



INSTITUTO DE ECONOMÍA  
FACULTAD DE CIENCIAS ECONÓMICAS  
Y ADMINISTRATIVAS

## Tarea N°3

### Econometría I

Valentina Andrade, Bernardita Salinas y Catalina Mussuto

martes 07, diciembre 2021

## Resumen

El siguiente reporte tiene por objetivo presentar los análisis realizados en la Tarea N°3 del ramo Econometría I dictado por el profesor Juan Urquiza. Esta tarea consta de dos partes pero que tienen el propósito de evaluar dos estrategias de identificación para abordar la causalidad. La primera parte trabajamos con el artículo ‘The Colonial Origins of Comparative Development’ (Acemoglu, Johnson y Robinson, 2001) para evaluar el rendimiento del uso de variables instrumentales (IV) y Mínimos Cuadrados en dos etapas (2SLS) en la estimación de efecto causal de la calidad de las instituciones sobre el ingreso per cápita. La segunda parte abordamos la famosa “Ley del Etiquetado” en Chile y que ya ha sido evaluado en famosos papers como “Equilibrium effects of food labeling policies” de Barahona, Otero, Otero y Kim (2020). En nuestro caso nos enfocaremos en uso de datos panel y utilizando métodos de Difference-in-Difference, pero usamos una discusión econométrica más amplia para hacer más robustas nuestras conclusiones. En términos sustantivos nos encontramos con que (aquí decir los resultados). El código del reporte, así como los análisis adicionales [se pueden obtener en el siguiente link o en el Apéndice de este informe](#)

## Índice general

|  |           |
|--|-----------|
| <b>Ejercicio 1</b>   | <b>4</b>  |
| Breve introducción . . . . .   | 4         |
| <b>Pregunta 1</b>  | <b>4</b>  |
| <b>A. Explorar la relación entre índice de protección y PIB</b>      | <b>5</b>  |
| B. Estimar la ecuación por MCO . . . . .                             | 6         |
| C. Discusión del sesgo de la estimación MCO . . . . .                | 6         |
| <b>D. Variable omitida de capital humano</b>                         | <b>11</b> |
| <b>E. Dirección de sesgo</b>   | <b>13</b> |
| <b>Pregunta 2</b>  | <b>14</b> |
| <b>A. Condiciones de instrumentos</b>                                | <b>14</b> |
| B. Mínimos Cuadrados en Dos Etapas manualmente . . . . .             | 16        |
| C. Estimación de Mínimos Cuadrados en Dos Etapas - Directo . . . . . | 19        |
| <b>Pregunta 3</b>  | <b>21</b> |
| A. Test de Haussman . . . . .  | 21        |
| B. Interpretación de Test de Haussman . . . . .                      | 22        |
| C Test de Hausman manual . . . . .                                   | 22        |

|  |           |
|--|-----------|
| <b>D. Comparación de VI y MCO</b>                              | <b>23</b> |
| Diferencias entre los estimadores                              | 25        |
| <b>E. Hipótesis de sesgo por ECV variable dependiente</b>      | <b>26</b> |
| <b>F. Hipótesis de sesgo por ECV en variable independiente</b> | <b>26</b> |
| <b>G. Hipótesis por sesgo de variable omitida</b>              | <b>27</b> |
| <b>Ejercicio 2</b>   | <b>29</b> |
| Variables . . . . .  | 29        |
| Descriptivos . . . . .   | 29        |
| <b>Pregunta 1</b>  | <b>33</b> |
| 1.1 Interpretación . . . . .                                   | 33        |
| <b>Pregunta 2</b>  | <b>34</b> |
| <b>Pregunta 3</b>  | <b>35</b> |
| A. Valor esperado de los productos . . . . .                   | 36        |
| B. Estimación del modelo DiD . . . . .                         | 36        |
| <b>Referencias</b>   | <b>41</b> |
| <b>Apéndice</b>  | <b>42</b> |
| Código en STATA Ejercicio 1 . . . . .                          | 42        |
| <b>Código STATA ejercicio 2</b>                                | <b>48</b> |

# Ejercicio 1

## Breve introducción

- Objetivo del paper

En “*The colonial origins of comparative development: An empirical investigation*” de [Acemoglu, D., Johnson, S., & Robinson, J. A. \(2001\)](#) se preguntan **cuáles son las causas de las diferencias en los ingresos per cápita entre los países**, indicando que cada vez se releva más importancia de *determinantes institucionales y derechos de propiedad* sobre la performance económica (*institutions matter literature*).

Su **hipótesis central** es que países con mejores instituciones, más seguridad en derechos de propiedad y menos políticas que distorsionan, se promueve más la inversión en capital físico y humano, y esos factores utilizados eficientemente producen un mayor nivel de ingresos. Además, Acemoglu (1995) años antes muestra que en países donde se refuerzan los derechos de propiedad, se promueve más la inversión en capital humano.

Ahora bien, desde el inicio indican que este puede existir un problema de **endogeneidad**. Es esperable que la **calidad de las instituciones** sea endógena a la performance económica, toda vez que es probable que las economías más ricas puedan conducir de mejor manera mejores instituciones (*efecto simultaneidad*). A su vez, también existen otras características que pueden hacer que los países difieran en sus ingresos per cápita (*variable omitida*) como el clima, dificultad para establecer instituciones por los niveles de violencia, etc. Sin embargo, indican que pese a que solo algunos países fueron colonizados no hay sesgo de selección pues la pregunta de interés es el efecto de las políticas de colonización *condicional a ser colonizado*.

Es en este problema que aparece la importancia de las **variables instrumentales**. Un hecho expresivo que muestran es que **instituciones más extractivas** o **más protectoras** se establecieron **dadas ciertas condiciones en que estaban las colonias**. Particularmente, como mostraremos que la fuente de variación exógena a la cuál recurrirán son las *tasas de mortalidad de los colonos* como un instrumento para las *instituciones actuales*, y a partir de ahí, encuentran grandes efectos de la calidad institucional sobre el ingreso per cápita.

*mortalidad colonos  $\implies$  instalacin instituciones  $\implies$  instituciones actuales  $\implies$  performance econmica actual*

## Pregunta 1

Como primer paso, vamos a estimar el modelo *MCO* que relaciona PBI per cápita de los países con que tan buenos son sus derechos de propiedad en los países<sup>1</sup>.

$$\ln(PIB_{percpita})_i = \beta_o + \beta_1 \cdot avexpr_i + u_i$$

donde definiremos como

---

<sup>1</sup>La notación que siguen los autores es la siguiente pero equivalente a la empleada por nosotras  $\log y_i = \mu + \alpha R_i + X_i' \gamma + \epsilon_i$

- $i$ : subíndice del país
- $\ln(PIB_{percapita})$ : logaritmo del PBI per cápita
- $ave\text{expr}$ : índice de protección contra riesgo de expropiación del país.
- $u$ : término de error

## A. Explorar la relación entre índice de protección y PIB

Antes de mostrar el resultado de la regresión podemos evidenciar que el efecto positivo de la calidad institucional sobre la performance económica planteada por los autores es cierta. De manera más precisa en la *Figura 1* que relaciona el índice de protección contra riesgo de expropiación de los países y el PBI per cápita es positiva, es decir a mayor protección contra el riesgo de expropiación tiende a haber un mayor PBI per cápita.

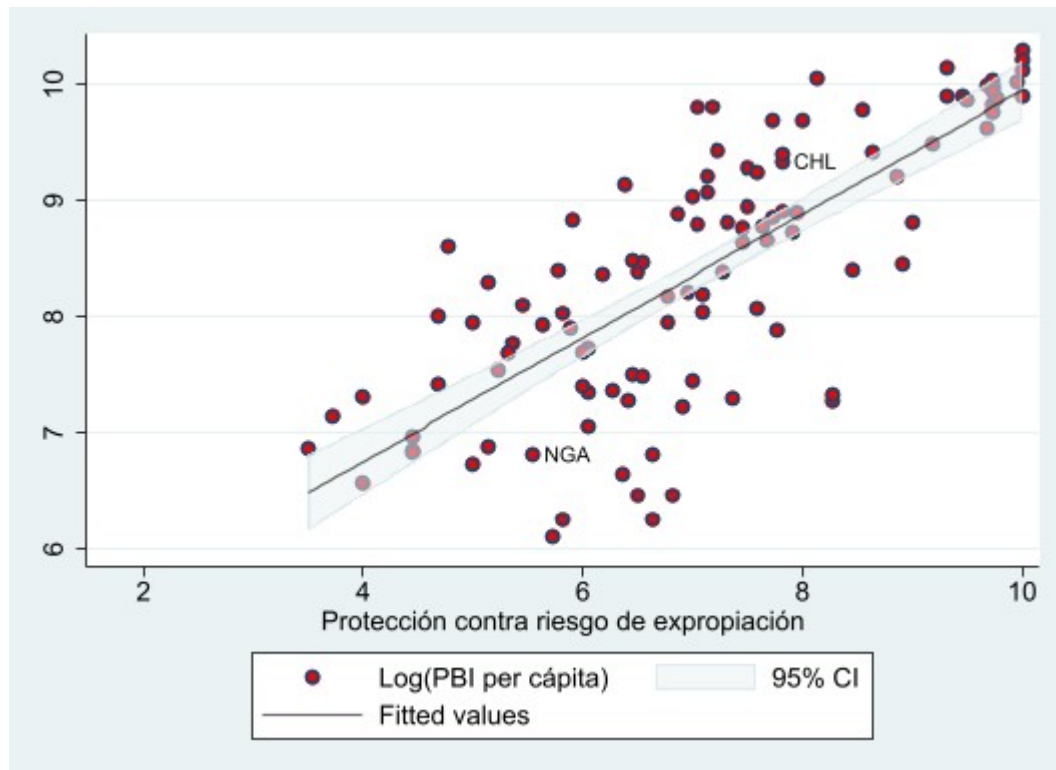


Figura 1: Gráfico de dispersión entre ingresos per cápita y protección a expropiación

Cada uno de los puntos corresponden a países, por lo que podemos discutir la posición que ocupan dos países de interés para los autores: Chile y Nigeria. De manera gráfica se observa que Chile se encuentra en la esquina superior derecha del gráfico y que cuenta con una mayor protección contra el riesgo de expropiación y con un mayor PBI per cápita que Nigeria, que se encuentra en el sector inferior izquierdo del gráfico. ¿Qué significa esto? Nigeria en la muestra está en el percentil 25 más bajo en la medida institucional (5,6 aproximadamente) y Chile está en el percentil 75 de índice (1,8 aproximadamente). La diferencia entre ambos muestra ser en promedio de 1,14 en el logaritmo del ingreso per cápita. Si llevamos esto a números Chile es cerca de 11 veces más rico que Nigeria.

Respecto a la distribución de estos dos casos cabe destacar que en la tendencia de la relación, Chile en términos observados se encuentra por sobre esta (por lo que estaría subestimado su predicción), mientras que Nigeria se encuentra bajo esta lo que podría significar que existen otras variables no observadas en esta relación (de todas formas no hay presencia de outliers hasta ahora).

## B. Estimar la ecuación por MCO

Ahora, en la *Tabla 1*. reportamos la regresión simple que estima el efecto de la protección contra el riesgo de expropiación sobre el logaritmo del PIB per cápita.

**Tabla 1.** Modelos de regresión lineal que estiman log GDP per cápita

| Predictores   | Modelo - MCO        |
|---|---------------------|
| Proteccion contra riesgo de expropiacion                              | 0.532***<br>[0.041] |
| Constant  | 4.626***<br>[0.301] |
| Observations  | 111                 |
| Adjusted R-squared  | 0.608               |
| Note: Run on 27 Nov 2021 using data from input\base_country-codes.dta |                     |
| Standard errors in brackets   |                     |
| *** p<0.01, ** p<0.05, * p<0.1  |                     |

### B.1 Interpretar

En términos generales, los resultados del modelo reportado en la *Tabla N°1* nos permite evidenciar con un 99% de confianza que, manteniendo todo lo demás constante, existe una relación positiva entre la calidad institucional y la performance económica. De manera más específica, el tamaño efecto es muy grande considerando que un aumento de un 1% en el índice de protección contra riesgo de expropiación del país se traducirá en un aumento del 53,2% en el PIB per cápita ( $\beta \cdot 100$ ,  $p < 0.001$ ). Este efecto se corresponde por el reportado en la Tabla N°2 por los autores donde obtienen una estimación de un 54% para todo el mundo y un 52% para una parte de la muestra.

En términos económicos el resultado tiene sentido en la medida en que, tal como indicamos en un inicio existe evidencia de que en los países donde se refuerzan los derechos de propiedad, se da más seguridad a la inversión en capital lo que aumenta la producción (sobre todo si es en capital humano) (Acemoglu, 1995)

## C. Discusión del sesgo de la estimación MCO

### C.1 Sesgo

Si bien los resultados de la *Tabla 1* y la *Figura 1* se muestran una fuerte correlación entre las instituciones y la performance económica, existen una serie de razones enumeraremos que nos

permiten afirmar que no podemos interpretar una relación causal.

1. **Efecto causal en la otra dirección:** es posible plantear que son las economías más ricas las que pueden habilitar aparatos institucionales más robustos o inclusive tener preferencias hacia fortalecer las instituciones para hacer más eficiente la economía.
2. **Variables omitidas:** existen una serie de determinantes a las diferencias por ingreso que naturalmente podrían estar correlacionados con las instituciones. Uno de ellos es la estructura productiva de los países y los tipos de materias que se producían. Por ejemplo, si un país en esa época tenía una economía más basada en la explotación de recursos naturales versus una que estaba principalmente basada en la agricultura o transporte, es probable que el tipo de instituciones que se formen sean distintas. Así también ese tipo de variables también tendrían un impacto sobre el PIB (evidentemente los retornos de capital de la industria extractiva, agricultura y transporte son distintos). Comentaremos más sobre este punto más adelante.
3. **Sesgo de investigador:** tal como indican los autores, las medidas sobre instituciones han sido construidas *ex post* y al hacer los análisis hoy es esperable que se produzca un sesgo en ver estas “mejores instituciones” en países más ricos.
4. **Error de medición:** inicialmente pensamos que podía existir un error asociado a la forma en que se mide la performance económica, esto es el log del PIB siendo que podría ser relevante su relación con la fuerza productiva si es que nos estamos enfocando en **derechos de propiedad** bajo el argumento de la seguridad y estabilidad para la inversión. Ahora bien, los autores prueban sus estimaciones con las medidas de Hall y Jones (1999) con el PIB por trabajador obteniendo resultados similares. A su vez, la técnica empleada ya ha sido antes probada por otras investigaciones como las de Mauro (1995) y Hall y Jones (1999)
5. **Cluster de instituciones:** el índice que se ocupa no tiene que ver realmente solo con derecho de propiedad, sino que reúne una serie de características institucionales que quizás no reflejan un mismo indicador. Por ello, se produciría un **sesgo de atenuación** en las estimaciones por MCO en la medida en que hay un error de medición en la variable dependiente. Antes de construir un índice de esas características se debe al menos pasar una prueba de consistencia interna (*alfa de Cronbach*), explorar las varianzas compartidas a partir de un análisis factorial o, en caso de querer incluir su error de medición en la regresión, hacer un modelo de ecuaciones estructurales.

Pensemos, nuevamente en países donde su PIB principalmente está basado en transporte de exportaciones e importaciones (países portuarios). En estos casos, estos países no necesitan potenciar instituciones respecto a derechos de propiedad y más si sobre derecho a comercio. Si sus derechos de propiedad no están reforzados *no necesariamente su crecimiento será menor*. De hecho, se ha demostrado que en muchos casos países de este estilo tienen un desarrollo económico considerable.

