

# Evaluación de Impacto - TP1

Leonardo A. Caravaggio

## 1. Experimentos aleatorios

### 1.1. Intervenciones para reducir la corrupción

En esta subsección se analiza el trabajo de [Olken \(2007\)](#). Se trata de un estudio de campo sobre los efectos del control local o externo en la ejecución de una política pública de construcción de carreteras en Indonesia.

- A. ¿Cuáles son los tratamientos?

Se evalúan dos posibles soluciones al problema de la corrupción. El aumento del control por parte del estado y por parte de la propia gente beneficiaria de la política pública. Es decir que estos dos son los tratamientos.

El primer tratamiento se asigna informando a los pueblos que tenían asignado el presupuesto, antes de que las obras inicien, que el proyecto iba a ser monitoreado por una agencia del gobierno central (elevando la probabilidad de ser monitoreados de 4 % al 100 %).

El segundo tratamiento se asigna de dos maneras diferentes: creando “reuniones de contabilidad” en las que el pueblo evalúa la política, y con un formulario anónimo.

- B. ¿Cómo se construye la variable dependiente?

La corrupción se mide como la diferencia entre el costo estimado por ingenieros una vez que se concluye la obra y el costo realmente gastado de acuerdo a lo informado por el pueblo. Es decir, una suerte de dinero perdido.

- C. ¿Cuál es la unidad de análisis y a qué nivel se asignan los distintos tratamientos?

La unidad de análisis es el pueblo que lleva adelante la obra y que recibe uno u otro tratamiento, asignado en forma aleatoria.

- D.1 Presenten evidencia de que tratados y controles están balanceados en las siguientes variables: *zkadesedyears*, *zkadesage*, *zpop*, *zpercentpoorpra*, *zdistancekec*, y *podeszhill* ¿Qué concluimos?

En el Cuadro 1 puede observarse que las medias son muy similares para todas las variables consideradas. Solo el caso de *zpercentpoorpra* muestra un p-valor menor a 0.5 para la prueba de diferencia de medias.

- D.2 Muestren la correlación entre estas variables y el tratamiento ¿Cómo se explican estos resultados?

Cuadro 1: Análisis de Diferencia de Medias

	mu_1	mu_2	P-Value
zkadesedyears	11.33956	11.51418	.4290202
zkadesage	43.45342	43.98221	.4272319
zpop	4.472738	4.272557	.3534819
zpercentpoorpra	.3953512	.4302302	.040185
zdistancekec	5.40679	5.43617	.9351882
podeszhill	.3540373	.4255319	.0721581

Nota: mu\_1 y mu\_2 representan la media para el caso de los tratados y no tratados respectivamente. Se presenta también el P-Value de la prueba de diferencia de medias, donde  $p < 0.05$  implica medias estadísticamente equivalentes.

Cuadro 2: Regresiones lineales

	Coefficiente	P-Value
zkadesedyears	.0059618	.4275603
zkadesage	.001984	.4273137
zpop	-.0071212	.3191312
zpercentpoorpra	.2009709	.0388413
zdistancekec	.000373	.9350604
podeszhill	.0749827	.0726362

Nota: Se presentan los coeficientes de la estimación lineal entre las variables seleccionadas y el tratamiento. Se presenta también el P-Value de los mismos.

En el Cuadro 2 se observa que solamente *zpercentpoorpra* tiene una relación lineal significativa con el tratamiento a un nivel de 0.05.

- D.3 Regresen el tratamiento en función de las variables de los puntos anteriores ¿Qué concluimos mirando el estadístico F?

El Cuadro 3 muestra el estadístico F y el correspondiente p-valor de la regresión entre *audit* y todas las variables propuestas. Un p-valor mayor a 0.05 indica que no es posible rechazar la hipótesis nula de la variable dependiente no es explicada en conjunto por los regresores.

Cuadro 3: Prueba F

	F	P-Value
r1	1.619589	.1392169

Nota: Se presenta el estadístico F de la regresión.

- Combinen esta base con *jperoaddata.dta* (con el comando *merge*). El indicador de las villas es *desaid*. Esta nueva base contiene las variables dependientes que usa Olken. Realicen las siguientes regresiones usando *Indiffeall4mainancil* como variable dependiente:
- E.1 Una regresión solo con *audit* como variable independiente. Expliquen su elección respecto a los errores estándar e interpreten el coeficiente de *audit*.

