

# Reporte Final - Trabajo Empírico

Sebastián Ocampo Palacios

15/12/2020

## Trabajo Empírico

### Síntesis

La última entrega del trabajo realiza ajustes al modelo original de regresión de mínimos cuadrados ordinarios para el gasto público destinado a Áreas Naturales Protegidas. En particular, el proyecto realiza un reajuste el índice étnico para evitar problemas de colinealidad y heterosticidad, descarta variables que no reflejan un cambio en la proporción y agrega otras que impulsan la significancia. El resultado final es un efecto negativo por parte de las variables de diversidad étnica y el porcentaje de población en tercera edad. El modelo sigue una regresión de mínimos cuadrados mal especificada (ante la falta de alternativas viables) que cumple con todos los supuestos para obtener los mejores estimadores linealmente insesgados.

El modelo de mínimos cuadrados multivariado se elige por ofrecer una interpretación sencilla sobre los estimados, pero enfrenta el problema de ser una especificación incorrecta como herramienta de análisis por tratarse de variables de proporción. Esto es, el modelo no permite entender a la relación como una curvatura similar a la encontrada en el modelo de Probit y Logit. De momento, no hay alternativas útiles para resolver este problema.

### Modelo Original de Regresión.

El modelo inicial es una regresión lineal de mínimos cuadrados en donde se evalúa para el porcentaje de la población afrodescendiente e indígena en lugar del índice de diversidad étnica (bajo su definición original de  $ETHNIC = 1 - \sum(raza_i)$ ). La decisión de usar las variables del porcentaje poblacional por separado está fundamentada a que arrojan un efecto mayor de significancia estadística. Esto podría deberse a una mala especificación del índice, situación que sería oportuna corregir ante la posibilidad de colinealidad entre ambas variables.

Además, el modelo contempla —siguiendo la metodología de Alesian y Baqir— la variable de Proporción Ingreso a la Mediana como indicador del ingreso per cápita de cada estado. Esto es, la proporción del ingreso per cápita medio para cada estado respecto de la mediana, permitiendo así entender cuáles están por encima y por debajo del valor que divide en dos partes equivalentes al intervalo de ingreso per cápita. Es importante notar que al considerar esta variable se eliminan las observaciones para dos estados (Tabasco y Coahuila) por falta de observaciones. Con ese motivo, se realiza siempre al último.

La Tabla I contiene el modelo de regresión original en donde estudiamos la regresión de tres variables independientes que podrían explicar la proporción de gasto público destinado a la secretaría de medio ambiente. En la tabla observamos que la variable *Porcentaje de población afrodescendiente* tiene un impacto negativo y estadísticamente significativo (90%) en la proporción de gasto público, inclusive tras la inserción de el resto de variables explicativas. Este resultado contrasta con el intercepto y la población indígena, pues pierden su significancia estadística tras la inserción del resto de variables (situación deseable para el caso del intercepto). Destaca también el nivel de  $R^2=0.3$  en comparación con Adjusted  $R^2 = 0.1$  por sugerir una mala especificación de las variables.

Table 1: Resultados de Regresión Inicial

	<i>Dependent variable:</i>				
	Gasto Ambiental Ajustado al Ingreso Estatal				
	(1)	(2)	(3)	(4)	(5)
Población Afrodescendiente	-0.004 (0.003)	-0.01* (0.003)	-0.01* (0.003)	-0.01* (0.003)	-0.01* (0.003)
Población Indígena		0.001* (0.000)	0.001* (0.000)	0.000 (0.000)	0.000 (0.000)
Porcentaje Población 60+			0.001 (0.004)	0.001 (0.004)	0.000 (0.004)
Distancia Mínima				-0.000 (0.000)	-0.000 (0.000)
Proporción Ingreso a la Media					-0.000 (0.000)
Constant	0.000*** (0.000)	0.000** (0.000)	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)
Observations	30	30	30	30	28
R <sup>2</sup>	0.1	0.2	0.2	0.2	0.3
Adjusted R <sup>2</sup>	0.02	0.1	0.1	0.1	0.1
Residual Std. Error	0.000 (df = 28)	0.000 (df = 27)	0.000 (df = 26)	0.000 (df = 25)	0.000 (df = 22)
F Statistic	1.7 (df = 1; 28)	2.5* (df = 2; 27)	1.6 (df = 3; 26)	1.7 (df = 4; 25)	1.8 (df = 5; 22)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Respecificación del modelo