## C.2 Ejemplo de variable omitida

Arriba propusimos algunos ejemplos generales no solo de variables omitidas, sino que incluso podrían estar correlacionadas con el factor institucional (lo que podría sesgar su efecto). Ahora solo nos enfocaremos en tres ejemplos de variable omitida respecto al PIB y que puedan estar siendo

captadas en el término de error. En términos generales estas refieren a las **condiciones de la colonia** que producen diferencias entre los países. Para hacer más interesante la discusión decidimos incorporar esas variables en los modelos reportados en la Tabla 1.1

1. *Condiciones ambientales*: investigaciones en la misma línea (Hall y Jones, 1999) plantearon que existe un efecto directo del clima de las colonias sobre la performance económica. El argumento es que en países donde el medio ambiente no era favorable para la colonización, la formación de instituciones extractivas era más probable, y el extractivismo producía una merma en la capacidad productiva de los países (no se incentiva el desarrollo económico). Gallup et al (1998) propuso que una forma de incorporar el clima es aladiendo la *latitud* de las zonas (en valor absoluto). Como podemos ver en la Tabla 1.1, la latitud no es significativo una vez que controlamos por las instituciones, tipo de país y tipo de relación colonial.
2. *Tipo de institución instalada*: el tipo de institucionalidad que se instala en sus orígenes es relevante para entender las diferencias en la producción. Planteamos dos tipos de ellas
  - a. “*Neo-europeas*”: son instituciones que intentan replicar instituciones europeas con mayor énfasis en propiedad privada y chequeos de poder de gobiernos (eg. Australia, Nueva Zelanda, USA y Canadá). Eso promueve un estado de bienestar, mayor infraestructura (Acemoglu et al (2001; p.1375)
  - b. “*Extractivas*”: se instalan instituciones que no proveen protección a la propiedad privada y no generan balances en la expropiación. El propósito central es expropiar para acumular (eg. belgas en el Congo, encomienda en América Latina, esclavismo en África). Se desarrolló una política económica basada en los commodities.

Entonces, probamos esa hipótesis y notamos que *sí tienen un impacto estadísticamente significativo sobre la producción*. Como vemos en la Tabla 1.1 (Modelo 1 - Zonas) tomamos la variable construida por los autores (países asiáticos, países africanos y países neo-europeos). Podemos ver que pese a que la protección contra el riesgo de expropiación sigue siendo significativo, la dummy de pertenecer a un país africano (extractivo) es también estadísticamente significativas y con tamaños efectos considerables. En promedio, y controlando por el resto de las variables del modelo, pertenecer a un país africano reduce en 92,5% la producción. De hecho tal como se predice en el artículo casi al final, los países africanos son más pobres 145 veces más pobres que América, incluso después de descontar el efecto de las instituciones (que corresponde a la misma estimación que realizamos nosotras).



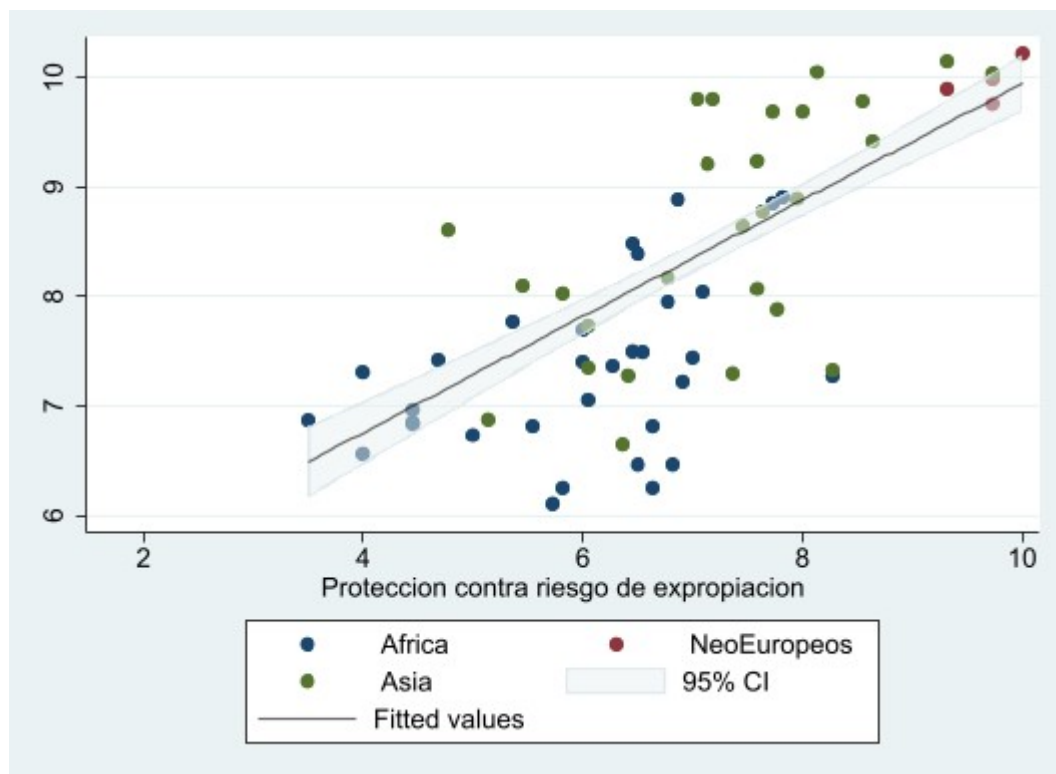


Figura 2: Gráfico de dispersión entre ingresos per cápita y protección a expropiación según países

Podemos reforzar este argumento con la Figura 2. Notemos la distribución del PIB, según país y según índice de protección a la propiedad privada. El argumento sobre el colonialismo extractivo parece ser sustantivo. Históricamente se ha evidenciado que la situación e instituciones coloniales persisten incluso después de la independencia pues estas lógicas formaron la base de las instituciones actuales, pero sobre todo en los países donde se desarrolló un colonialismo extractivo (en Brasil el esclavismo duró hasta 1886 y en México el trabajo forzado hasta su revolución en 1910) (para más véase Acemoglu, 1995).

### 3. Zonas geográficas:

Ahora bien, existía un problema que no podíamos soslayar, y es que estas variables que consideraban el tipo de institucionalidad no diferenciaban entre países africanos y latinos. Por ello hicimos un *match* con la base de datos del Banco Mundial con los nombres de los países en la fecha y obtuvimos su continente. Resultado de eso construimos el “Modelo 1- Continentes” y la Figura 3. Como podemos ver, el resultado respecto de África es importante. Notemos que en promedio, y controlando por las instituciones y condiciones climáticas, los países colonizados en África tienen en promedio un producto mucho menor que los de América (con un 99% de confianza). Esto inclusive considerando que en ambas zonas se desarrolló extractivismo, pero no es menor indicar que al parecer América es más heterogéneo en su forma colonial y por eso se da esta diferencia sustancial (se puede ver esto también en la Figura 3). De todas formas no podemos pasar por alto la ya indicada comparación de Chile y Nigeria, donde el resultado es abismal.

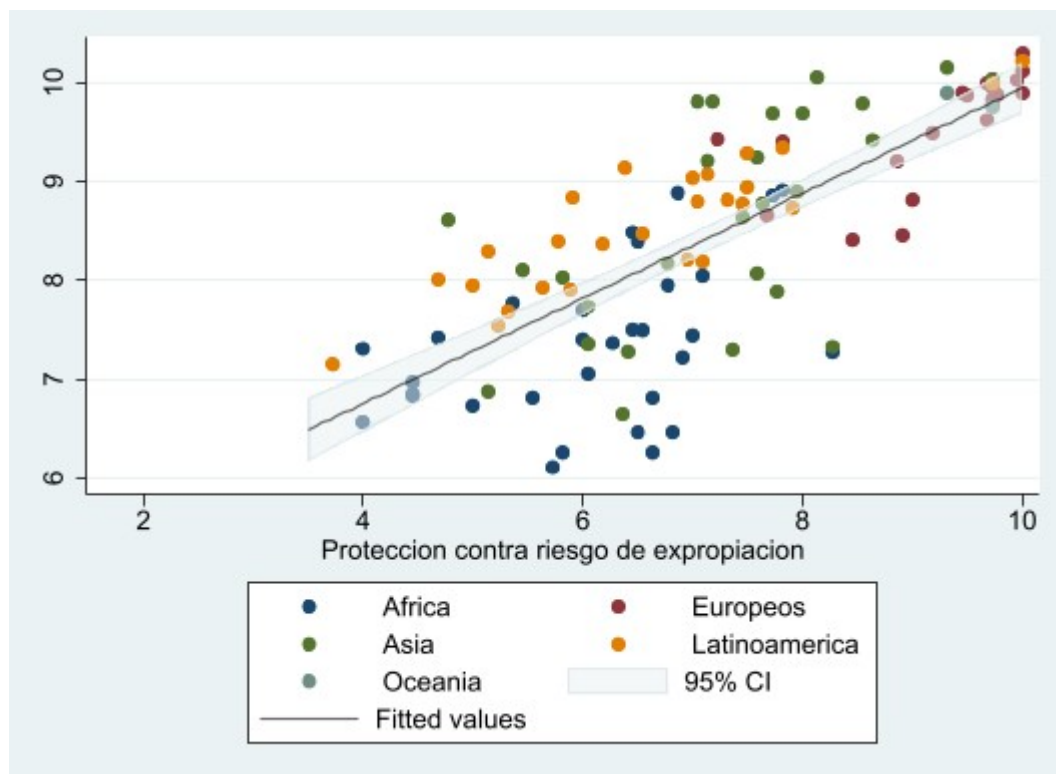


Figura 3: Gráfico de dispersión entre ingresos per cápita y protección a expropiación - Datos propios

**Tabla 1.1.** Modelos de regresión lineal que estiman log GDP per cápita - Con zona y Continente

| Predictores  | Modelo 1 -MCO       | Modelo 1 - Zonas     | Modelo 1- Continentes |
|--|---------------------|----------------------|-----------------------|
| Proteccion contra riesgo de expropiacion                         | 0.532***<br>[0.041] | 0.388***<br>[0.052]  | 0.419***<br>[0.055]   |
| Continente (ref: Americas)                                       |                     |                      |                       |
| Africa   |                     |                      | -0.986***<br>[0.170]  |
| Asia   |                     |                      | -0.285<br>[0.177]     |
| Europa   |                     |                      | -0.367<br>[0.274]     |
| Oceania (otros)  |                     |                      | -0.107<br>[0.488]     |
| Latitud  |                     | 0.323<br>[0.446]     | 0.764<br>[0.542]      |
| Pais africano  |                     | -0.925***<br>[0.166] |                       |
| Pais asiatico  |                     | -0.159<br>[0.155]    |                       |
| Paises neo-europeos  |                     | 0.184<br>[0.336]     |                       |
| Constant   | 4.626***<br>[0.301] | 5.871***<br>[0.345]  | 5.657***<br>[0.370]   |
| Observations   | 111                 | 111                  | 109                   |
| Adjusted R-squared   | 0.608               | 0.701                | 0.701                 |
| Note: Run on 30 Nov 2021 using data from input\country-codes.dta |                     |                      |                       |
| Standard errors in brackets                                      |                     |                      |                       |
| *** p<0.01, ** p<0.05, * p<0.1                                   |                     |                      |                       |

## D. Variable omitida de capital humano

El capital humano es un factor relevante en esta estimación por dos razones. Primero, por el sesgo que podría producir por el efecto *directo* que tiene sobre la performance económica, y en segunda instancia, por la correlación que puede tener con el factor institucional, lo que estaría atenunado el efecto del coeficiente principal ( $\beta_1$ ) del índice institucional. Analizaremos los dos por separado, y en la segunda parte discutiremos la dirección de su sesgo.

1. *Efecto del capital humano sobre la producción*: investigaciones contemporáneas a la de los autores (Barro, 2001) plantean que países con mayor capital humano tienen un mayor crecimiento económico, en la medida en que este produce condiciones para producir de manera más eficiente y sostenida a largo plazo. Si bien la macroeconomía neoclásica lo ha incorporado como una variable residual (véase en modelos como el de Ramsey, 1928; cf. Solow, 1965 o *residuos de Solow*) cada vez se ha argumentado más como un factor más que promueve la productividad total (Cruz & Ahmed, 2018), y que con la crisis de la sociedad industrial el retorno del capital humano de hecho se hizo más sustantivo que el capital físico (en Hanushek & Woessmann, 2012) toda vez que la productividad empezó a crecer a escalas mayores cuando el sector productivo pasó a ser menos manual y requerir más altos niveles de calificación.
2. *Relación de las instituciones y el capital humano*: por un lado, el capital humano no solo promueve una mayor innovación y desarrollo, sino que también un mayor capital humano permite *habilitar* de mejor forma las instituciones. Esto quiere decir que, en la medida en

que exista más medios y conocimientos en la sociedad, estos no solo se transmitirán hacia los canales productivos (la economía), sino que también a la capacidad de construir instituciones complejas y adecuadas a contextos de protección de la propiedad, comercio y estabilidad financiera. Ahora bien esta también es una relación en la otra dirección (y por eso planteamos que existe una *correlación* entre ambas variables), dado que como indican los autores en 1995, países donde se refuerzan los derechos de propiedad, se promueve más la inversión en capital humano dado que por la seguridad que estos regímenes normativo-institucionales establecen, los capitales están dispuestos a invertir a *largo plazo* y probablemente en factores que inicialmente *tienen un mayor costo* que aquellos como el capital físico.

Imaginemos que el modelo real que estima la producción es la siguiente

$$\ln(PIB_{percapita})_i = \beta_o + \beta_1 \cdot aveexpr_i + \beta_2 \cdot capital_{humanoi} + u_i$$

Donde  $aveexpr \wedge capital_{hum}$  son vectores de variables independientes. Luego estimamos la regresión lineal como

$$\ln(\widehat{PIB}_{percapita})_i = \hat{\beta}_o + \hat{\beta}_1 \cdot aveexpr_i + \hat{\beta}_2 \cdot capital_{humanoi} + \hat{u}_i$$

Veremos matricialmente el efecto que podría tener omitir una variable relevante, dado que llegaremos a la conclusión que la estimación no solo afecta en su relación con la variable dependiente sino que también las independientes

$$\begin{bmatrix} aveexpr'aveexpr & aveexpr' \cdot capital_{hum} \\ capital_{hum}' \cdot aveexpr & capital_{hum}' \cdot capital_{hum} \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} aveexpr' \cdot \ln(PIB_{percapita}) \\ capital_{hum}' \cdot \ln(PIB_{percapita}) \end{bmatrix}$$

Donde podemos evidenciar a partir de la siguiente matriz que podemos obtener tanto  $\hat{\beta}_1$  como  $\hat{\beta}_2$

$$(aveexpr'aveexpr) \cdot \hat{\beta}_1 + aveexpr' \cdot capital_{hum} \hat{\beta}_2 = aveexpr' \ln(PIB_{percapita}) \Rightarrow \hat{\beta}_1 = (aveexpr'aveexpr)^{-1} \cdot aveexpr' \cdot (\ln(PIB_{percapita}) - capital_{hum} \hat{\beta}_2)$$

En este punto parece ser evidente que es el *set* de variables de la regresión la que define el valor de los coeficientes de regresión **y no únicamente al que está relacionado**. Por ejemplo, en el  $\hat{\beta}_1$  está definido por su relación que tiene con el  $\ln(PIB)$  **menos** un vector que parcializa o corrige esta relación  $(aveexpr'aveexpr)^{-1} \cdot aveexpr \cdot capital_{hum} \hat{\beta}_2$  donde aparece el capital humano y su efecto parcial sobre el  $\ln(PIB)$  también. Entonces, existirá un sesgo en dos líneas, como indicamos intuitivamente antes.