Se utilizan errores robustos porque se supone heterocedasticidad. En el Cuadro 4 se observa el coeficiente de la regresión planteada.

Cuadro 4: Regresión lineal

	Coef	P-Value
r1	-.0917645	.0022987

Nota: Se presenta el coeficiente y el P-Value de la regresión simple con errores estándar robustos.

- E.2 Una regresión que, además de *audit*, incluya como variables de control las que usaron para analizar el balance entre tratados y controles. Expliquen qué ocurre con el coeficiente de *audit* y por qué.

Al incluir controles, tanto el coeficiente como el p-valor del tratamiento *audit* apenas se modifican. Esto puede verse en el Cuadro 5.

Cuadro 5: Controles

	Coeficiente	P-Value
<i>audit</i>	-.0748141	.0141881
<i>zkadesedyears</i>	-.0058537	.3587619
<i>zkadesage</i>	-.0037133	.0475869
<i>zpop</i>	.0069239	.2605051
<i>zpercentpoorpra</i>	-.0506758	.4685071
<i>zdistancekec</i>	-.0014615	.6893502
<i>podeszhill</i>	-.0524245	.1175045

Nota: Se presentan los coeficientes de la estimación lineal y también el P-Value de los mismos.

- E.3 Agreguen a la regresión del punto e.ii efectos fijos por enumerador (*z7enumcode1*) ¿Cuál es la variación que identifica nuestro coeficiente de interés? ¿Es correcta esta especificación? Justificar.

Posteriormente se realiza un ejercicio incluyendo efectos fijos por enumerador. La salida de esta especificación puede observarse en el Cuadro 6. En realidad, se habla de efectos fijos cuando se trata de Datos de Panel ([Hansen, 2014](#), p. 327). En este caso sería simplemente un control por cada enumerador, donde al menos uno de ellos será eliminado por multicolinealidad.

- E.4 Agreguen a la regresión del punto e.ii efectos fijos por subdistrito (*kecnum*) ¿Cuál es la variación que identifica nuestro coeficiente de interés? ¿Es correcta esta especificación? Justificar.

El resultado de la salida de la regresión con efectos fijos por distrito puede verse en el Cuadro 7. El comentario del punto anterior es válido también aquí.

- Olken encuentra que el tratamiento *audit* afecta cómo las conexiones con autoridades de las villas o jefes de proyectos llevan a efectos positivos sobre el empleo (pp. 226-229). Parecería que se da un efecto sustitución de una forma de corrupción directa por otra más parecida al nepotismo. Podríamos pensar, entonces, que las posibilidades de este efecto sustitución podrían moderar el efecto de *audit*. Es decir, este tratamiento podría mostrarse más efectivo (sobre la variable dependiente que venimos usando) en aquellos lugares donde haya

Cuadro 6: Efectos Fijos

	(1)
	Efectos Fijos por Enumerator
audit	-0.0756** (0.0293)
zkadesedyears	-0.00833 (0.00629)
zkadesage	-0.00266 (0.00187)
zpop	0.00566 (0.00569)
zpercentpoorpra	-0.0390 (0.0700)
zdistancekec	-0.00140 (0.00351)
podeszhill	-0.0227 (0.0358)
2.z7enumcode1	-0.00314 (0.0746)
3.z7enumcode1	0.417*** (0.0751)
4.z7enumcode1	0.0242 (0.0873)
5.z7enumcode1	-0.0862 (0.0801)
6.z7enumcode1	0.0350 (0.0760)
7.z7enumcode1	0.0348 (0.0746)
8.z7enumcode1	0.0173 (0.0611)
9.z7enumcode1	-0.0284 (0.0680)
10.z7enumcode1	0.165** (0.0736)
11.z7enumcode1	0.213*** (0.0719)
12.z7enumcode1	0.0154 (0.0576)
Constant	0.449*** (0.143)
Observations	528
R-squared	0.138

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Nota: Se presentan los coeficientes de la estimación lineal incluyendo efectos fijos por enumerador y errores estándar robustos. También se presenta el P-Value de los coeficientes.

