d}_4 POP_i^2 + \mathbf{d}_5 AID_i^2 \\ &\quad + \mathbf{d}_6 INC_i^2 + \mathbf{d}_7 (POP_i AID_i) + \mathbf{d}_8 (POP_i INC_i) + \mathbf{d}_9 (INC_i AID_i) + \mathbf{m}_i\end{aligned}$$

Calcule el R^2 de la última regresión, en este caso $R^2 = 0.564404$

Calcule $W_a = nR^2 = 28.2202$

Compare W_a con $c_{9,0.01}^2 = 21.66604759$

Claramente podemos rechazar la hipótesis nula de la no presencia de heteroscedasticidad. Así podemos concluir que existe Heteroscedasticidad.

3. Empleando la misma información del punto anterior, responda las siguientes preguntas:

a) Calcule la matriz de varianzas y Covarianzas de los estimadores MCO por el método sugerido por White (1980). Y discuta la significancia de los coeficientes.

Como lo habíamos discutido, esta matriz es calculada automáticamente por el EasyReg. En este caso tenemos que:

$$EstVar \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{b}} \end{bmatrix} = EstVar \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{b}}_3 \\ \hat{\mathbf{b}}_4 \end{bmatrix} = \begin{bmatrix} 145624.176 & -40.2266 & -614.124 & 0.0257 \\ & 0.6238 & -0.406 & -0.0001 \\ & & 8.699 & -0.0001 \\ & & & 0.0000 \end{bmatrix}$$

En la Tabla 1, bajo la denominación de Ecuación 2 se muestran los t-calculados teniendo en cuenta la corrección de White por Heteroscedasticidad.

Como se ve en la Tabla 1, todos los coeficientes asociados con las pendientes son significativos, mientras que el coeficiente del intercepto no es significativamente diferente de cero. Pero ha que tener mucho cuidado con esta conclusión, pues como vimos existe un serio problema de multicolinealidad que puede estar inflando los t-calculados. Lo cual nos puede estar llevando erróneamente a concluir que dichos coeficientes son diferentes de cero cuando en realidad no lo son.

b) Pruebe la Hipótesis nula $\mathbf{b}_2 = \mathbf{b}_3$. Discuta su decisión.

El estadístico de Wald, teniendo en cuenta la corrección por Heteroscedasticidad, para probar la hipótesis nula $\mathbf{b}_2 = \mathbf{b}_3$ ($\mathbf{b}_2 - \mathbf{b}_3 = 0$) es, que es mucho más grande que los valores críticos. Así podemos rechazar la hipótesis nula. Es decir el modelo tiene sentido.

4. Basado en sus resultados y en los gráficos, identifique la variable independiente que está provocando el problema de heteroscedasticidad (Explique brevemente). Ahora corrija el modelo empleando el método de mínimos cuadrados ponderados. Escriba el modelo corregido y muestre que este no posee problemas de heteroscedasticidad. Estime el modelo por el método de mínimos cuadrados ponderados. Reporte sus resultados en la misma tabla que reportó las estimaciones anteriores.

No es difícil argumentar que la población es la variable que puede estar creando el problema de heteroscedasticidad en los errores, pues tanto AID como INC dependen a su vez de la población. Además el test de Goldfeld y Quandt nos permite asumir que $\mathbf{s}_i^2 = \mathbf{s}^2 POP_i^2$. Así, bajo este supuesto, podemos transformar el modelo original en uno homoscedástico dividiendo a ambos lados del modelo por POP_i . Es decir:

$$EXP_i = \mathbf{b}_1 + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i + \mathbf{e}_i \quad (2)$$

$$\frac{EXP_i}{POP_i} = \mathbf{b}_1 \frac{1}{POP_i} + \mathbf{b}_2 \frac{POP_i}{POP_i} + \mathbf{b}_3 \frac{AID_i}{POP_i} + \mathbf{b}_4 \frac{INC_i}{POP_i} + \frac{\mathbf{e}_i}{POP_i} \quad (3)$$