1. Un sesgo que depende de la relación que tiene la variable omitida con la variable de interés, donde si el coeficiente de esa variable llegara a ser significativo ( $\hat{\beta}_2$ ), *evidentemente* tendrá un impacto en la estimación de  $\hat{\beta}_1$ .
2. Un sesgo que se produce por la correlación entre las variables, en donde solo sería cero si y solo si estas son ortogonales.

En ambos casos su omisión está sesgado las estimaciones. En el primer caso produce un problema importante de variable omitida que se va al término de error, entonces la estimación por MCO es menos precisa. En el segundo caso genera un problema peor: el término de error, al omitir esta variable, a su vez correlaciona con uno de los predictores. Eso ocasiona que además de ser menos preciso MCO, el coeficiente no es ni sesgado ni consistente. Revisaremos la dirección de ese sesgo a continuación.

## E. Dirección de sesgo

Matemáticamente podemos calcular el sesgo a partir de la esperanza del estimador, y verificar si este es igual al coeficiente real. Imaginemos que estimamos una regresión sin el capital humano, y entonces el coeficiente asociado al factor institucional será distinto (como ya demostramos). Luego, la esperanza del nuevo estimador es

$$E(\tilde{\beta}_1) = \beta_1 + \delta_2 \cdot \beta_2, \quad \text{donde} \quad \delta_2 = (ave\!xpr'ave\!xpr)^{-1} \cdot ave\!xpr \cdot capital_{hum}$$

El sesgo sería entonces

$$E(\tilde{\beta}_1 - \beta_1 | X_i) = E(\beta_1 + \delta_2 \cdot \beta_2 - \beta_1 | X_i) = \beta_2 \cdot (ave\!xpr'ave\!xpr)^{-1} \cdot ave\!xpr \cdot capital_{hum}$$

Entonces como esperamos hipotéticamente que el efecto del capital humano sea positivo y estadísticamente significativo ( $\beta_2 > 0$ ) y tenga una correlación positiva con la fortaleza de los derechos de propiedad el sesgo  $\beta_2 \cdot (ave\!xpr'ave\!xpr)^{-1} \cdot ave\!xpr \cdot capital_{hum} > 0$ . Entonces el  $\tilde{\beta}_1$  que estimamos al inicio estaría sobre estimando el efecto de  $\beta_1$  real, dado que  $\tilde{\beta}_1 > \beta_1$

## Pregunta 2

Pudimos notar anteriormente que el  $\hat{\beta}_{1MCO}$  este sesgado, que por el sesgo de variable omitida (y también el de simultaneidad que indicamos), *avexpr* correlacione con el término de error y por tanto sea endógena. Los autores buscaron una fuente de variación exógena basada en la teoría institucional sobre la colonización que permitiese captar los elementos omitidos que indicamos en nuestras propuestas de sesgo en la parte **1.C**, esta es la probabilidad de muerte de los colonos al llegar a las colonias hace cientos de años (*logem4*) entre 1970 a 1990 (datos de Phillip Curtin).

En el artículo explican que cuando los colonos llegaban a países donde la mortalidad era alta instalaban solo instituciones extractivas (malas). En cambio, cuando llegaban a países donde la mortalidad era baja, muchas veces mandaban gente a vivir ahí por lo que desarrollaban instituciones de mayor calidad. Como dichas instituciones persisten en el tiempo, ellas terminan afectando los derechos de propiedad de los países hoy. Es decir, la mortalidad de los colonos en el pasado influye en que tan bueno son los derechos de propiedad hoy.

Formalmente

### Ecuación reducida

$$avexpr_i = \pi_o + \pi_1 \cdot logem4 + v$$

### Regresión por Mínimos Cuadrados

$$\ln(PIB_{percapita})_i = \beta_o + \beta_1 \cdot avexpr_i + u_i$$

donde

- $i$ : subíndice del país
- $\ln(PIB_{percapita})$ : logaritmo del PBI per cápita
- *avexpr*: índice de protección contra riesgo de expropiación del país.
- *logem4*: la probabilidad de muerte de los colonos al llegar a las colonias hace cientos de años
- $u$ : término de error

## A. Condiciones de instrumentos

**1. Exogeneidad:** El primer supuesto consiste en que el instrumento debe ser exógeno, esto quiere decir que el instrumento no debe correlacionar con el término del error ( $Cov(Z, u) = 0$ ). Es decir, el instrumento no debe ser un predictor en la regresión (sin efecto parcial en la performance económica actual) y no está correlacionada con variables omitida.

La variable instrumental, condicional a los controles incluidos en la regresión, la tasa de mortalidad de los colonos europeos de hace 100 años **no tiene un efecto sobre el PIB per cápita hoy**, contrario sobre su efecto sobre el desarrollo de instituciones actuales (relevancia). Una de las preocupaciones sobre la *exogeneidad* tenía que ver con que en medio ambientes malos (donde se desarrollan más enfermedades etc.) puede tener un efecto directo sobre la performance económica, y en ese caso, se debería incorporar el efecto de las enfermedades sobre el PIB. El tema es que ese tipo de muertes en la época tenían que ver con enfermedades que hoy están erradicadas tales como la malaria y fiebre amarilla, por lo que es difícil que tengan un efecto actual sobre el PIB hoy.

Sumado a este argumento, los autores (p.1372) prueban que los resultados no cambian incluyendo otros controles.

**2. Relevancia:** El segundo supuesto que debe cumplirse para que el instrumento sea válido es que éste sea relevante, esto significa que el instrumento debe estar correlacionado con la variable endógena  $X$  ( $Cov(Z, X) \neq 0$ ).

En el caso de la tasa de mortalidad de los colonos podemos decir que la evidencia muestra una fuerte relación entre esta y las recientes instituciones en la medida en que explica sobre un 25% de la variación. Podemos ver la relación negativa en la *Figura 4* donde a mayor tasa de mortalidad de colonos menor será el índice promedio del riesgo de expropiación. En términos sustantivos es probable que en países donde se hayan aplicado instituciones más extractivas las condiciones ambientales y sociales no hayan sido las óptimas para el proceso de colonización. Por ejemplo, si colonos vieron resistencia de las comunidades de las colonias, es probable que hayan tenido que tener enfrentamientos bélicos produciendo la muerte de ellos. A su vez, es probable que entre menos colonos capaces de establecer el orden occidental (pues morían por enfrentamientos o condiciones ambientales), menos probable que hayan podido instalar instituciones<sup>2</sup>.

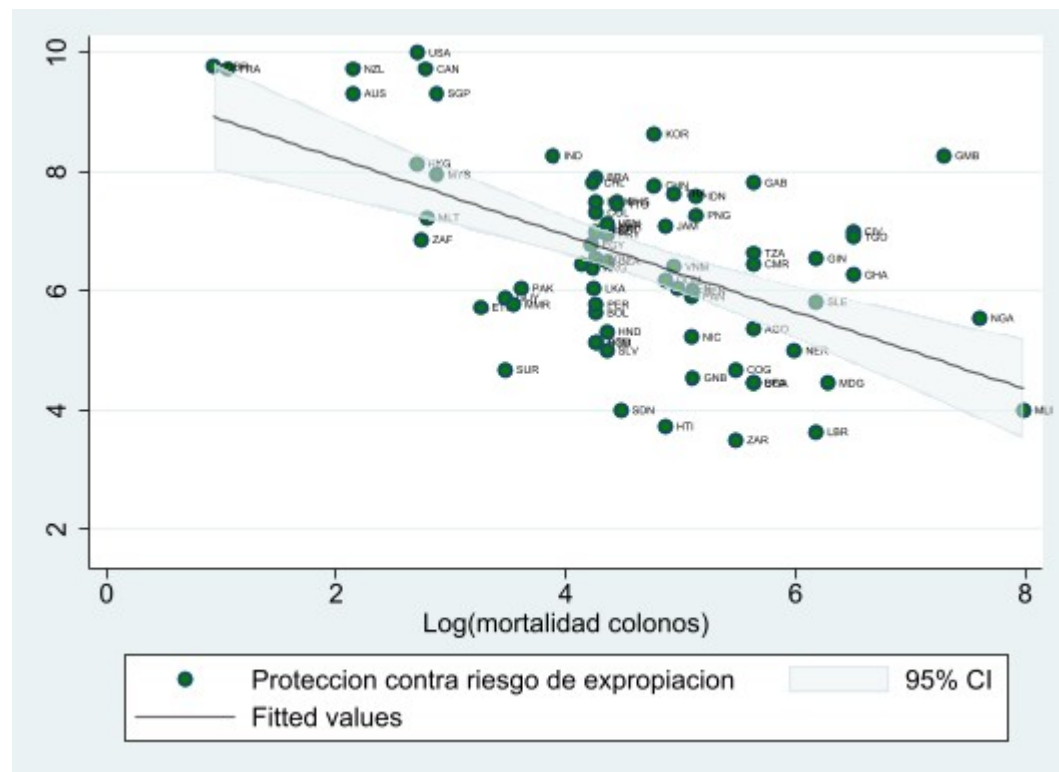


Figura 4: Gráfico de correlación entre protección a expropiación y probabilidad de muerte de los colonos

<sup>2</sup>De hecho podemos mirar este ejemplo mirado desde Chile. En Chile la corona española siempre tuvo de trinchera el Bío-Bío, mientras que el Centro-Norte del país fue colonizado y “pacificado” rápidamente. Esta situación tiene consecuencias hasta hoy en las instituciones actuales y en su relación con la producción. En la Araucanía, donde se aplica una producción extractiva, el conflicto aún perdura. Es probable que la tasa de mortalidad de españoles sea mayor que la que se tuvo en Valparaíso. Y si luego comparamos la forma institucional de ambas regiones hoy notaremos sus diferencias.

Evidentemente pueden haber más instrumentos e incluso que compartan más varianza, pero es muy importante que **a la vez se cumpla la exogeneidad** (lo que es una dificultad, como indican en el artículo). De hecho discuten desde las teorías del desarrollo económico-institucional otros instrumentos como la cantidad de europeos en las colonias hoy, otros vectores de covariables que afectan a las variables de interés, y como reportan en la tabla 3 a partir de 10 regresiones distintas es que la tasa de mortalidad de los colonos (M para ellos) es el que mejor responde a la técnica de identificación pues no afecta al producto.

## B. Mínimos Cuadrados en Dos Etapas manualmente

### B.1 Procedimiento de estimación

#### Ecuación reducida

$$aveexpr_i = \pi_o + \pi_1 \cdot logem4 + v$$

#### Regresión por Mínimos Cuadrados en dos etapas

$$\ln(PIB_{percpita})_i = \beta_o + \beta_1 \cdot aveexpr_i + u_i$$

Como se puede observar en nuestra *Tabla 2.1*, con un 99% de confianza podemos plantear que existe un impacto negativo de la muerte de colonos sobre la calidad actual de las instituciones. De manera más precisa cuando aumenta en 1% la tasa de mortalidad de los colonos, el índice de protección a la propiedad privada se ve afectado en una disminución de un 0,00647 puntos del índice ( $\frac{\pi_1}{100}$ ). Si consideramos la distribución de este indicador, el impacto es relativamente muy bajo, toda vez que se mide en una escala de 1 al 10, y donde recién un aumento en cerca de un 160% de las tasas de mortalidad disminuirían en 1 punto el índice de riesgo de expropiación.

**Tabla 2.1.** Modelos de regresión lineal que estiman variable endógena

| Predictores  | First Step           |
|--|----------------------|
| Log(mortalidad colonos)  | -0.647***<br>[0.115] |
| Constant   | 9.528***<br>[0.548]  |
| Observations   | 74                   |
| Adjusted R-squared   | 0.295                |
| Note: Run on 30 Nov 2021 using data from input\country-codes.dta |                      |
| Standard errors in brackets                                      |                      |
| *** p<0.01, ** p<0.05, * p<0.1                                   |                      |

Más allá de eso, si podemos evidenciar que hay un efecto relevante cuando las muertes explican en cerca de un 30% la varianza del índice institucional. A su vez, podemos chequear la relevancia y



fortaleza del instrumento en el resto de los puntos. Desde ya declaramos que nuestros resultados se condicen con los reportados por los autores en la tabla 4 del artículo.

## B.2 Relevancia de los instrumentos

El supuesto de relevancia en este caso implica que la correlación entre el instrumento, es decir, la probabilidad de muerte de los colonos al llegar a las colonias hace cientos de años (*logem4*), y la variable endógena, que en este caso es el índice de protección contra riesgo de expropiación del país, debe ser diferente de 0. Ya habíamos delineado que como se puede observar en el gráfico (Figura 4), la correlación entre protección a expropiación y probabilidad de muerte de los colonos es negativa, por lo que podría ser un instrumento relevante.

Una manera más segura de probar el efecto que tiene el instrumento es tomar el coeficiente de la ecuación reducida y probar si este es igual o no a cero, pese a que su efecto parcial reportado sea *-0.64 aproximadamente*.

$$H_o : logem4 = 0 H_1 : logem4 \neq 0$$

**Tabla 2.2** Test restricciones múltiples relevancia del instrumento

| Instrumento |                 |
|-------------|-----------------|
| F(q, n-k-1) | F(1,72) = 31.51 |
| p-value     | 0.0000          |

En la Tabla 2.2 podemos notar dos hechos. El primero es que, con un 99% de confianza rechazamos la hipótesis nula de que *logem4* no tiene un efecto significativo sobre el índice de protección al derecho de propiedad. El segundo es la *calidad* del instrumento en términos de **fuerza**. Se dice que un instrumento es débil cuando su correlación con la variable endógena es cercana a 0. Esto genera diferentes consecuencias, tales como que las estimaciones de VI/MC2E serán muy poco precisas y la inferencia usual deja de ser válida, de modo que los estadísticos t y F serán poco fiables.

Dado que, para que sea un instrumento fuerte la correlación entre el instrumento y la variable endógena debe estar lejana a 0, se puede concluir que si el instrumento es fuerte, va a cumplir también con el supuesto de relevancia. Para comprobar esto, se aplica una regla básica simple, la cual indica que debemos verificar que el estadístico  $F \geq 10^3$ , y si esto se cumple el instrumento es fuerte. En la tabla 2.2, podemos observar que el estadístico  $F$  es igual a *31.51*, por lo que el instrumento es fuerte.

## B.3 Segunda etapa

La ecuación de la segunda etapa consiste en tomar la ecuación original que habíamos planteado al inicio, pero ocuparemos nuestra variable de interés instrumentada ( $ave\hat{xpr}_i$ ) en el logaritmo de las muertes (proceso que hicimos en la primera etapa y reportamos en ecuación reducida)

<sup>3</sup>en caso de que tengamos un instrumento solamente, no debemos comparar simplemente t. Más bien debemos transformar  $t^2 = F$  y luego evaluar el límite indicado.

## Regresión por Mínimos Cuadrados en dos etapas

$$\ln(PIB_{percpita})_i = \beta_o + \beta_1 \cdot ave\hat{xpr}_i + u_i$$

Como se puede ver en la misma ecuación hemos tomado el valor predicho de el índice institucional y lo hemos ocupado como un regresor en nuestro segundo modelo. Los resultados los reportamos en la *Tabla 2.3*.