Cuadro 7: Efectos Fijos por Subdistrito

	Coeficiente	P-Value
audit	.2397418	.090665
zkadesedyears	-.0058008	.4063959
zkadesage	-.0028151	.1534786
zpop	-.0082475	.2205715
zpercentpoorpra	-.105686	.1491791
zdistancekec	-.003291	.4704185
podeszhill	-.0309567	.5282209

Nota: Se presentan los coeficientes de la estimación lineal incluyendo efectos fijos por subdistrito y errores estándar robustos. También se presenta el P-Value de los coeficientes. Por una cuestión de espacio no se reportan los coeficientes de los clusters.

más posibilidades para realizar esta sustitución. La base ties.dta contiene dos variables para medir las oportunidades de sustitución: famprojhead (proporción de la villa con conexiones familiares con el líder del proyecto) y famvilgovt (proporción de la villa con conexiones familiares con el gobierno de la villa). Combinen ties.dta con la base que usaron en las secciones anteriores.

- F. Planteen un modelo de regresión con el cual podamos testear estos moderadores. No se preocupen por incluir controles en esta especificación. Expliquen por qué este modelo sería adecuado y qué esperaríamos encontrar en términos del signo de los efectos estimados si este efecto sustitución fuese cierto. Presenten los resultados en una nueva tabla.

Si el efecto sustitución planteado fuera correcto, los coeficientes de las covariables de la regresión planteada tendrían que haber resultado significativos, e incluso haber reducido o eliminado el efecto del coeficiente del tratamiento. Esto no es así, como se observa en el Cuadro 8.

Cuadro 8: Efecto Sustitución

	Coeficiente	P-Value
audit	-.0916749	.0023338
famprojhead	.0302641	.8272932
famvilgovt	.0751831	.1896139

Nota: Se presentan los coeficientes de la estimación lineal con errores estándar robustos. También se presenta el P-Value de los coeficientes.

El Cuadro 9 muestra la media de cada una de las variables propuestas en función de los valores de la variable de tratamiento para el caso de las invitaciones. En todos los casos las medias son estadísticamente equivalentes.

El Cuadro 10 muestra la salida de la regresión utilizando el tratamiento de invitaciones. El mismo presenta un efecto no significativo.

- G.1 ¿Cuál es el MDE de esta intervención para un nivel de significatividad de 5 % y un power de 80 %?

El Minimum Detectable Effect (MDE) de esta intervención para un nivel de significatividad de 5 % y un power de 80 % es de 0,22. Para obtener un MDE de 0.1 desvíos estándar con un nivel de significatividad de 5 % y un power de 80 % sería necesaria una muestra de 3136 observaciones.

Cuadro 9: Análisis de Diferencia de Medias

	mu_1	mu_2	P-Value
zkadesedyears	11.45894	11.40152	.8045801
zkadesage	43.09662	44.01515	.189099
zpop	4.218797	4.463	.2817888
zpercentpoorpra	.4047759	.4150721	.5649464
zdistancekec	5.551691	5.352381	.5998928
podeszhill	.3864734	.3879093	.9726306

Nota: mu\_1 y mu\_2 representan la media para el caso de los tratados y no tratados respectivamente. Se presenta también el P-Value de la prueba de diferencia de medias, donde  $p < 0.05$  implica medias estadísticamente equivalentes.

Cuadro 10: Controles con tratamiento Invitaciones

	Coeficiente	P-Value
undfpm	-.0248021	.4286703
zkadesedyears	-.0065928	.2988745
zkadesage	-.0039699	.034729
zpop	.0074787	.2378095
zpercentpoorpra	-.0626138	.3760045
zdistancekec	-.0013126	.723312
podeszhill	-.0615267	.0663954

Nota: Se presentan los coeficientes de la estimación lineal con errores estándar robustos. También se presenta el P-Value de los coeficientes.

## 1.2. Duración de mandatos y producción legislativa

En esta subsección se analiza el trabajo de [Dal Bó y Rossi \(2011\)](#).

- [Dal Bó y Rossi \(2011\)](#) analizan un experimento natural en el Congreso Argentino para evaluar si los límites en los mandatos afectan el esfuerzo y outcomes de los legisladores. Lean el trabajo y contesten las siguientes preguntas:

A. ¿Por qué no podemos simplemente analizar correlaciones entre países con mandatos más largos y más cortos?

Si comparáramos el esfuerzo legislativo entre países con diferentes duraciones de mandatos, podría ser que la diferencia encontrada no se deba al largo del mandato sino a características propias de cada uno de los países.