$$\frac{EXP_i}{POP_i} = \mathbf{b}_1 \frac{1}{POP_i} + \mathbf{b}_2 + \mathbf{b}_3 \frac{AID_i}{POP_i} + \mathbf{b}_4 \frac{INC_i}{POP_i} + \frac{\mathbf{e}_i}{POP_i} \quad (4)$$

$$EXP_i^* = \mathbf{b}_2 + \mathbf{b}_1 z_i + \mathbf{b}_3 w_i + \mathbf{b}_4 g_i + \mathbf{e}_i^* \quad (5)$$

donde $EXP_i^* = \frac{EXP_i}{POP_i}$, $z_i = \frac{1}{POP_i}$, $w_i = \frac{AID_i}{POP_i}$, $g_i = \frac{INC_i}{POP_i}$, y $\mathbf{e}_i^* = \frac{\mathbf{e}_i}{POP_i}$. Noten que el nuevo modelo ya es homoscedástico pues $Var[\mathbf{e}_i^*] = Var\left[\frac{\mathbf{e}_i}{POP_i}\right] = \frac{1}{POP_i^2} Var[\mathbf{e}_i] = \frac{1}{POP_i^2} \mathbf{s}^2 POP_i^2 = \mathbf{s}^2$. Entonces el modelo (5) es homoscedástico y por tanto los estimadores MCO aplicados al modelo (5) serán MELI.

La estimación de este modelo se reporta en la Tabla 1 bajo el nombre de Ecuación 3.

5. Empleando el modelo corregido presentado en el punto 4, escriba y estime un modelo que capture la siguiente hipótesis: "En promedio, ceteris paribus, el gasto público estatal es diferente en cada una de las 4 regiones geográficas". Muestre que su modelo si captura la hipótesis. Estímelo y escriba sus resultados en la misma tabla que ha empleado anteriormente para escribir sus resultados.

Claramente, el siguiente modelo captura la anterior hipótesis.

$$EXP_i = \mathbf{b}_1 + \mathbf{a}_1 N_i + \mathbf{a}_2 S_i + \mathbf{a}_3 E_i + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i + \mathbf{e}_i \quad (6)$$

Para demostrar que este modelo si captura la hipótesis, vemos cual es el valor esperado de este modelo.

$$E[EXP_i] = \mathbf{b}_1 + \mathbf{a}_1 E[N_i] + \mathbf{a}_2 E[S_i] + \mathbf{a}_3 E[E_i] + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i$$

$$E[EXP_i] = \begin{cases} \mathbf{b}_1 + \mathbf{a}_1 + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i & si \ i \notin N \\ \mathbf{b}_1 + \mathbf{a}_2 + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i & si \ i \notin S \\ \mathbf{b}_1 + \mathbf{a}_3 + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i & si \ i \notin E \\ \mathbf{b}_1 + \mathbf{b}_2 POP_i + \mathbf{b}_3 AID_i + \mathbf{b}_4 INC_i & si \ i \notin W \end{cases}$$

Así el modelo si presenta diferentes gastos públicos (todo lo demás constante) en las 4 áreas geográficas. Pero el modelo (6) presenta heteroscedasticidad. Como lo demostramos anteriormente, el problema se soluciona al dividir el modelo por POP. Así el siguiente modelo no tendrá heteroscedasticidad.

$$EXP_i^* = \mathbf{b}_2 + \mathbf{a}_1 N_i^* + \mathbf{a}_2 S_i^* + \mathbf{a}_3 E_i^* + \mathbf{b}_1 z_i + \mathbf{b}_3 w_i + \mathbf{b}_4 g_i + \mathbf{e}_i^* \quad (7)$$

donde $N_i^* = \frac{N_i}{POP_i}$, $S_i^* = \frac{S_i}{POP_i}$ y $E_i^* = \frac{E_i}{POP_i}$. Los resultados de la estimación de este modelo se presentan en la Tabla 1 bajo el nombre Ecuación 4.