**Tabla 2.3.** Modelos de regresión lineal que estiman log GDP per cápita

| Predictores  | Second Stage        |
|--|---------------------|
| Predicción lineal de índice de riesgo a expropiación             | 0.872***<br>[0.099] |
| Constant   | 2.350***<br>[0.653] |
| Observations   | 81                  |
| Adjusted R-squared   | 0.490               |
| Note: Run on 30 Nov 2021 using data from input\country-codes.dta |                     |
| Standard errors in brackets                                      |                     |
| *** p<0.01, ** p<0.05, * p<0.1                                   |                     |

## B.4 Interpretación de las la estimación por variables instrumentales

Como se puede ver en la *Tabla 2.3*, con un 99% de confianza que, manteniendo todo lo demás constante, existe una relación positiva entre la calidad institucional y la performance económica. De manera más específica, el tamaño efecto es muy grande considerando que un aumento de un 1% en el índice de protección contra riesgo de expropiación del país se traducirá en un aumento del 87,2% en el PIB per cápita ( $\beta \cdot 100$ ,  $p < 0.001$ ). Este efecto es muy similar con la base muestral de los autores (ocupan varios *datasets*), dando un valor de 94,0% (aunque como veremos en el siguiente punto nuestra estimación mejorará al hacer el proceso de dos etapas en un único proceso).

Antes, tiene profundo sentido económico el resultado. Como habíamos mencionado, el desarrollo institucional genera condiciones de posibilidad para el desarrollo de los factores productivos. Por ejemplo, si existen instituciones que aseguren el derecho de propiedad tanto los capitales nacionales como extranjeros confiarán en inversión a largo plazo, sobre todo el factores productivos que podrían significar un capital inicial muy alto pero a largo plazo más eficiente (como la innovación e inversión en ciencia). Si suponemos perfecta movilidad de los capitales y paridades de tasas de interés descubierta, notaremos que un aumento en la balanza comercial generará un aumento en el PIB nacional.

A su vez, instituciones con mayor fortaleza no solo aseguran condiciones de producción, sino también condiciones de reproducción de la fuerza de trabajo. Y aquí el instrumento juega un rol relevante: Acemoglu et al (2001) indican que existieron diferencias en las formas de colonización, donde las colonizaciones más neo-europeas **también instalaron sistemas de bienestar**. Un mayor nivel de

bienestar, pese a que significa inicialmente un gasto para el gobierno, generará un retorno del empleo más eficiente y por tanto una productividad marginal del trabajo mayor que si estos no existieran. En ese sentido, el instrumento nos permitió ajustar con la tasa de mortalidad esa variación o factor exógeno dentro del desarrollo institucional, en la medida en que ese efecto negativo de mayores tasas de muerte genera peores instituciones estaba atenuando el real efecto de las instituciones sobre la performance económica.

## C. Estimación de Mínimos Cuadrados en Dos Etapas - Directo

Como indicamos anteriormente, nuestras estimaciones son similares y en la misma dirección que los autores pero presentamos pequeñas diferencias. Estas se deben a que la primera estimación que hicimos tiene algunos errores que comentaremos a continuación luego de presentar su resultado en la *Tabla 2.4*.

### C.1 Presentación de la estimación

**Tabla 2.4.** Modelos de regresión por Variables Instrumentales que estiman el efecto institucional sobre la performance económica

| Predictores  | First-stage          | Instrumental variables 2SLS |
|--|----------------------|-----------------------------|
| Proteccion contra riesgo de expropiacion                         |                      | 0.868***<br>[0.123]         |
| Log (mortalidad de colonos)                                      | -0.631***<br>[0.116] |                             |
| Constant   | 9.514***<br>[0.548]  | 2.370***<br>[0.826]         |
| Observations   | 70                   | 70                          |
| Adjusted R-squared   | 0.294                | 0.305                       |
| F statistic  | 29.8***              |                             |
| Wald test  |                      | 49.67***                    |
| Note: Run on 28 Nov 2021 using data from input\country-codes.dta |                      |                             |
| Standard errors in brackets                                      |                      |                             |
| *** p<0.01, ** p<0.05, * p<0.1                                   |                      |                             |

Notemos que la estimación del efecto de el índice institucional ha cambiado relativamente poco pero si a tres cifras significativas (inicialmente era 0.872, y ahora es 0.868). Esta diferencia se debe a que en la estimación simultánea de Mínimos Cuadrados en 2 Etapas (Ecuación reducida y luego 2SLS) estiman los residuos de la regresión de manera **compuesta**. Una parte considera el residuo de la regresión de la ecuación reducida  $\hat{v}$  y el otro el residuo de la regresión  $\hat{u}$ . Entonces, distingue entre los factores no observados, aleatorios entre una ecuación y otra. Esto tendrá consecuencias en la suma de los residuos al cuadrado ( $u'u$ ) que permite obtener los errores estándar de los coeficientes de regresión. Si la estimación simultánea considera a ambos como residuos, la suma de residuos será mayor, y por ende la precisión de la estimación de los coeficientes mejor. Mientras que si estimamos por separado, los residuos de la ecuación reducida no serán considerados en la ecuación de la segunda etapa por lo que los errores estándar serán menos precisos.

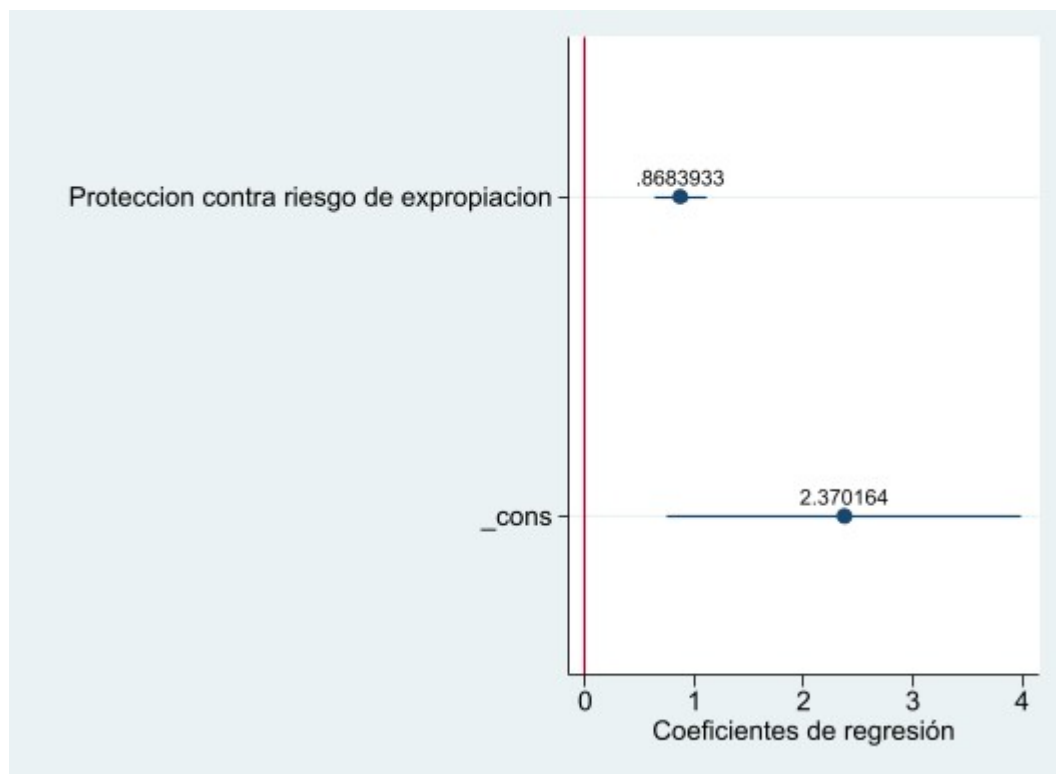


Figura 5: Gráfico de coeficiente del efecto de la determinante institucional sobre la performance económica

Sumado a ello, podemos afirmar con un 99% de confianza que, manteniendo todo lo demás constante, existe una relación positiva entre la calidad institucional y la performance económica. De manera más específica, el tamaño efecto es muy grande considerando que un aumento de un 1% en el índice de protección contra riesgo de expropiación del país se traducirá en un aumento del 86,8% en el PIB per cápita ( $\beta \cdot 100$ ,  $p < 0.001$ ). Como podemos ver en el *forestplot* de la *Figura 5* este coeficiente es estadísticamente significativo y bastante preciso dado que no está cercano al cero (línea roja) y su intervalo de predicción es bastante pequeño (tamaño de línea azul).

## Pregunta 3

En este apartado discutiremos los resultados obtenidos de la regresión por variables instrumentales ( $\hat{\beta}_{1IV}$ ) y Mínimos Cuadrados Ordinarios ( $\hat{\beta}_{1MCO}$ ). Para ello, haremos un examen de la evaluación de las estimaciones a partir de las pruebas de endogeneidad de variables (*contraste de Hausman*), discusión sobre propiedades de los estimadores como el sesgo y consistencia de ambos estimadores. Al final del apartado haremos algunos ejercicios analíticos sobre errores de medición.

### A. Test de Haussman

Nos permite evaluar si efectivamente las variables eran endógenas o no. Esto quiere decir que si realmente se necesitaba o no construir un modelo de regresión por variables instrumentales. Intuitivamente lo que implica ello es que uno compara la estimación por Mínimos Cuadrados y la de Variables Instrumentales, y si las estimaciones son las mismas, la variable en cuestión no era endógena (esto, evidentemente luego de considerar un instrumento relevante y fuerte).

Analíticamente lo que se hace es mirar la relación que se establece entre los errores de la ecuación reducida ( $v$ ) y la general ( $u$ ). Si estos no están relacionados ( $Cov(v, u) = 0$ ), entonces la parte que no explican los instrumentos a la variable endógena en cuestión tampoco afecta a la variable dependiente.

En concreto, tomaremos los residuos de la ecuación reducida ( $\hat{v}$ ) y estimaremos la ecuación original de Mínimos Cuadrados Ordinarios con ( $\hat{v}$ ) como predictor. Si resulta ser que este coeficiente no es significativo ( $\delta = 0$ ), entonces la “parte endógena” de la variable cuestionada realmente no explican en nada a nuestra variable de interés formal. En términos de hipótesis testaremos que

$$H_o : \delta = 0 \quad H_1 : \delta \neq 0$$

En ese sentido,  $\hat{\beta}_{1IV}$  es consistente bajo la hipótesis nula (coeficiente obtenido por regresión por variables instrumentales),  $\hat{\beta}_{1MCO}$  es inconsistente bajo hipótesis alternativa (coeficiente obtenido en MCO).

Matemáticamente se calcula de la siguiente forma, tomando como prueba de contraste la distribución  $Chi^2$  que hemos resumido en la *Tabla 3.1*

$$Chi^2(1) = (\hat{\beta}_{1IV} - \hat{\beta}_{1MCO})' \frac{(Var(\hat{\beta}_{1IV}) - Var(\hat{\beta}_{1MCO}))}{(\hat{\beta}_{1IV} - \hat{\beta}_{1MCO})}$$

**Tabla 3.1** Test de Hausman

|            | $hat\beta_{1IV}$ | $hat\beta_{1MCO}$ | $\hat{\beta}_{1IV} - \hat{\beta}_{1MCO}$ | $SE(\hat{\beta}_1)$ |
|------------|------------------|-------------------|--|---------------------|
|            | .8683933         | .5318713          | .336522                                  | .1163249            |
| $Chi^2(1)$ | 8.37             |                   |  |                     |
| p-value    | 0.0038           |                   |  |                     |

## B. Interpretación de Test de Haussman

Como vemos en la Tabla 3.1, podemos ver que con un 95% de confianza no podemos rechazar la hipótesis nula de exogeneidad de nuestra variable de calidad institucional. Esto quiere decir que el error asociado al índice institucional (su parte endógena) sí explica en algo a la variable dependiente, y en consecuencia, no controlar endogeneidad a partir de una fuente de variación exógena produciría que este componente se correlacione con el error del modelo de regresión original ( $u$ ), produciendo estimadores sesgados e inconsistentes.

En el fondo tendríamos matemáticamente un error compuesto que no solo depende de un error aleatorio, sino que también de un error asociado a la variable endógena

$$\hat{u} = \delta \hat{v} + error$$

## C Test de Hausman manual

Los pasos para hacer manualmente la prueba de Hausman son los siguientes

**Paso 1.** Estimar la ecuación reducida del instrumento. Notemos que en este caso no tenemos otros regresores en el modelo así que no controlamos esta predicción contra todo el resto de variables exógenas

$$\hat{ave}pr = \hat{\pi}_o + \hat{\pi}_1 \cdot logme4 + \hat{v}$$

**Paso 2.** Estimar la regresión original de Mínimos Cuadrados Ordinarios considerando el error de la ecuación reducida ( $\hat{v}$ )

$$\ln(PIB_{percapita})_i = \beta_o + \beta_1 \cdot avepr_i + \delta \hat{v}_i + u_i$$

**Paso 3.** Calcular el estadístico de contraste  $Chi(1)$

$$Chi^2(1) = (\hat{\beta}_{1IV} - \hat{\beta}_{1MCO})' \frac{(Var(\hat{\beta}_{1IV}) - Var(\hat{\beta}_{1MCO}))}{(\hat{\beta}_{1IV} - \hat{\beta}_{1MCO})}$$

Reemplazando los valores de la tabla 3.1 obtenemos 8.37.

**Paso 4.** Evaluar la hipótesis nula donde  $H_o$  indica que  $\delta = 0$  (exogeneidad) y  $H_1 : \delta \neq 0$  (endogeneidad).

Si miramos la distribución  $Chi^2(1)$  con una restricción podremos ver que se rechaza la hipótesis nula con un 95% de confianza.

También reportamos el modelo auxiliar con los residuos de la primera etapa incorporados. Notemos que también incorporamos la significancia de los residuos en la primera etapa.

**Tabla 3.2** Modelo de regresión auxiliar que resume la estimación del modelo de interés y la relación con los residuos de la ecuación reducida

| Predictores  | Modelo Auxiliar      |
|--|----------------------|
| Proteccion contra riesgo de expropiacion                         | 0.856***<br>[0.082]  |
| Residuals First Step   | -0.495***<br>[0.099] |
| Constant   | 2.478***<br>[0.547]  |
| Observations   | 70                   |
| Adjusted R-squared   | 0.679                |
| F test for Residuals First Step                                  | F(1,67) = 25.17***   |
| Note: Run on 30 Nov 2021 using data from input\country-codes.dta |                      |
| Standard errors in brackets                                      |                      |
| *** p<0.01, ** p<0.05, * p<0.1                                   |                      |

## D. Comparación de VI y MCO

Para hacer una comparación sustantiva hicimos dos ejercicios. Primero, compararemos en términos absolutos ambas estimaciones, y en segunda instancia, hicimos una simulación a partir de iteraciones de la muestra obteniendo la distribución de densidad de los estimadores de MCO y las variables instrumentales, y con ello discutiremos algunas de las propiedades de insesgamiento, consistencia y eficiencia.

Primero, notemos que la estimación del efecto del factor institucional sobre la performance económica en Mínimos Cuadrados Ordinarios (*OLS*) es menor que la de Variables Instrumentales ( $\hat{\beta}_{1MCO} = 0.532 < \hat{\beta}_{1IV} = 0.868$ ), resultado consistente con lo reportado por los autores. Esto quiere decir que la estimación por *OLS* planteaba que el efecto institucional era menor del que realmente era, una vez que instrumentamos la calidad institucional por la tasa de mortalidad de los colonos(*logem4*). En ese sentido, se puede plantear que

$$\hat{\beta}_{1MCO} < Cov(avepxr, u) \cdot \hat{\beta}_{1IV}$$

Donde esa correlación entre el factor institucional y el error (su problema de endogeneidad que ya probamos), hacía que el efecto se **subestimara**, por lo que estábamos ante la presencia de un sesgo negativo.