- B. b) ¿Por qué no comparamos, por ejemplo, a los senadores de Buenos Aires antes y después de que se extiendan los mandatos?

Si comparáramos el esfuerzo legislativo entre dos momentos distintos del tiempo, por ejemplo el cambio en la duración de los mandatos de Buenos Aires, podría ser que la diferencia encontrada no se deba al largo del mandato sino a características propias del momento del tiempo.

- C. Expliquen brevemente el experimento natural y la estrategia de identificación.

Teniendo en cuenta esto, se aprovecharon dos situaciones particulares ocurridas en Argentina. La primera situación con la renovación total de las cámaras posterior a

la dictadura militar 1973-1983. La segunda, la reforma consitucional de 1994. En estas dos situaciones los legisladores fueron elegidos por voto popular y luego se asignó aleatoreamente el largo del mandato de cada uno de ellos. De esta manera la única diferencia entre unos y otros legisladores es el tratamiento, es decir unos con un mandatos de dos años, otros de cuatro y otros de seis dependiendo el caso.

- D. ¿Qué checks realizan los autores para mostrar que la asignación aleatoria se realizó correctamente? Expliquen los checks que son distintos a los que aparecen en ([Olken, 2007](#)).

Los autores muestran las medias de distintas variables para los tratados y los no tratados, y los resultados del test de diferencia de medias.

- E. ¿Por qué los autores no usan efectos fijos por legislador en la Tabla 5?

En lugar de utilizar efectos fijos por legislador, los autores utilizan errores estándar clusterizados al nivel del legislador y reportan la información colapsada por legislador.

- F. ¿Cómo se justifica la elección de los errores estándar en esta especificación?

La regla general de la clusterización propone hacerlo a nivel del tratamiento, en este caso sería clusterizar a nivel del legislador. Sin embargo, es posible pensar que a nivel provincial los legisladores trabajen en conjunto, buscando realizar esfuerzos que redunden en posibilidad de reelegir no solo para ellos sino también para otros candidatos de su partido y provincia.

- G. En la sección de checks de robustez 4.1 los autores clusterean los errores estándar al nivel de la provincia ¿Le parece adecuada esta especificación? ¿Cuáles son los problemas usar clustered standard errors cuando no son necesarios?

La idea de clusterizar al nivel de la provincia es que los legisladores podrían compartir los esfuerzos por ser reelegidos con sus compañeros de partido y provincia.

Usar errores estándar agrupados (clustered standard errors) cuando no son necesarios puede llevar a conclusiones erróneas en el análisis y tener un impacto negativo en la interpretación de los resultados. Cuando se utilizan errores estándar agrupados en ausencia de agrupación real en los datos, los errores estándar pueden ser más grandes de lo necesario. Esto puede llevar a la sobreestimación de la significancia estadística de las variables explicativas y a la subestimación de los intervalos de confianza.

Un problema asociado sería si al hacer un agrupamiento tan general quedaran identificados pocos clusters. En esos casos es preferible usar errores estándar comunes ([Angrist y Pischke, 2009](#), p. 234).

- H. ¿Hay problemas de spillovers? ¿Y de compliance? Justificar.

Podría entenderse como spillovers (externalidades) la situación de un legislador no asignado al tratamiento pero con un compañero de partido que sí recibió el tratamiento y entonces los esfuerzos que realice podrían ser aprovechados por el que sí recibió el tratamiento. La existencia de spillovers representa una violación al supuesto SUTVA de Rubin ([Cunningham, 2021](#), p. 153).

En principio no debería haber casos de compliance significativos, pero podría pensarse el caso de un legislador que a pesar de haber sido asignado a un período legislativo largo sabía desde el principio que no iba a completar su mandato por alguna situación. En ese caso sería un asignado al tratamiento pero que no lo recibió completamente.

- I. En la nota de la Tabla 10 los autores afirman que: “Standard errors clustered at the legislator level are in parentheses (for collapsed data, these are equivalent to robust standard errors).” Expliquen esta afirmación.

Los intervalos de confianza agrupados a nivel del legislador tienen en cuenta la correlación entre las observaciones, que es lo mismo que hace un intervalo de confianza robusto.

En otras palabras, los intervalos de confianza agrupados a nivel del legislador son una forma de calcular intervalos de confianza robustos para los datos colapsados.