En la tabla de correlación sobre las variables de Gasto Estatal del reporte intermedio encontramos que no existen correlaciones mayores a 0.7 en las variables no manipuladas (vemos que el índice étnico sí tiene una correlación alta con la variable de población indígena). Esto nos alienta a no preocuparnos por la colinealidad y refuerza la idea de evitar usar el índice étnico y los porcentajes de población simultáneamente. No obstante, decidimos hacer un reajuste al índice étnico (de ahora en adelante índice de diversidad étnica) para asegurarnos que los porcentajes de población afrodescendiente e indígena no mitigen sus efectos mutuamente ( $cor(Af, In) = 0.36$ ). Aún más, el nuevo índice de diversidad étnica nos permite adherirnos a la proposición de Alesina, Baqir y Easterly e incluir al sector de población homogénea que no habíamos incluido anteriormente. Definimos el índice como  $Ethnic = 1 - \sum (raza_i)^2$  con  $i \in (Homogenea, Indigena, Afrodescendiente)$ . La interpretación de los resultados dependientes de este índice, indica que de pasar de una sociedad completamente homogénea ( $Ethnic = 0$ ) a una sociedad completamente diversa ( $Ethnic = 1$ ), habrá un cambio en la proporción del gasto en X puntos porcentuales. Con esto, al realizar un modelo de mínimos cuadrados, obtenemos los siguientes resultados:

Table 2: Resultados de Regresión Inicial

	<i>Dependent variable:</i>			
	Gasto Ambiental Ajustado al Ingreso Estatal			
	(1)	(2)	(3)	(4)
Diversidad Étnica	-0.7*** (0.2)	-0.7*** (0.2)	-0.7*** (0.2)	-0.9*** (0.2)
Aprox. Distancia Mínima		1.0 (2.2)	0.9 (2.2)	0.9 (2.3)
Porcentaje Población 60+			-146.1 (114.9)	-220.2* (113.6)
Proporción Ingreso a la Mediana				-0.1 (0.1)
Constant	0.6*** (0.1)	0.5* (0.2)	0.5** (0.2)	0.7** (0.3)
Observations	30	30	30	28
R <sup>2</sup>	0.3	0.3	0.3	0.5
Adjusted R <sup>2</sup>	0.3	0.2	0.3	0.4
Residual Std. Error	0.1 (df = 28)	0.2 (df = 27)	0.1 (df = 26)	0.1 (df = 23)
F Statistic	11.0*** (df = 1; 28)	5.5** (df = 2; 27)	4.3** (df = 3; 26)	5.3*** (df = 4; 23)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

En primer lugar observamos un aumento significativo en el impacto de la diversidad étnica sobre la proporción del gasto. Específicamente, el transitar a una sociedad enteramente diversa ( $Ethnic = 1$ ) tiene como efecto una reducción de 0.7 unidades porcentuales en la proporción de gasto público en la categoría ambiental. Esta situación se amplifica al incluir el resto de los controles, resultando en nueve puntos porcentuales de disminución ante un cambio en el valor del índice de diversidad étnica con una significancia estadística del 99%. También observamos que la distancia sigue sin tener un impacto significativo en la regresión, pero salta su valor positivo -sugiere, sin significancia estadística, que un cambio positivo en la distancia de 0 a la distancia máxima () aumentará la proporción de gasto. El porcentaje de población 60+ destaca por su valor exorbitante y su significancia estadística de 90%. Aún más, parece ser el caso de un error la forma funcional

del modelo por sugerir un incremento negativo de 220% ante una población compuesta de puros adultos mayores. Si bien se revisará esto en la prueba de forma funcional, establecemos la interpretación como que ante una población de adultos mayores no habría gasto (acotando la reducción al -100%) Finalmente, la proporción del ingreso respecto de la mediana tienen un efecto negativo pero no es estadísticamente significativo.

## Análisis de Supuestos

Para revisar la posibilidad de colinealidad consultamos la tabla de correlaciones corregidas en donde se incluye la variable de Diversidad Étnica. Observamos que ninguna de las variables alcanza un valor de correlación mayor a 0.7, con excepción de las variables que componen el índice de diversidad con el índice y, casi, la relación que es el objetivo de estudio (índice de diversidad-prop. gasto). No nos preocupamos por la colinealidad en consecuencia, pues parece que el índice de diversidad sirvió como corrección.

	Porcentaje Población 60+	Población Autoadscrita Indígena	Población Aut. Afrodescendiente	Aprox. Distancia Mínima	Gasto Ajustado al Ingreso	Proporción Ingreso a la Mediana	Diversidad Étnica
Porcentaje Población 60+	1.00000000	-0.2241635	0.1967105	0.02559111	-0.34029926	-0.3017254	0.1311876
Población Autoadscrita Indígena	-0.22416345	1.00000000	0.3632207	0.26341231	-0.16448862	-0.2724671	0.4091167
Población Aut. Afrodescendiente	0.19671047	0.3632207	1.00000000	0.17006035	-0.47159200	-0.4774368	0.8552102
Aprox. Distancia Mínima	0.02559111	0.2634123	0.1700604	1.00000000	0.05547511	-0.4206667	0.1194644
Gasto Ajustado al Ingreso	-0.34029926	-0.1644886	-0.4715920	0.05547511	1.00000000	0.2464288	-0.6120588
Proporción Ingreso a la Mediana	-0.30172543	-0.2724671	-0.4774368	-0.42066673	0.24642883	1.00000000	-0.5390408
Diversidad Étnica	0.13118759	0.4091167	0.8552102	0.11946438	-0.61205876	-0.5390408	1.00000000