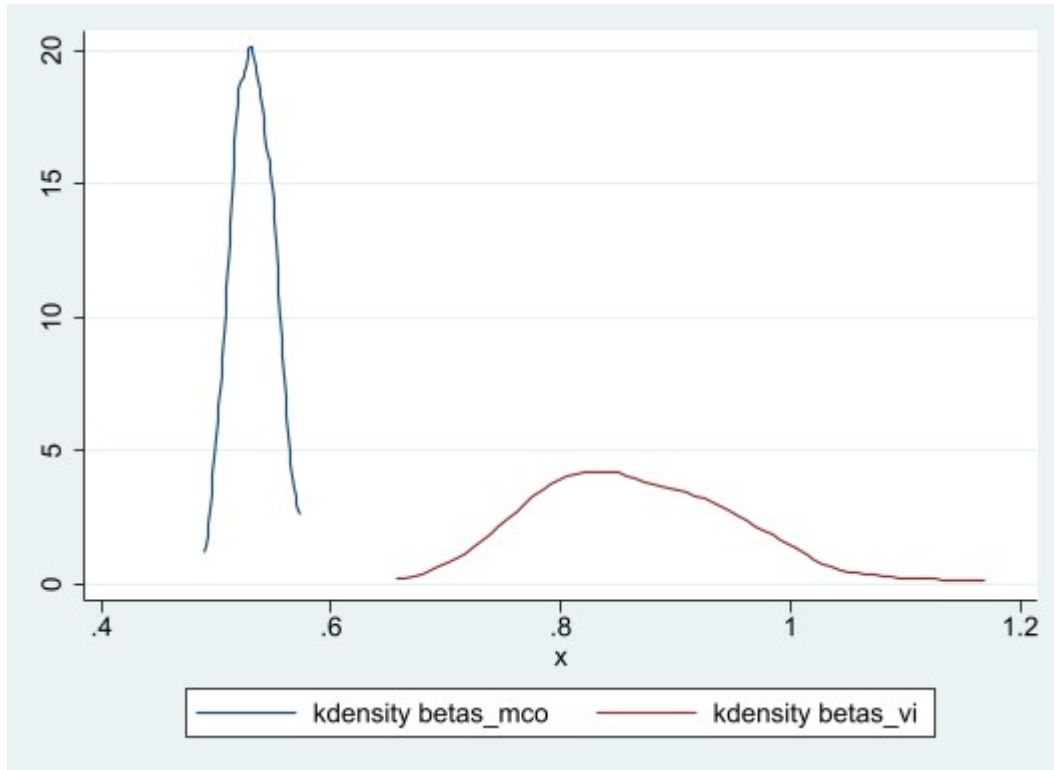


Figura 6: Gráfico de distribución de densidad de los estimadores por MCO y variables instrumentales

Esto es muy relevante en términos de estimación pues el efecto que deberíamos interpretar es el obtenido en variables instrumentales. Como podemos ver en la Figura 6 si bien las estimaciones por variables instrumentales son menos precisas a medida que hacemos variar el tamaño muestral, pero siempre tienen el parámetro estimado. Mientras que en Mínimos Cuadrados Ordinarios, pese a que aumentemos la muestra o la disminuyamos la estimación **siempre será insesgada e inconsistente en presencia de endogeneidad** lo que nos impedirá hacer cualquier tipo de interpretación causal. Podemos reforzar esta idea sobre la distribución de los coeficientes en la *Tabla 3.3* donde podemos ver que los valores de ambas difieren mucho.

Entonces, dado que probamos la endogeneidad de la variable institucional, desde ya las estimaciones *OLS* son insesgadas e inconsistentes. Notemos que si bien no probamos la prueba de Sargan pues solo tenemos un instrumento para una variable, los autores si lo hacen en su artículo demostrando exogeneidad estricta (p.1393)

**Tabla 3.3** Comparación distribución coeficientes de regresión

| $\hat{\beta}$        | Media    | Desviación estandar | Mínimo   | Máximo   |
|----------------------|----------|---------------------|----------|----------|
| $\hat{\beta}_{1MCO}$ | .5319788 | .0175523            | .4900478 | .5723662 |
| $\hat{\beta}_{1IV}$  | .8661319 | .0903656            | .6571248 | 1.169443 |

Así podemos concluir con nuestros dos ejercicios que sí encontramos lo que estábamos esperando al inicio de nuestra investigación. Es probable que la forma en que se mide y ocupa la calidad



institucional contenga un error de medición<sup>4</sup> de esta variable institucional o variables omitidas que producen un *sesgo de atenuación* que hace que el impacto de la protección al derecho de propiedad se subestime. Si volvemos al ejemplo de Nigeria y Chile, por ejemplo, el efecto institucional es mucho más potente que ser 7 veces más un país rico. Por último, nos parece relevante reforzar la importancia de las diferencias institucionales sobre las hipótesis sobre la importancia sobre la performance económica. De manera similar a como hicimos nosotras, los autores aplican una prueba de robustez con iteraciones pero discriminando los grupos según tipo de institucionalidad instalada, mostrando que efectivamente los resultados son significativos. . Esto refuerza el argumento de que inclusive países Africanos son pobres no por su cultura o geografía, sino que por la **existencia de malas instituciones**.

## Diferencias entre los estimadores

¿Por qué razón  $\beta_{MCO} \wedge \beta_{IV}$  pueden diferir? Revisaremos tres casos de análisis y discutiremos su factibilidad, también, de cara al artículo. Lo primero es recordar cuál es el objetivo de construir regresiones por Variables Instrumentales (*IV*) y no seguir con la estimación por Mínimos Cuadrados Ordinarios (*MCO*): existe un problema de endogeneidad en el modelo original estimado que buscaremos corregir a partir del uso de instrumentos. Las causas son muchas, pero las más comunes indicadas en Wooldrige (2017) son:

1. Variable omitida
2. Error de medición, donde los errores clásicos de variable (ECV) son
  - ECV N°1  $Cov(X_1^*, e^o) = 0$
  - ECV N°2  $Cov(X_1, e^o) = \sigma_{e^o}^2$
  - ECV N°3  $Cov(X_1, y) = -\beta_1 \sigma_{e^o}^2$
3. Sesgo de selección (o tratamiento endógeno)
4. Simultaneidad

Básicamente, lo relevante para plantear endogeneidad es que no se cumpla un principio fundamental en el teorema de Gauss Markow, y esto es, que los errores *no dependan* de ningún otro factor observable, y que más bien sean aleatorios y mínimos (ni afectados por problemas en los predictores, en la selección de la muestra, en la medición de las variables, etc.)

Entonces, utilizaremos instrumentos que permitan corregir cualquiera de esos cuatro problemas, esto es, encontrar formas de extraer esa fuente de variación endógena de la regresión, pues si no las estimaciones **son insesgadas e inconsistentes**.

Para responder las preguntas que vienen partiremos de la siguiente base que  $plim(\hat{\beta}_1)$  converge asintóticamente a  $\frac{Cov(ave\text{expr}, \ln(pib))}{Var(ave\text{expr})}$

---

<sup>4</sup>los autores también comentan esta posibilidad

## E. Hipótesis de sesgo por ECV variable dependiente

Se plantea un error de medición ( $e^o$ ) sobre el  $\ln(PIB)$ . Entonces el modelo real sería

$$\ln(PIB_{percapita})_i = \beta_o + \beta_1 \cdot aveexpr_i + \delta \hat{v}_i + u_i + e_i^o$$

En base a la contextualización que hicimos anteriormente, lo lógico sería evaluar qué pasa si este error compuesto ( $u_i + e_i^o$ ) produce un problema de endogeneidad o no. Desarrollamos esto y obtuvimos

$$Var(u_i + e_i^o | aveexpr_i) = Var(u_i | aveexpr_i) + Var(e_i^o | aveexpr_i) + 2Cov(u_i, e_i^o | aveexpr_i)$$

Como no está en juego la propiedad respecto a la independencia del error  $u$  respecto a los predictores, lo único relevante que queda por definir tiene que ver con la relación que tomaría el error de medición con  $aveexpr$ . Por enunciado se plantea error clásico de variable (nosotras tomaremos el *ECV N°2*), esto implica entonces que  $Cov(e_i^o, aveexpr_i) = \sigma_{e^o}^2$  o en otras palabras que el error de medición es independiente del índice institucional. Matemáticamente uno puede demostrar también que eso implica que la relación entre los errores tampoco lo está.

En síntesis, ante un error de medición en la variable dependiente la endogeneidad no se pone en juego si se cumple error clásico de variable  $Cov(e_i^o, aveexpr_i) = \sigma_{e^o}^2$ . En este escenario planteado la estimación será insesgada, asintóticamente es consistente, aunque menos eficiente pues ahora tendremos una varianza mayor

Este resultado tiene sentido con lo presentado en el artículo. Los autores toman otra medición de la performance económica que ocuparon años antes Hall y Jones (1999), y tanto las estimaciones por MCO como las de IV siguieron siendo muy similares (solo cambiaron los errores estándar, mostrando ser más robusta su forma de medición original). Es decir, no es factible pensar que la diferencia de los estimadores entre ambos métodos se deba a esta causa.

Para una demostración detallada seguimos que

$$\begin{aligned} \frac{Cov(aveexpr, \ln(pib))}{Var(aveexpr)} &= \frac{Cov(aveexpr, \ln(pib)^* + e^o)}{Var(aveexpr)} = \\ &= \frac{Cov(aveexpr, \ln(pib)^*)}{Var(aveexpr)} + \frac{Cov(aveexpr, e_o)}{Var(aveexpr)} = \\ &= \beta_1 + \frac{Cov(aveexpr, e_o)}{Var(aveexpr)} \end{aligned}$$

Y como indicamos, por error clásico de variable el último término es cero, entonces el límite en probabilidad del estimador de MCO con error de medición en la variable dependiente sigue siendo consistente e insesgado, por lo que no debería ser distinto que el de variables instrumentales.

## F. Hipótesis de sesgo por ECV en variable independiente

Este argumento es plausible, y no solo se fundamenta matemáticamente sino que también en términos sustantivos. Partamos de la intuición matemática: si la variable de interés tiene un error de medición y ese error de medición no está relacionado a la variable real (*ECV N°1*), y aunque

tampoco a la utilizada (*ECV N°2*) existirá algo que acompaña a la variable dependiente que sesgará su real valor, por lo que las estimaciones de su efecto van a estar “atenuadas” por ese error de medición. Más claramente, si tenemos

$$\begin{aligned} \frac{Cov(aveexpr, \ln(pib))}{Var(aveexpr)} &= \\ \frac{Cov(aveexpr^* + e^o, \ln(pib))}{Var(aveexpr^* + e^o)} &= \\ \frac{Cov(aveexpr^*, \ln(pib)) + Cov(e^o, \ln(pib))}{Var(aveexpr^*) + Var(e^o) + 2Cov(aveexpr^*, e^o)} &= \\ \frac{\beta_1 Var(aveexpr^*)}{Var(aveexpr^*) + Var(e^o)} \end{aligned}$$

Con lo que obtenemos el sesgo de atenuación que habíamos comentado más arriba y que demostramos matricialmente. Notemos que si ese error de medición existe, la estimación por Mínimos Cuadrados Ordinarios es sesgada e inconsistente. Este sesgo de atenuación puede ser el que esté generando que se *subestime* el verdadero efecto que tiene la calidad institucional sobre el PIB.

En términos sustantivos tiene sentido. Pensemos que ese sesgo de medición sobre el indicador institucional implica que se ocupan medidas que recogen distintos elementos que no tienen que ver solo con los derechos de propiedad, sino que también implícitamente instituciones relacionadas a estados de bienestar. En ese sentido, existirá una relación endógena con factores no observados que inciden en la producción (como fortaleza de los sistemas de salud). Pero una vez que tomamos como instrumento las tasas de mortalidad, es probable que esa fuente endógena de variación sea captada por este instrumento, en la medida en que capta también el tipo de colonialismo que se implementó en la época (neo-europeo o extractivista). Solo para cerrar, en ese sentido es probable que este error de medición si pueda ser la causa de la diferencia entre las estimaciones MCO y Variables Instrumentales, algo que también los autores abordan en parte de la discusión sobre la robustez del modelo.

## G. Hipótesis por sesgo de variable omitida

La discusión planteada por la historia es de hecho uno de los puntos que discuten los autores a la hora de discutir la **exogeneidad** del instrumento. Para que se entienda el punto es que, en el caso planteado, esto implicaría que la tasa de mortalidad de los colonos podría ser un predictor relevante en la performance económica. La razón es simple: si las enfermedades del pasado persisten hoy, la tasa de mortalidad del pasado sería un buen predictor del crecimiento poblacional y su evolución, lo que tendría efectos en el crecimiento económico y performance económica de hoy. Entonces, esta tasa de mortalidad podría ser relevante en la estimación del modelo real hoy. Si eso es así, entonces el instrumento **no cumpliría con ser exógeno pues tendría un efecto parcial sobre  $\ln(\text{PIB})$** , y si lo “omitieramos” pasaría a ser parte del término de error de la regresión, y por tanto, a fin de cuentas correlacionaría con este ( $Cov(\logme4, u) \neq 0$ ). Ahora bien, precisamente una de las preocupaciones sobre la *exclusión* tenía que ver con que medio ambientes malos ( que producen enfermedades, más muertes) puede tener un efecto directo sobre la performance económica por su incidencia actual y en las tasas de crecimiento poblacional, y en ese caso, se debería incorporar el efecto de las enfermedades sobre el PIB o inclusive ese factor de las tasas de mortalidad. Ahora bien,

tal como ellos discuten esas muertes tenían que ver con enfermedades que hoy están erradicadas como la malaria y fiebre amarilla, por lo que su efecto en el tiempo de disciparía en el tiempo<sup>5</sup>. En términos de robustez, también se probó esto indicado en los resultados no cambian incluyendo otros controles (p.1372) junto que también se hace un contraste de Sargan y demuestran que el instrumento es exógeno.

---

<sup>5</sup>De hecho, no tiene que ver con la tarea pero a mayores distancias de tiempo, pese a que pueda existir una relación de estas cosas en el tiempo, en Greene (2002, capítulo 19.3.2) se demuestra como estos eventos se discipan en el tiempo.

## Ejercicio 2

El gobierno de Chile está decidiendo si seguir con la **política de etiquetado frontal a productos altos en azúcar**. Para ello quiere saber si esta política ha **reducido el consumo de productos altos en azúcar**. En principio, no es obvio que esto tendrá un efecto en consumo ya que la información nutricional en el dorso del paquete ya existía, por lo que el etiquetado no brinda más información, sino que hace *más visible* la información que ya estaba en el producto.

El contexto: a partir de 2016, Chile aprobó una ley en la cual los productos que tienen más de  $22.5g$  ( $c/100g$ ) de azúcar tienen que incluir una etiqueta que dice “ALTO EN AZUCAR” en su frente.



La base de datos labels.dta tiene información de consumo en supermercados para 100 productos para los años 2014 (antes de la política) y 2018 (después de la política). Además, tenemos una variable que nos dice cuanto azúcar contiene cada uno de estos productos (suponemos que esto no cambio en el tiempo) cada 100g.