## 2. Diferencias en Diferencias

### 2.1. Consecuencias políticas de los recursos naturales

- [Sances y You \(2022\)](#) analizan las consecuencias políticas de la industria del fracking en Estados Unidos. Aquí vamos a concentrarnos en los efectos sobre las contribuciones de campaña. La unidad de análisis es el código postal (ZIPC) / ciclo electoral (year). Las variables dependientes son las contribuciones totales (ltotamount), contribuciones al partido Republicano (ltotamountr), y contribuciones al partido Demócrata (ltotamountd). Estas variables están en logaritmos:  $\log(1+x)$ . La primera variable independiente que vamos a usar es el logaritmo de los pozos de fracking (lnwells). Como primer paso, vamos a realizar una estimación por paneles con efectos fijos:
- A.1 Propongan una ecuación a estimar que tenga efectos fijos por código postal y efectos temporales por año. Presenten los resultados en una tabla e interpreten el coeficiente de interés. Expliquen su elección respecto a los errores estándar.

La ecuación que incluye efectos fijos temporales por año y por código postal tomaría la siguiente forma:

$$y_{it} = \alpha_i + \delta_t + \beta X_{xi} + \epsilon_{it} \quad (1)$$

Donde  $\delta_t$  controla por todos los factores temporales que afectan a todas las unidades por igual y  $\alpha_i$  captura todos los determinantes de  $y_{it}$  que no varían en el tiempo.  $y_{it}$  en este caso es el logaritmo de las contribuciones totales (ltotamount) y se utiliza como única variable independiente  $X_{xi}$  el logaritmo de los pozos de fracking (lnwells).

- Realicen unas estimaciones como las del punto anterior, pero restringiendo el análisis a los estados que concentran la mayor parte de la industria del fracking (highfracking = 1). Presenten estos resultados en la misma tabla que el punto anterior.



Cuadro 11: Regresiones Efectos Fijos

	lnwells	lnwells*
Coefficiente	.056085	.0848796
P-Value	.00011	9.96e-08

Nota: Se presentan los coeficientes de la estimación lineal con efectos fijos. Se presenta también el P-Value de los mismos. El primer par se refiere a la estimación sobre toda la base y el par indicado con asterísco a la estimación para los estados que concentran la mayor parte de la industria del fracking (highfracking = 1).

La salida de estas estimaciones para el total de la base y para los estados que concentran la mayor parte de la industria del fracking (highfracking = 1) puede verse en el Cuadro 2.1.

- Para analizar si la industria del fracking tuvo un efecto sobre las contribuciones de campaña, los autores proponen una estrategia de identificación que se basa en usar el 2005 como año en el que la industria del fracking se volvió económicamente viable a gran escala (año del tratamiento). La variable frack toma el valor de 1 para todos los códigos postales que, en algún momento de la muestra, tuvieron al menos un pozo de fracking. En esta sección, vamos restringir el análisis a los siete estados que concentran la mayor parte de la explotación de fracking.
- B. Realicen una estimación de diferencias en diferencias usando el año 2005 como el momento en que se asigna el tratamiento. Realicen esta estimación para las contribuciones a Republicanos y Demócratas por separado. Presenten los resultados en una nueva tabla e interpreten. Muestren evidencia sobre tendencias paralelas. Interpreten estos resultados.

El Cuadro 12 presenta una estimación de diferencias en diferencias (DID) usando el año 2005 como el momento en que se asigna el tratamiento. Se muestra el resultado para las contribuciones totales en la especificación (1), para las contribuciones a los Republicanos en la especificación (2) y a los Demócratas en la especificación (3).

La Figura 1 presenta la modificación en las tendencias paralelas de las contribuciones al partido Republicano a partir del año 2005. Puede observarse como hasta el año 2005 las contribuciones al partido Republicano eran muy parecidas para los lugares con pozos de fracking y sin pozos de fracking. En cambio, a partir de ese año los lugares con pozos de fracking tienen mayor contribución al partido republicano. Esta modificación en la tendencia no se observa para el caso del partido Demócrata, tal como muestra la Figura 2.

- Bertrand et al (2004) muestran que las estimaciones por paneles pueden sufrir de serios problemas de autocorrelación. Si bien este problema puede ser más complejo con pocos clusters, vamos a realizar un check para asegurarnos que nuestra inferencia no está siendo afectada.