Al realizar la prueba de Breuch-Pagan encontramos un coeficiente de 4.8988 (con 4 grados de libertad asociados a las variables restantes) y un p-value = 0.2978 que no permite descartar la hipótesis nula. El rechazo de la hipótesis nula sugeriría la existencia de una relación entre la varianza de los residuales y las transformaciones de cada valor independiente. Esto es, el modelo de mínimos cuadrados parece no tener problemas de heterosticidad.<sup>1</sup> Al realizar las pruebas de Durbin-Watson y Breusch-Godfret encontramos que parece existir un problema de autocorrelación contemporánea. En el caso de la prueba Durbin-Watson, la función arroja un coeficiente dentro de la zona de rechazo (2.079), pero no es estadísticamente significativo para un p-value de 0.5501. Así también el test de Breusch-Godoffrey obtiene un coeficiente de 0.092627 pero estadísticamente insignificativo p-value = 0.7609. Entonces, no parece que exista un problema de autocorrelación contemporánea estadísticamente significativo para las variables elegidas

Continuamos al analizar la normalidad de los residuales. Para esto, realizamos la prueba Jarque Vera que arroja los valores de 0.35248 para un insignificante p-value de 0.8384. De forma similar, la prueba de Shapiro-Wilk resulta insignificante (0.96748, p-value = 0.515). El rechazo de evidencia sobre un problema de no-normalidad en residuales nos ahorra problemas en la definición y el uso de intervalos de confianza. Finalmente, encontramos que la especificación funcional es errónea. Esto era claro desde el efecto en el gasto por parte de la población mayo a sesenta años. Antes de proceder a resolver este problema, revisamos la prueba que nos aseguró el fatal destino de nuestro modelo: RESET TEST de Ramsey.<sup>2</sup> La prueba RESET arroja un p-value de 0.006255, indicando que con una confianza del 99% hay evidencia de un error en la forma funcional.

## Forma Funcional

El problema de la forma funcional surge en la naturaleza de los datos en el modelo. Particularmente, en su carácter como tasas o proporciones. Hansjörg Plieninger indica en su sitio que al utilizar un modelo de

<sup>1</sup>El modelo también pasa la prueba de términos cruzados de White, siendo BP = 1.9273, df = 2, p-value = 0.3815.

<sup>2</sup>Muy adecuado nombre para una prueba que nos regresa al lugar oscuro del mal manejo econométrico.

proporciones es necesario asegurarse de que las variables se encuentren dentro de niveles intermedios para evitar obtener proporciones fuera del límite de  $[0,1]$ <sup>3</sup>. Esto no ocurre para la variable que mide la proporción del ingreso respecto de la mediana estatal, pues naturalmente replica tasas mayores a uno para valores por encima de la mediana. Como alternativa, Plieninger sugiere el uso de un regresión beta o un modelo quasibinomial de regresión: ambos no salen significativos, probablemente por un mal plantamiento propio.

## Resultados

Los resultados obtenidos sugieren una disminución sobre el gasto ambiental dado un incremento en la diversidad étnica, esto no es sorprendente según el modelo. No obstante, contrasta con la intención de este análisis al descartar la posibilidad de que la localización de las áreas naturales protegidas y su relación con los pueblos indígenas generen un aumento en el gasto destinado a la protección ambiental (por medio de la Secretaría de Ambiente). Estos resultados son importantes por sugerir que el peso de la diversificación étnica equivale a una reducción en el gasto aún en bienes públicos gozados principalmente por sectores particulares de la población.

Los resultados de este trabajo son una particularización de los obtenidos por Alesiana, Baqir e Easterly, que si bien no estudian el gasto ambiental, encuentran una tendencia en la disminución de los bienes públicos aún cuando estos son productivos y beneficios para todos los integrantes de la sociedad. Otro ejemplo es la investigación de Abhijit Banerjee y Rohini Somanathan sobre la provisión histórica de bienes públicos en India, en donde identifican como un efecto de sustitución a la reducción de ciertos bienes públicos en pro de otros<sup>4</sup>. Tampoco toman en consideración el gasto ambiental.

## References

Banerjee, Abhijit, and Rohini Somanathan. "The Political Economy of Public Goods: Some Evidence from India." *Journal of Development Economics* 82, no. 2 (2007): 287–314.

Plieninger, Hansjörg. "Regression Modeling with Proportion Data (Part 1). It Was Simple." Accessed December 16, 2020. <https://hansjoerg.me/2019/05/10/regression-modeling-with-proportion-data-part-1/>.

---

<sup>3</sup>Hansjörg Plieninger, "Regression Modeling with Proportion Data (Part 1). It Was Simple," accessed December 16, 2020, <https://hansjoerg.me/2019/05/10/regression-modeling-with-proportion-data-part-1/>

<sup>4</sup>Abhijit Banerjee and Rohini Somanathan, "The Political Economy of Public Goods: Some Evidence from India," *Journal of Development Economics* 82, no. 2 (2007): 287–314