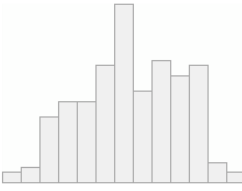
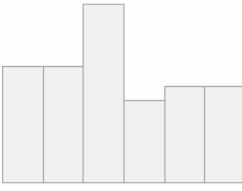
### Variables

- *product*: identificador del producto
- *azucar*:  $\frac{g}{100g}$  de azúcar
- *lnconsumo*: logaritmo del consumo de producto, donde este se mide en miles de dólares ( $\ln(\text{consumo})$ )
- *year*: año
- *post*: marca temporal a la aplicación de la política (0 = antes del 2016; 1 = después del 2016)
- *trat*: productos que tienen más de 22.5 gr. de azúcar cada 100 gr tienen sello (Con sello = 1), en caso contrario sin sello (Sin sello = 0)

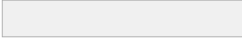
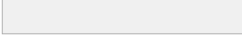
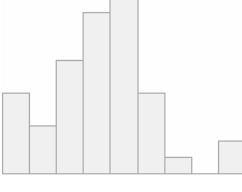
### Descriptivos

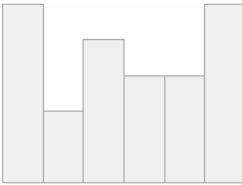
**Tabla 1.** Estadísticos descriptivos para la muestra

| Variable            | Etiqueta | Estadísticas / Valores | Frec. (% sobre válidos) | Gráfico   |
|---------------------|----------|------------------------|-------------------------|---|
| trat<br>[character] |          | 1. Con sello           | 112 (56.0%)             |  |
|                     |          | 2. Sin sello           | 88 (44.0%)              |   |
| post<br>[character] |          | 1. Antes               | 100 (50.0%)             |  |
|                     |          | 2. Despues             | 100 (50.0%)             |   |

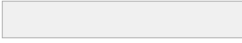
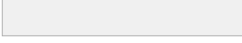
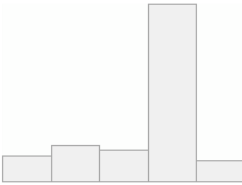
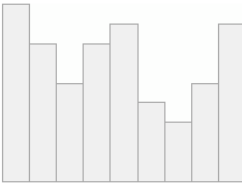
| Variable               | Etiqueta | Estadísticas / Valores  | Frec. (% sobre válidos) | Gráfico   |
|------------------------|----------|---|-------------------------|---|
| lnconsumo<br>[numeric] |          | Media (d-s) : 5.6 (0.5)<br>min < mediana < max:<br>4.2 < 5.6 < 6.7<br>RI (CV) : 0.8 (0.1) | 189 valores distintos   |  |
| azucar<br>[numeric]    | Calorias | Media (d-s) : 24.4 (8.6)<br>min < mediana < max:<br>10 < 23.5 < 40<br>RI (CV) : 13 (0.4)  | 29 valores distintos    |  |

**Tabla 1.1** Estadísticos descriptivos para la muestra para subgrupo: Sin sello, post = Antes



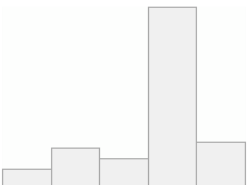
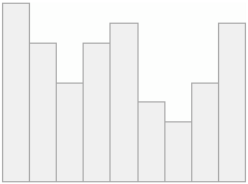
| Variable               | Etiqueta | Estadísticas / Valores  | Frec. (% sobre válidos) | Gráfico   |
|------------------------|----------|---|-------------------------|---|
| trat<br>[character]    |          | 1. Sin sello  | 44 (100.0%)             |  |
| post<br>[character]    |          | 1. Antes  | 44 (100.0%)             |  |
| lnconsumo<br>[numeric] |          | Media (d-s) : 5.3 (0.4)<br>min < mediana < max:<br>4.6 < 5.3 < 6.4<br>RI (CV) : 0.4 (0.1) | 43 valores distintos    |  |

| Variable            | Etiqueta | Estadísticas / Valores  | Frec. (% sobre válidos) | Gráfico   |
|---------------------|----------|---|-------------------------|---|
| azucar<br>[numeric] | Calorias | Media (d-s) : 16.5 (3.8)<br>min < mediana < max:<br>10 < 16.5 < 22<br>RI (CV) : 7 (0.2) | 12 valores distintos    |  |

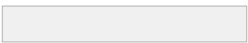
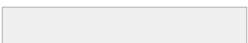
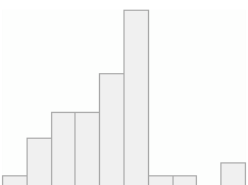
**Tabla 1.2** Estadísticos descriptivos para la muestra para subgrupo: trat = Con sello, post = Antes

| Variable               | Etiqueta | Estadísticas / Valores   | Frec. (% sobre válidos) | Gráfico   |
|------------------------|----------|--|-------------------------|---|
| trat<br>[character]    |          | 1. Con sello   | 56 (100.0%)             |    |
| post<br>[character]    |          | 1. Antes   | 56 (100.0%)             |    |
| Inconsumo<br>[numeric] |          | Media (d-s) : 6 (0.5)<br>min < mediana < max:<br>4.7 < 6.1 < 6.7<br>RI (CV) : 0.4 (0.1)    | 52 valores distintos    |  |
| azucar<br>[numeric]    | Calorias | Media (d-s) : 30.7 (5.6)<br>min < mediana < max:<br>23 < 30.5 < 40<br>RI (CV) : 10.2 (0.2) | 17 valores distintos    |  |

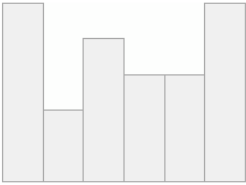
**Tabla 1.3** Estadísticos descriptivos para la muestra para subgrupo: trat = Con sello, post = Despues

| Variable               | Etiqueta | Estadísticas / Valores   | Frec. (% sobre válidos) | Gráfico   |
|------------------------|----------|--|-------------------------|---|
| trat<br>[character]    |          | 1. Con sello   | 56 (100.0%)             |  |
| post<br>[character]    |          | 1. Despues   | 56 (100.0%)             |  |
| lnconsumo<br>[numeric] |          | Media (d-s) : 5.6 (0.5)<br>min < mediana < max:<br>4.2 < 5.7 < 6.3<br>RI (CV) : 0.4 (0.1)  | 56 valores distintos    |  |
| azucar<br>[numeric]    | Calorias | Media (d-s) : 30.7 (5.6)<br>min < mediana < max:<br>23 < 30.5 < 40<br>RI (CV) : 10.2 (0.2) | 17 valores distintos    |  |

**Tabla 1.4** Estadísticos descriptivos para la muestra para subgrupo: trat = Sin sello, post = Despues

| Variable               | Etiqueta | Estadísticas / Valores  | Frec. (% sobre válidos) | Gráfico   |
|------------------------|----------|---|-------------------------|---|
| trat<br>[character]    |          | 1. Sin sello  | 44 (100.0%)             |  |
| post<br>[character]    |          | 1. Despues  | 44 (100.0%)             |  |
| lnconsumo<br>[numeric] |          | Media (d-s) : 5.3 (0.4)<br>min < mediana < max:<br>4.6 < 5.3 < 6.3<br>RI (CV) : 0.5 (0.1) | 43 valores distintos    |  |



| Variable            | Etiqueta | Estadísticas / Valores  | Frec. (% sobre válidos) | Gráfico   |
|---------------------|----------|---|-------------------------|---|
| azucar<br>[numeric] | Calorias | Media (d-s) : 16.5 (3.8)<br>min < mediana < max:<br>10 < 16.5 < 22<br>RI (CV) : 7 (0.2) | 12 valores distintos    |  |

## Pregunta 1

Si bien solo analizaremos el efecto después de la política, estimamos la regresión por Mínimos Cuadrados Ordinarios para la submuestra del año 2014 (antes de la política), 2018 (después de la política) y una general (o conocida en la literatura como *Pooled OLS*) de modo que se vea evidente luego en el resto de nuestro desarrollo que es necesario tomar otra estrategia para estimar el efecto causal de la política. Por ahora mostremos:

Las ecuaciones son las siguientes

$$\ln(\text{consumo})_{i2014} = \beta_o + \beta_1 \cdot \text{trat}_{i2014} + u_{i2014}$$

$$\ln(\text{consumo})_{i2018} = \beta_o + \beta_1 \cdot \text{trat}_{i2018} + u_{i2018}$$

$$\ln(\text{consumo})_i = \beta_o + \beta_1 \cdot \text{trat}_i + \beta_2 \cdot \text{post}(Ao = 2018) + u_i$$

**Tabla 1.5.** Modelos de regresión lineal que estiman log consumo para el periodo anterior y posterior a la política

| Predictores   | Modelo -Antes política | Modelo - Después política | Modelo Pooled        |
|---|------------------------|---------------------------|----------------------|
| Con sello (>22.5 gr azucar) (ref= Sin)                                | 0.655***<br>[0.091]    | 0.307***<br>[0.091]       | 0.481***<br>[0.065]  |
| Año (ref= 2014)   |                        |                           | -0.242***<br>[0.065] |
| Constant  | 5.318***<br>[0.068]    | 5.271***<br>[0.068]       | 5.416***<br>[0.059]  |
| Observations  | 100                    | 100                       | 200                  |
| Adjusted R-squared  | 0.341                  | 0.095                     | 0.250                |
| Note: Run on 30 Nov 2021 using data from input\base_tarea3_labels.dta |                        |                           |                      |
| Standard errors in brackets   |                        |                           |                      |
| *** p<0.01, ** p<0.05, * p<0.1  |                        |                           |                      |

### 1.1 Interpretación

Con un 99% de confianza, podríamos decir que la diferencia promedio esperada en el consumo entre los productos con sello “Alto en Azúcar” y sin sello es de 0.307 para el año 2018, *ceteris paribus*.

Esto significa que en la regresión indicada (*Modelo - Después de la política*) se indica que aquellos productos que contienen la etiqueta se consumen 30.7% más que aquellos que no la contienen.

## Pregunta 2

Este efecto podría parecer poco intuitivo, pues uno podría llegar a mal interpretar dos cosas erróneas: (1) que la política tiene un efecto no esperado en la medida en que hace que el consumo de los productos con sellos con azúcar sea mayor (*efecto no deseado*); o (2) que pese a la política, el consumo sigue siendo alto (*sin efecto*).

La primera opción ya podemos descartar mirando una comparación simple entre el nivel de consumo promedio de productos con azúcar antes y después de la política, y notaremos que este disminuye (*notar que no significa que sea el efecto causal, pero ya nos debe dar intuición que es al menos cuestionable ese punto*). Mientras que el segundo punto no podríamos evaluarlo sin considerar un contrafactual, esto es, evaluar el cambio en el consumo de los productos antes y después de la política, considerando qué hubiese pasado con el consumo si esta no se hubiese implementado.

En síntesis, el resultado encontrado en el punto anterior no tiene sentido, ya que el objetivo de la política es reducir el consumo de productos altos en azúcar, es decir que el incorporar una etiqueta a un producto debiera reducir su consumo. De acuerdo a los resultados encontrados, el hecho de que un producto cuente con una etiqueta aumentaría sus ventas

Ahora bien, no solo intuitivamente podemos notar que difícilmente este es el efecto de la política de los sellos en el consumo de productos altos en azúcares. Para entender mejor la razón plantearemos la formulación del problema de causalidad en términos del modelo de resultados potenciales. Sea  $T_i$  el parámetro que nos permite construir el grupo control y tratamiento en base a los niveles de azúcar

$$T_i = \begin{cases} 0 & \text{Sin sello azucar} < 22.5gr(\text{Control}) \\ 1 & \text{Con sello azucar} > 22.5gr(\text{Tratamiento}) \end{cases} \quad (1)$$

A su vez, tenemos un hito exógeno que es la implementación de la política sobre los sellos que nos permite hacer una marca temporal

$$Post_t = \begin{cases} 0 & \text{Antes } year \leq 2016 \\ 1 & \text{Después } year > 2016(\text{Tratamiento}) \end{cases} \quad (2)$$

Entonces, por construcción para poder saber el cambio en el consumo de azúcares (que medimos en logaritmo del consumo), denotaremos el modelo de la siguiente forma ( $lnconsumo_i^{P_t}$  será su resultado potencial)

$$Resultados_{potenciales} = \begin{cases} lnconsumo_i^0 & T_i = 0 \\ lnconsumo_i^0 & T_i = 1 \\ lnconsumo_i^1 & T_i = 0 \\ lnconsumo_i^1 & T_i = 1 \end{cases}$$

El consumo calculado en el ejercicio anterior **no considera los resultados entre los dos periodos de tiempo**, y por ello el consumo observado para el 2018 es

$$ln(consumo)_i = lnconsumo_{0i}^1 + (lnconsumo_{1i}^1 - lnconsumo_{0i}^1) \cdot T_i$$

Entonces, cuando calculamos el valor esperado del consumo obtenemos el consumo esperado en el 2018 para los productos con sello y sin sello

$$E(\ln consumo_i^1 | T_i = 1) - E(\ln consumo_i^1 | T_i = 0)$$

El problema de esta estimación es que estaremos omitiendo aspectos importantes: primero, cuál era el nivel anterior de consumo de esos productos (su efecto *fijo* como veremos más adelante), y en segunda instancia, si hay otros factores que también podría haber *covariado* con su cambio. Estos dos problemas comentados corresponden al sesgo de selección y sesgo de variable omitida. Como podemos ver algebraicamente, en base al modelo de resultados potenciales deberíamos tener los siguientes valores para estimar un efecto causal

- $E(\ln consumo_i^0 | T_i = 0)$
- $E(\ln consumo_i^0 | T_i = 1)$
- $E(\ln consumo_i^1 | T_i = 0)$
- $E(\ln consumo_i^1 | T_i = 1)$

Algebraicamente podemos obtener las siguientes diferencias

$$[E(\ln consumo_i^0 | T_i = 1) - E(\ln consumo_i^0 | T_i = 0)] - [E(\ln consumo_i^1 | T_i = 1) - E(\ln consumo_i^1 | T_i = 0)]$$

Donde la primera parte de la diferencia corresponde al efecto de la variación en el tiempo ( $c + d$ ), mientras que la segunda el efecto fijo ( $c$ ), con lo cuál podemos obtener el estimador de diferencias en diferencias (por ello son dos diferencias). En el ejercicio propuesto diremos que este es  $d$ .

En términos sustantivos, esta explicación es **muy relevante para entender los procedimientos que seguimos**. Pensemos que el coeficiente que reportamos al inicio podría estar sesgado debido al sesgo de selección que ocurre porque los productos que reciben el tratamiento de la ley de etiquetados son distintos a los que no los reciben. Entonces, esa diferencia de comparación necesita ser controlada en la estimación.

### Pregunta 3

Del modelo de resultados potenciales podemos obtener una forma en modelos de regresión

$$\ln(consumo)_i^t = a + b \cdot T_i + c \cdot Post_t + d \cdot (T_i \cdot Post_t) + u_{it}$$

Hemos descompuesto los efectos antes y después de la política tanto para el grupo de control como tratamiento, ejercicio que podemos precisar dado que tenemos datos panel. Esto nos permitirá observar de manera simple el consumo de productos antes y después de la política de sellos. Como podremos ver en la siguiente tabla tendremos que  $b$  será el efecto fijo (o *between*),  $b + d$   $\wedge$   $a + d$  el efecto tiempo (o *within*) y  $d$  el efecto causal de la política (u *overall*).