C Realicen el mismo análisis que en el punto anterior como un DD 2x2, es decir, tomando solo los años 2000 y 2014. Presenten sus resultados en la tabla anterior e interpreten.

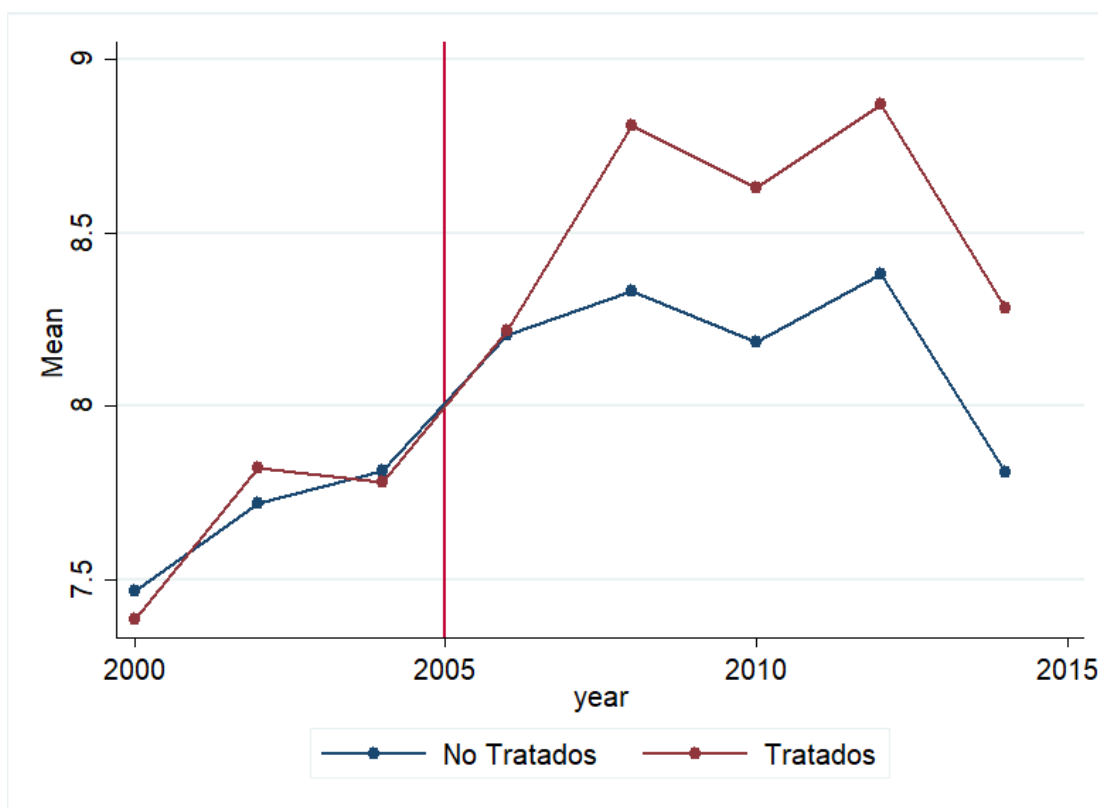
Cuadro 12: Diferencias en Diferencias

	(1) DID total	(2) DID Republicanos	(3) DID Demócratas
did	0.0573** (0.0287)	0.281*** (0.0466)	-0.176*** (0.0526)
o.frack	-	-	-
treatment	-0.342*** (0.0231)	0.412*** (0.0375)	0.451*** (0.0423)
lnwells	0.0751*** (0.0126)	0.0885*** (0.0204)	0.111*** (0.0231)
2002.year	-0.0615*** (0.0219)	0.325*** (0.0356)	0.563*** (0.0402)
2004.year	0.353*** (0.0217)	0.470*** (0.0352)	0.672*** (0.0398)
2006.year	0.376*** (0.0225)	0.320*** (0.0366)	0.221*** (0.0413)
2008.year	0.961*** (0.0219)	0.811*** (0.0355)	0.797*** (0.0401)
2010.year	0.538*** (0.0217)	0.602*** (0.0353)	0.352*** (0.0398)
2012.year	1.125*** (0.0211)	0.992*** (0.0343)	0.673*** (0.0387)
2014o.year	-	-	-
Constant	8.452*** (0.0155)	7.285*** (0.0253)	6.714*** (0.0285)
Observations	37,842	37,842	37,842
R-squared	0.122	0.091	0.040
Number of zipcode_n	5,719	5,719	5,719

Standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Figura 1: Tendencias Paralelas



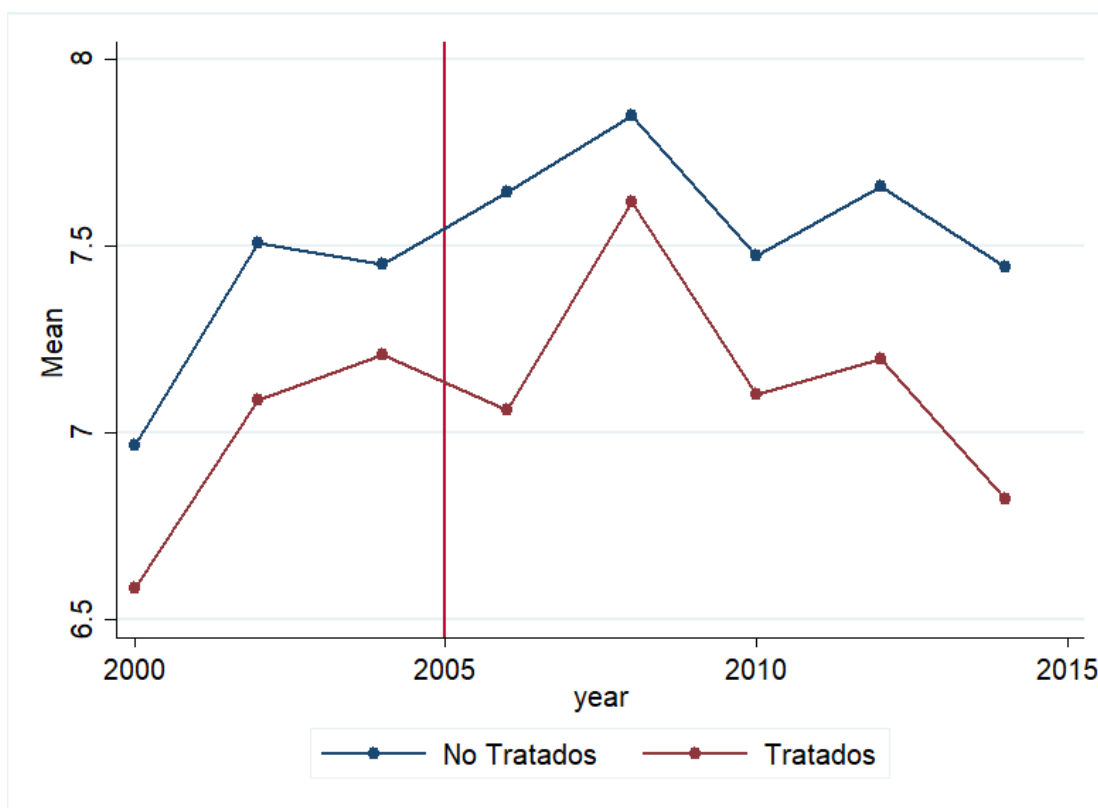
Nota: El gráfico presenta la modificación en las tendencias paralelas de las contribuciones al partido Republicano a partir del año 2005.

- Al trabajar a un nivel de agregación pequeño como son los códigos postales, existen potenciales problemas de spillovers. Para evaluar esta posibilidad, vamos a realizar un análisis adicional agregando estos datos a nivel estatal, y concentrándonos en las contribuciones al partido Republicano. D.1 Supongamos que en nuestra especificación al nivel del código postal hubiese spillovers. Expliquen qué es lo que estaría ocurriendo si ese fuese el caso ¿Cuál sería la dirección de ese sesgo? ¿Cómo afectaría la interpretación de nuestros resultados?

Si existieran spillovers los resultados o efectos observados en un código postal afectarían a los códigos postales circundantes. Esto significa que las variables o factores que están siendo estudiados en un código postal pueden tener un impacto en los códigos postales vecinos, lo que podría distorsionar la interpretación de los resultados. Si se asume que hay spillovers en la especificación a nivel de código postal, esto implicaría que las observaciones en un código postal específico no son independientes de las observaciones en los códigos postales cercanos.