**Tabla 3.1** Resumen del modelo de Diferencias en Diferencias para la estimación del cambio en el consumo

| Tiempo $Post_t/$<br>Productos $T_i$ | Sin sello - Control<br>$E(\lnconsumo^0)$ | Con<br>sello-Tratamiento<br>$E(\lnconsumo^1)$ | Diferencia |
|-------------------------------------|--|---|------------|
| Antes $T_i = 0$                     | (a)                                      | (a + b)                                       | (b)        |
| Después $T_i = 1$ )                 | (a + c)                                  | (a + c + b + d)                               | (b + d)    |
| Diferencia                          | (c)                                      | (c + d)                                       | (d)        |

## A. Valor esperado de los productos

Con el resumen de nuestras posibilidades presentamos el valor esperado de los productos tratados y control antes y después del cambio de la ley. Notemos, nuevamente, que se indica “Sin sello” y “Con sello” pues superan los límites indicados por la ley desde 2016. Si superan el valor de 22.5 g/100 g de azúcar, entonces tendrán sello. En caso contrario no. Esto implica discutir las implicancias de estimaciones sesgadas como la de *OLS pooled*. Lo que nosotras vimos hasta antes es que la diferencia *ex post* de los productos sin sello y con sello disminuyeron su consumo en -0.307. Ahora bien notemos que tanto el grupo que no tiene sello como el que si tiene sello disminuyeron su consumo en el tiempo (el primero 0.047 y el segundo 0.395). Por ello, no se puede plantear que ni la diferencia entre los grupos ni la diferencia temporal *constituyan el efecto causal de la política*.

**Tabla 3.2** Valores esperados del logaritmo del consumo antes y después del cambio de ley

| Tiempo $Post_t/$<br>Productos $T_i$ | Sin sello -Control<br>$E(\lnconsumo^0)$ | Con sello -<br>Tratamiento<br>$E(\lnconsumo^1)$ | Diferencia $d$    |
|-------------------------------------|---|---|-------------------|
| Antes $T_i = 0$                     | 5.318 (a)                               | 5.973 (a + b)                                   | -0.652 (b)        |
| Después $T_i = 1$ )                 | 5.271 (a + c)                           | 5.578 (a + c + b + d)                           | -0.307 (b + d)    |
| Diferencia $d$                      | 0.047 (c)                               | 0.395 (c + d)                                   | <b>-0.348</b> (d) |

Más bien, mostraramos que podemos obtener el efecto causal cuando comparamos antes y después del inicio de la política al nivel de consumo de los productos comparando los grupos que ahora tienen sello de aquellos que no. Más adelante mostraremos algunos supuestos para que esta afirmación sea robusta, pero por ahora es evidente que si consideramos tanto la diferencia inicial y posterior tanto de los grupos que tienen sellos por tener mucha azúcar y aquellos que no podremos tener una aproximación *mucho más certera* de cómo afectó esta política al consumo. Este efecto, como ya lo habíamos anticipado corresponde a  $d$  (la diferencia en la diferencia), que en la regresión Difference in Difference consiste el término de interacción entre el tiempo y el tratamiento (lo que a esta altura consiste algo evidente dado que estamos estimando el cambio en el tiempo según cada grupo y si eso es significativo y en qué dirección lo hacen).

## B. Estimación del modelo DiD

Conociendo los cimientos de estos modelos con datos panel que nos permiten hacer análisis respecto a política, diremos primero que el efecto fijo (la diferencia entre los grupos que antes habíamos conocido como sesgo de selección) nos permitirá estimar el efecto de la política a partir de la variación del consumo de los productos con sello y sin sello ( $\Delta_{2018-2014} \lnconsumo | T_i = 1 - \Delta_{2018-2014} \lnconsumo | T_i = 0$ )

## B.1 Presentación de tabla

**Tabla 3.1.** Modelos DiD

| Predictores  | Modelo DiD-Fixed Effects |
|--|--------------------------|
| Con sello (>22.5 gr azucar) (ref= No) -omitted                       | -0.651***                |
| Marca temporal (ref= antes de politica)                              | -0.0470***<br>[0.0086]   |
| Interacción  | -0.3483***<br>[0.0115]   |
| Constant   | 5.6854***<br>[0.0040]    |
| Observations   | 200                      |
| Number of product (group)  | 100                      |
| R-squared  |                          |
| Whitin   | 0.0047000000000000002    |
| Between  | -0.65200000000000002     |
| F(2,98)  | 1358.46***               |
| Note: Run on 1 Dec 2021 using data from input\base_tarea3_labels.dta |                          |
| Standard errors in brackets  |                          |
| *** p<0.01, ** p<0.05, * p<0.1                                       |                          |

## B.2 Interpretación $b$

Primero nos interesa puntualizar que los efectos fijos pueden tener varios componentes, dado que estos representan la heterogeneidad no observada de las observaciones que puede incluir factores como, por ejemplo, qué tipo de bien es el producto  $i$  (en términos económicos, un bien normal, superior, inferior), **sus niveles de azúcar promedio**, su relación comercial exterior, su elasticidad respecto a algún agregado económico o facilidad de disponer de su stock en mercado de bienes. Estos factores, en la medida en que sean más o menos constantes a corto plazo y no los pueda captar a través de datos observables (por ejemplo, cuentas del supermercado) podrían producir diferencias importantes en la línea base de comparación de los productos. El coeficiente  $b$  representa estos efectos fijos, en particular aquellos que representan la diferencia *entre* aquellos productos con etiqueta y aquellos que no. Son las diferencias *sistemáticas entre los grupos* (o variación *between*). No representan un tiempo determinado pues esta diferencia no refiere a un tiempo u otro pues se centra en las diferencias entre los grupos que no varían. Mientras que  $c$  representa el otro efecto fijo, que es la diferencia *dentro* de los grupos en el tiempo. También corresponde a un efecto fijo en la medida en que nos indica un base line de cambio, y ese es constante. Por ejemplo, una elasticidad respecto a su dificultad de poder comprar ciertos productos. Entonces, es una característica no observable, que podemos hipotetizar o intentar calcular matemáticamente y que nos permite decir, en general, como covaría la demanda de un bien respecto a otro factor. Estos dos tipos de efectos fijos nos permiten “limpiar” aún mejor el efecto causal de una política, dado que el sesgo de selección no solo estará controlado por diferencias sistemáticas entre los grupos (por ejemplo, ser un bien necesario para la subsistencia o de consumo en más grandes cantidades en promedio), sino que

también por diferencias sistemáticas dentro del grupo respecto a su cambio (es muy elástico a algún agregado).

Entendiendo cómo los modelos DiD panel nos habilitan para hacer luego ese análisis causal, enfatizamos que **el coeficiente  $b$**  representa el efecto fijo entre los aquellos grupos que tienen sello y aquellos que no tienen sello. El coeficiente estimado nos indica con un 99% de confianza que en promedio el consumo en los productos sin sello es menor que el de aquellos con sello, controlando por el resto de las variables del modelo. De manera más concreta, la diferencia de la esperanza del consumo entre productos sin sellos es de un 65,2% menor que de aquellos con sello.

### B.3 Interpretar $d$

El efecto causal se evidencia tanto en la tabla 3.2 mirando  $d$  pues es la comparación entre los grupos en el tiempo, pero también lo podemos evidenciar en la tabla del modelo de regresión que reportamos. Evidentemente los valores son similares, solo que en la tabla del modelo se reporta más información asociada a cifras significativas, se consideran el cálculo de residuos y de los errores estándar también para hablar de significancia estadística.

De manera concreta, la política de los sellos implementada produce un cambio importante en el consumo y es el esperado. Esto dado que el considerando la diferencia de los grupos y su cambio en el tiempo podemos decir que con la introducción de esta, manteniendo todo el resto constante, el consumo promedio de los productos con sello disminuyó en cerca de un 34,8%. Entonces, el efecto causal de la política es efectiva, toda vez que esta disminución es producto de la instalación de la política, controlando por características “base” de los productos comparativamente e internamente.

Investigaciones recientes sobre el tema llegan a resultados similares. [Barahona, Otero, Otero, & Kim, J. \(2020\)](#) muestran que la ley de etiquetados en Chile tuvo un efecto importante en el consumo de productos poco saludables. En particular ellos estudian la demanda del mercado de los cereales, mostrando que cuando se implementó la política los consumidores sustituyeron estos cereales etiquetados por otros sin etiquetas. Evidenciaron que este cambio se daba de hecho aún más en productos que se pensaba que eran más sanos de lo que eran. De manera mucho más compleja que lo que propusimos aquí, pero ellos también construyen contrafactuales para estimar el cambio de la política de etiquetados, esto es, *qué hubiese pasado con el consumo de los cereales en ausencia de la política*. Entonces, la potencia de estas técnicas parece ser bastante robusta. Pasaremos a discutir entonces porqué no implementar Mínimos Cuadrados Ordinarios.

### B.4 Comparando MCO con DiD

Para comparar de manera sustantiva Mínimos Cuadrados Ordinarios y los modelos de Difference-in-Difference, estimamos ambos modelos: uno general (o *pooled MCO* según Wooldridge(2017)) y el modelo *DiD*. La primera regresión consiste básicamente en que juntamos los años, corremos la regresión por mínimos cuadrados y desde ya nos produce una severa limitación. Nos obliga a suponer que todo efecto fijo no está correlacionado con la variación de una variable de interés. Recordemos la ecuación general por MCO

$$\ln(\text{consumo})_{it} = \beta_o + \beta_1 \cdot T_{it} + v_{it}$$

Entonces ese error  $v_{it}$  no considera que pueden existir una serie de características asociadas a que la política le afecte o no le afecte (por ejemplo, **ser un producto alto en azúcares**). Evidentemente,

el  $\beta_1$  será un coeficiente sesgado dado que existe un **sesgo de heterogeneidad entre los grupos**. Entonces, en los modelos de diferencia en diferencia podemos “limpiar” ese sesgo de diferencia entre los grupos, a su vez de “limpiar” el efecto que hay de la **heterogeneidad en el cambio** (por ejemplo, hay un producto más elástico que otro). Matemáticamente lo podemos plantear así

$$E(v_{it}|T_{it}) = E(a_i + u_{it}|T_{it}) = E(a_i|T_{it}) + E(u_{it}|T_{it})$$

Sabemos que  $E(u_{it}|T_{it}) = 0$ , pero ya hipotetizamos que la correlación entre  $a_i$  y  $T_{it}$  es distinto de cero, por lo que  $E(u_{it}|T_{it}) \neq 0$ . ¿Qué podemos ver en la *Tabla 3.4*? Notemos que las diferencias son *abismales*. Con Mínimos Cuadrados Ordinarios podríamos llegar a concluir que luego de la política de los sellos el nivel de consumo de los productos con sello ahora es mayor, lo que no tendría sentido alguno pues la diferencia promedio en el consumo entre productos con sello respecto a los sin sello “aumentaría en un 48,1%”. El error es que sabemos que eso solo representa el valor esperado considerando todos los tiempos a la vez y sin las diferencias base entre los productos que hoy tienen sellos y aquellos que no. Mientras que si controlamos por como era antes el consumo de estos productos, y como es hoy, junto con las diferencias que hacen que de base tengan niveles distintos, podremos saber que realmente la política disminuyó en cerca de un 34,8% el consumo de productos poco saludables. En síntesis, este sesgo de heterogeneidad en el consumo de productos saludables y no, antes y después de la política representa un impacto importante al punto tal que debemos desconfiar de las simples estimaciones por MCO.

**Tabla 3.4** Comparación de Pooled OLS y DiD models

| VARIABLES   | Modelo OLS Pooled    | Modelo DiD - Fixed Effect | ...4 |
|---|----------------------|---------------------------|------|
| Con sello (>22.5 gr azucar) (ref= No)                                 |                      | -0.651***                 |      |
| Marca temporal (ref= antes de politica) = 1                           |                      | -0.0470***<br>[0.0086]    |      |
| Interacción   |                      | -0.3483***<br>[0.0115]    |      |
| Con sello (>22.5 gr azucar) (ref= No)                                 | 0.481***<br>[0.065]  |                           |      |
| Año (ref = 2014)  | -0.242***<br>[0.065] |                           |      |
| Constant  | 5.416***<br>[0.059]  | 5.685***<br>[0.004]       |      |
| Observations  | 200                  | 200                       |      |
| Adjusted R-squared  | 0.250                | 0.929                     |      |
| Whitin  |                      | 0.0047000000000000002     |      |
| Between   |                      | -0.6520000000000000002    |      |
| F(2,98)   |                      | 1358.46***                |      |
| Number of product   |                      | 100                       |      |
| Note: Run on 30 Nov 2021 using data from input\base_tarea3_labels.dta |                      |                           |      |
| Standard errors in brackets   |                      |                           |      |
| *** p<0.01, ** p<0.05, * p<0.1  |                      |                           |      |

## B.5 Supuestos y Robustez de DiD

Ahora bien, como anunciamos requerimos de algunos supuestos para poder afirmar que el efecto estimado por DiD es causal. Para ello propondremos a su vez ciertas pruebas que se pueden aplicar

1. *Supuestos de MCO*: al igual que en Mínimos Cuadrados Ordinarios debemos cumplir al menos los cinco puntos del Teorema de Gauss-Markow para que los estimadores no sean sesgados. Entre ellos, la única diferencia tiene que ver con la esperanza de los errores idiosincráticos. En este caso, como estamos estimando cambios esperamos que el cambio en el error, dado el cambio alguno de los parámetros de nuestra regresión sea cero ( $E(\Delta u_i | \Delta X' i)$ ), donde  $u$  denota el error idiosincrático (factores no observados, error aleatorio) y  $X'$  el vector de regresores. El supuesto que nos permite afirmar homocedasticidad de los errores es algo que podríamos poner en duda dada la correlación serial de los errores, y por tanto también asumiremos que esto también se cumple de modo de no tener problemas con la *inferencia estadística*.
2. **Tendencias paralelas**: es el supuesto más importante en estos modelos. Lo que plantea es que el resultado potencial (en este caso el logaritmo del consumo de los productos) es igual en tratados (con sello) y no tratados (sin sello) antes del tratamiento. Es decir,

$$(lnconsumo_i^0 | T_i = 1) = (lnconsumo_i^0 | T_i = 0)$$

$$E(lnconsumo_i^0 - lnconsumo_i^1 | T_i = 1) = E(lnconsumo_i^0 - lnconsumo_i^1 | T_i = 0)$$

Este supuesto es relevante pues lo que nos dice es que en ausencia de tratamiento, las unidades de observación (los productos) hubiesen tenido el mismo cambio. Así, esta forma nos permite no solo clarificar que el efecto no se debe a otras tendencias, sino que también construir un contrafactual para estimar el cambio efectivo de la política (es decir, nos pone una “línea base” o *baseline* como aparece en la literatura).

## Robustez

1. *Placebo*: una forma de chequear el último supuesto es agregar más periodos anteriores, aplicar el modelo de diferencias en diferencias y encontrar que las diferencias entre los grupos no son significativas.
2. *Más grupos de control*: permite controlar el efecto fijo dado que si tomo dos grupos y su *baseline* no es lo suficientemente parecida, puede ocurrir que existe endogeneidad por variables omitidas. Por ello, una buena práctica es tomar más unidades para grupo de control y verificar si existen diferencias en las estimaciones (por ejemplo, con pruebas de exclusión como *Jackknife*).
3. *Efectos tendencias*: se puede probar con otras variables posibles efectos *espúreos*. Por ejemplo, podríamos tomar otras covariables como los precios de los productos para ver si estas también son afectadas por la política. Si no se vieron afectadas producto del cambio, entonces no hay problemas. En caso contrario deberíamos incorporarlas como regresores ( $X'_{itj}a$ ) lo que sin duda mejorará la precisión al controlar potenciales diferencias entre los grupos o efectos de otras variables.
4. **Interacciones**: una forma interesante de interpretación es tomar las interacciones, con lo que podremos conocer *en qué momento los grupos comenzaron a diferir* y si esto tiene sentido. Estos modelos, los *DDD* resultan muy interesantes a la hora de hacer predicciones de en cuánto tiempo demora la política influir y en qué tasa de cambio.



## Referencias

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5), 1369-1401

Acemoglu, D. (1995). Reward structures and the allocation of talent. *European Economic Review*, 39(1), 17-33.

Barro, R. J. (2001). Human capital and growth. *American economic review*, 91(2), 12-17.

Barahona, N., Otero, C., Otero, S., & Kim, J. (2020). Equilibrium effects of food labeling policies. Available at SSRN.

Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others?. *The quarterly journal of economics*, 114(1), 83-116.

# Apéndice

## Código en STATA Ejercicio 1

```
//=====//
//      Tarea N°3 - Parte 1
//      Econometria I - PUC
//      Author: Valentina Andrade
//      Last update: 27-11-2021
//=====//

// 0. Preliminars
clear all
* Crear el log file
// log using "output\log_assignment1.smcl", replace
*Instalar asdoc
net install asdoc, from("http://fintechprofessor.com") replace // Para exportar tablas
ssc install outreg2 // Tablas a latex
** net install spost13, from("https://jslsoc.siteshost.iu.edu/stata/") replace // Para lincom
set scheme s2color, permanently //Tema de graficos

// a. Directorio
global data_row = "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3\output\data"
global data_prepared "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3\output\data_prepared"
global latex = "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3\output\latex"
cd "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3"

// b. Leer datos
use "input\countrycodes", replace
use "input\base_tarea3_colonial", replace
// merge 1:1 country using "base_tarea3_colonial"
use "input\country-codes", replace

// 0. Procesamiento
// a. Tabla resumen de toda la base
summarize
l country

// b. Crear variables

* 1. Para filtrar etiquetas (country_d)
gen country_d = country if country == "CHL" | country == "NGA" //
```

```

* 2. Para crear variable otros (a diferencia del paper no hay como distinguir de America asi q
gen otros = 1 if africa == 0 & asia == 0 & rich4 == 0
replace otros = 0 if missing(otros)

* 3. Continente (proviene del merge)
    l continent
    egen cont = group(continent), label

**Nota: contiene a America

// c. Etiquetar
    * Se hace para las tablas en latex despues
// label var number "Identificador numérico del país" No esta
label var country "Pais"
label var avexpr "Proteccion contra riesgo de expropiacion"
label var logpgp95 "Log(PBI per capita) en PPP"
label var rich4 "Países neo-europeos"
label var logem4 "Log(mortalidad colonos)"
label var lat_abst "Latitud"
label var africa "Pais africano"
label var asia "Pais asiático"
label var loghjypl "log PIB por trabajador (Hall& Jones)"
label var baseco "muestra del paper original"
label var otros "Otros"
label var countryname "Nombre pais"
label var cont "Continente"

// Pregunta 1 -----
** 1.3 Graficos
    twoway (scatter logpgp95 avexpr, mlabel(country_d) mlabsize(small) mfcolor(red) mlabcolor(black))
    graph export "output/figure_modelo1_mco.jpg"

** 1.2 Estimacion modelos

*** Modelo simple - MCO
    regress logpgp95 avexpr //tablas en latex abajo
    estimate store modelo1_mco //ocuparlo despues
    estat vce // Para ver su matriz de correlaciones estimada

//C.2

// 1. Solo algunos paises
    twoway (scatter logpgp95 avexpr if africa == 1) (scatter logpgp95 avexpr if rich4 == 1) (s
    graph export "output/figure2_modelo1_.jpg"

```

```

// Regresion zonas

regress logpgp95 avexpr africa asia rich4 lat_abst
estimates store modelo1_vozone

// 2. Con todos - datos adicionales
    twoway (scatter logpgp95 avexpr if continent == "Africa") (scatter logpgp95 avexpr if c
graph export "output/figure3_modelo1_vocontinent.jpg"

// Modelo (se ocupa cont para que sea factor)
**Nota: como en el paper definimos "America" como categoria de referencia. Oceania es como "ot
regress logpgp95 avexpr ib2.cont lat_abst
estimates store modelo1_vocontinent

// Pregunta 2.

* A. Condiciones de instrumentos

* Exogeneidad
** Solo discutir exploratoriamente

* Relevancia
** Agregaremos un grafico tal como lo hicieron los autores (correlacion)
    twoway (scatter avexpr logem4, mlabel(country) mlabsize(tiny) mfcolor(green) mlabcolor(black%)
graph export "output/figure4_2ls_fs_relevance.jpg"

// B. 2SLS Manual

** B.1 First Step
regress avexpr logem4
estimates store modelo_2sl2_fs

*** Predecir avexpr (paso 2) y residuos (test de Haussman)
predict avexpr_p, xb
predict v_resid, residuals

// B.2 Relevancia
test logem4

// B.3 Second step
regress logpgp95 avexpr_p
estimates store modelo_2sl2_ss

// C. Interpretación
ivregress 2sls logpgp95 (avexpr = logem4), first

```

```

estimates store modelo_2sls2_first

ivregress 2sls logpgp95 (avexpr = logem4)
estimates store modelo_2sls2_second

** Graficos
    estimates restore modelo_2sls2_second
    coefplot, mlabel mlabcolor(black%96) mlwidth(medthick) mlabposition(12) xline(0) xtitle(0)
    graph export "output/figure5_2sls.jpg"

// Pregunta 3.

// A. Test Haussman
hausman modelo_2sls2_second modelo1_mco, alleqs
//esttab mco_true mco_endogena vi using "output/haussman.xls", ///

// B. Interpretación

// C. Haussman manual
* Ya habíamos predicho los residuos de la regresión de la primera etapa
tab v_resid
regress logpgp95 avexpr v_resid
estimates store modelo_hausman

// Significancia
test v_resid

// D. Comparacion de coeficientes
estimates restore modelo_2sls2_second
return list
matrix list r(table)
matrix beta_iv = r(table)[1,1]

**estimates restore modelo1_mco
**return list
**matrix list r(table)
**matrix beta_mco = r(table)[1,1]

**A. MCO
// 200 muestras aleatorias de tamaño 70
matrix betas_mco = (0,0)
forvalues i = 1/70{
    preserve
    quietly sample 70
    quietly regress logpgp95 avexpr
    matrix beta = (`i', r(table)[1,1])
    matrix betas_mco = (betas_mco\beta)
}

```

```

    restore
}

**B. VI
matrix betas_vi = (0,0)
forvalues i = 1/70{
    preserve
    quietly sample 70
    quietly ivregress 2sls logpgp95 (avexpr = logem4)
    matrix beta = (`i', r(table)[1,1])
    matrix betas_vi = (betas_vi\beta)
    restore
}

** C. Comparar
clear
matrix betas = (betas_mco, betas_vi)
svmat betas
drop if betas2 == 0
drop betas1 betas3
rename betas2 betas_mco
rename betas4 betas_vi
sum betas_mco betas_vi
twoway(kdensity betas_mco, n(70)) (kdensity betas_vi, n(70))
graph export "output/figure6_simulacion_vi_mco.jpg"
** mco sesgadas, vi poca eficiencia pero contiene el parámetro
sum betas_mco betas_vi

// Apendice
//I ----Tablas

// Pregunta 1 MCO
*Tablas:
foreach var in 1_mco {
    outreg2 [ modelo`var'] ///
    using "output/tab/tabla1_`var'.tex", ///
    label /*Las variables del modelo aparecen con etiquetas*/ ///
    tex(frag) /*Permite ingresar directamente el archivo generado al tex*/ ///
    dec(3) /*Agrega 3 decimales*/ ///
    adj /*agrega el r2 ajustado*/ ///
    addtext("Note: Run on $$_DATE, using data from $$_FN") /*Produce el texto en la tabla*/ //
    replace
}

```

```

*Tablas excel

foreach var in 1_mco {
    outreg2 [ modelo`var'] ///
    using "output/tab/tabla1_`var'.xls", ///
    label /*Las variables del modelo aparecen con etiquetas*/ ///
    tex(frag) /*Permite ingresar directamente el archivo generado al tex*/ ///
    dec(3) /*Agrega 3 decimales*/ ///
    adj /*agrega el r2 ajustado*/ ///
    bracket /*Genera paréntesis de corchetes*/ ///
    addtext("Note: Run on $$_DATE using data from $$_FN") /*Produce el texto en la tabla*/ ///
    replace
}

// Modelos adicionales
foreach var in 1_ {
    outreg2 [ modelo`var'mco modelo`var'vozone modelo`var'vocontinent ] ///
    using "output/tab/tabla1.1.xls", ///
    label /*Las variables del modelo aparecen con etiquetas*/ ///
    tex(frag) /*Permite ingresar directamente el archivo generado al tex*/ ///
    dec(3) /*Agrega 3 decimales*/ ///
    adj /*agrega el r2 ajustado*/ ///
    bracket /*Genera paréntesis de corchetes*/ ///
    addtext("Note: Run on $$_DATE using data from $$_FN") /*Produce el texto en la tabla*/ ///
    replace
}

// Pregunta 2

// VI manual
foreach var in 2sl2_ss 2sl2_fs {
    outreg2 [ modelo_`var' ] ///
    using "output/tab/tabla_`var'.xls", ///
    label /*Las variables del modelo aparecen con etiquetas*/ ///
    tex(frag) /*Permite ingresar directamente el archivo generado al tex*/ ///
    dec(3) /*Agrega 3 decimales*/ ///
    adj /*agrega el r2 ajustado*/ ///
    bracket /*Genera paréntesis de corchetes*/ ///
    addtext("Note: Run on $$_DATE using data from $$_FN") /*Produce el texto en la tabla*/ ///
    replace
}

// VI real
outreg2 [ modelo_2sls2_first modelo_2sls2_second ] ///
using "output/tab/tabla_modelo_2sls2.xls", ///
label /*Las variables del modelo aparecen con etiquetas*/ ///
tex(frag) /*Permite ingresar directamente el archivo generado al tex*/ ///

```

```

dec(3) /*Agrega 3 decimales*/ ///
adj /*agrega el r2 ajustado*/ ///
bracket /*Genera paréntesis de corchetes*/ ///
addtext("Note: Run on $$_DATE using data from $$_FN") /*Produce el texto en la tabla*/ ///
replace

// Pregunta 3
outreg2 [ modelo_hausman ] ///
using "output/tab/tabla_modelo_hausman.xls", ///
label /*Las variables del modelo aparecen con etiquetas*/ ///
tex(frag) /*Permite ingresar directamente el archivo generado al tex*/ ///
dec(3) /*Agrega 3 decimales*/ ///
adj /*agrega el r2 ajustado*/ ///
bracket /*Genera paréntesis de corchetes*/ ///
addtext("Note: Run on $$_DATE using data from $$_FN") /*Produce el texto en la tabla*/ ///
replace

log close

```

## Código STATA ejercicio 2

```

//=====//
//      Tarea N°3 - Parte 1
//      Econometria I - PUC
//      Author: Valentina Andrade
//      Last update: 27-11-2021
//=====//
// 0. Preliminars
clear all
* Crear el log file
// log using "output\log_assignment1.smcl", replace
*Instalar asdoc
net install asdoc, from("http://fintechprofessor.com") replace // Para exportar tablas
ssc install outreg2 // Tablas a latex
** net install spost13, from("https://jslsoc.siteshost.iu.edu/stata/") replace // Para lincom
set scheme s2color, permanently //Tema de graficos

// a. Directorio
global data_row = "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3"
global data_prepared "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3"
global latex = "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3\output"
cd "C:\Users\Valentina Andrade\Documents\GitHub\master\econometrics\reporte3"

```



```

// b. Leer datos
    use "input\base_tarea3_labels", replace

// 0. Procesamiento
// a. Tabla resumen de toda la base
    summarize

// b. Crear variables
    * 1. Para construir hito exogeno
gen post = 1 if year > 2016
replace post = 0 if missing(post)
    *1.1 Otros lags temporales para testear robustez(no hay)

    * 2. Para construir grupo control y tratamiento
gen trat = 1 if azucar > 22.5
replace trat = 0 if missing(trat)

//c. Etiquetas

    label var trat "Con sello (>22.5 gr azucar) (ref= No)"
    label var post "Marca temporal (ref= antes de politica)"
    label var lnconsumo "logaritmo del consumo de producto (miles USD)"

    label var azucar "1/100 gr azucar"
    label var year "anio"
    label var product "Identificador del producto"

// Pregunta 1. Estimacion MCO
foreach var in 2014 2018 {
    regress lnconsumo trat if year == `var'
    estimates store modelomco_`var'
}

**Adicional: Estimacion pooled
    regress lnconsumo trat i.year
    estimates store modelomco_pooled

**avplot trat

// Pregunta 2. (Explicaciones)

// Pregunta 3

/// A. Valor esperado
** E(lnconsumo| trat, antes de la política) y E(lnconsumo| trat, despues de la política)
** collapse (mean) lnconsumo, by(trat post)

```

```

** list

/// B. Estimacion DiD

/// 1. Especificacion
tsset product year

/// 2 Estimacion
** Estimacion con variables constantes o poco cambios
*** Imaginar que es var2 (sexo)
*** Si lo omite es completamente no habia nada de cambio y lo hace por colinealidad

xtreg lnconsumo o.trat##i.post,fe // unary operator to omit a variable or indicator (ante pro
estimates store modelo_did

*** Podrian ser poco significativa pues es facil que tenga mucha correlacion las estimaciones

*** Interpretacion: no olvidar que los coeficientes son de variacion/comparacion en el tiempo

// DiD ---- Otra opcion de codigo
**bysort product: egen treated = mean(azucar)
**replace treated = 1 if treated > 22.5

**generate post2 =0
**replace post2 = 1 if year > 2016

**didregress (lnconsumo treated post2) (azucar), group(product) time(post) nogteffects

// Apendice-----

/// Tablas

//// Pregunta 1
outreg2 [ modelomco_2014 modelomco_2018 modelomco_pooled] ///
using "output/tab/parte2/tabla1_mco.xls", ///
label /*Las variables del modelo aparecen con etiquetas*/ ///
tex(frag) /*Permite ingresar directamente el archivo generado al tex*/ ///
dec(3) /*Agrega 3 decimales*/ ///
adj /*agrega el r2 ajustado*/ ///
bracket /*Genera paréntesis de corchetes*/ ///
addtext("Note: Run on $$_DATE using data from $$_FN") /*Produce el texto en la tabla*/ ///
replace

// Pregunta 3

```

```

// B.2
    outreg2 [ modelo_did]   ///
    using "output/tab/parte2/tabla2_did.xls", ///
    label /*Las variables del modelo aparecen con etiquetas*/   ///
    tex(frag) /*Permite ingresar directamente el archivo generado al tex*/   ///
    dec(4) /*Agrega 3 decimales*/ ///
    adj /*agrega el r2 ajustado*/ ///
    bracket /*Genera paréntesis de corchetes*/   ///
    addtext("Note: Run on $$DATE using data from $$FN") /*Produce el texto en la tabla*/   ///
    replace

// B.3 Hacer grafico para ver tendencia

// B.4
    outreg2 [ modelomco_pooled modelo_did]   ///
    using "output/tab/parte2/tabla3_mco_did.xls", ///
    label /*Las variables del modelo aparecen con etiquetas*/   ///
    tex(frag) /*Permite ingresar directamente el archivo generado al tex*/   ///
    dec(3) /*Agrega 3 decimales*/ ///
    adj /*agrega el r2 ajustado*/ ///
    bracket /*Genera paréntesis de corchetes*/   ///
    addtext("Note: Run on $$DATE using data from $$FN") /*Produce el texto en la tabla*/   ///
    replace

```