Es decir, la presencia de un pozo de fracking en un determinado código postal podría no solo afectar a la población (y sus intenciones de votar) y las contribuciones en ese determinado código postal, sino también en los códigos postales vecinos. Siguiendo los resultados hasta aquí observados, sería esperable que, así como en el código postal en cuestión la participación en la elecciones baja y las contribuciones al partido Republicano aumenta, lo mismo suceda en los códigos postales vecinos. Se podría así estar subestimando la relevancia de la presencia de pozos de fracking ya que no solo afectarían al código postal en cuestión sino también a los circundantes.

Figura 2: Tendencias Paralelas



Nota: El gráfico presenta las tendencias paralelas sin modificación a partir del año 2005 de las contribuciones al partido Demócrata.

- Para trabajar a nivel estatal, vamos a tener que realizar algunos cambios a la muestra que tenemos. Lo que vamos a hacer es agregar los datos a nivel de STATE. Tengan en cuenta que: a) las contribuciones a Republicanos están en logs, así que tenemos que pasarla a niveles antes de agregarla, y luego volver a tomar el logaritmo;2 b) la variable *frack* va a tomar más valores luego de agregarla, recodifiquen esta variable para poder usarla en una especificación de diferencias en diferencias clásico. D.2 Realicen una estimación de diferencias en diferencias para medir el impacto del fracking sobre las contribuciones al partido Republicano a nivel estatal. Presenten los resultados en una nueva tabla.

Los resultados de esta estimación se muestran en el Cuadro 14. Se observe que se pierde la significatividad de los resultados alcanzados. Esto podría deberse a que se están analizando muy pocas unidades por año.

- D.3 Con pocos clusters, nuestra inferencia puede tener errores. Realicen la misma estimación que en el punto anterior, pero usen block bootstrap para calcular los errores estándar. Presenten estos resultados en la misma tabla ¿Qué concluimos?

## 2.2. Los efectos de la AUH

En esta sección se analiza el trabajo de [Edo, Marchionni, y Garganta \(2017\)](#).

Cuadro 13: Diferencias en Diferencias 2X2

	(1) DID total	(2) DID Republicanos	(3) DID Demócratas
did	0.0866 (0.0900)	0.550*** (0.143)	-0.600*** (0.157)
o.frack	-	-	-
treatment	-0.312*** (0.0317)	0.392*** (0.0503)	0.504*** (0.0555)
lnwells	0.0383 (0.0330)	-0.0419 (0.0524)	0.250*** (0.0577)
2014o.year	-	-	-
Constant	8.549*** (0.0190)	7.435*** (0.0301)	6.845*** (0.0332)
Observations	9,027	9,027	9,027
R-squared	0.026	0.038	0.030
Number of zipcode_n	5,131	5,131	5,131

Standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

- Muchos trabajos analizan distintos efectos de la Asignación Universal por Hijo (AUH) usando una estrategia similar de diferencias en diferencias. Ahora vamos a analizar un estudio sobre los efectos en la escolaridad. Lean el paper de [Edo et al. \(2017\)](#) y contesten las siguientes preguntas:

A. ¿Cuál es el problema de endogeneidad? ¿Cómo se definen tratados y controles?

Existe un problema de endogeneidad porque se intenta identificar el efecto de la AUH independientemente de los efectos de la Ley Nacional de Educación que hizo obligatoria la escuela secundaria en 2006.

- B. ¿Por qué los autores dicen que su estimación es un intention to treat (ITT)? ¿Cómo afecta esto la generalización de los resultados?

El estudio se trata de un ITT porque la Encuesta Permanente de Hogares (EPH) no incluye preguntas que permitan identificar a los beneficiarios de la AUH. Por este motivo los grupos de "tratamiento y control" se definen en función de la elegibilidad para la AUH. Podría haber personas que cumplan los criterios de elegibilidad para la AUH pero que sin embargo no la reciban.

- C. A diferencia de los casos que vimos de experimentos aleatorios, en la Tabla 3 vemos que tratados y controles son distintos en varias características observables ¿Es un problema dado su diseño de diferencias en diferencias? Fundamentar.

Era de esperar que tratados y controles difieran en varias características dado el carácter no aleatorio del experimento. Las diferencias no presentan un problema

Cuadro 14: Diferencias en Diferencias por Estado

	(1) DID total	(2) DID Republicanos	(3) DID Demócratas
did	0.227 (0.399)	0.0793 (0.335)	0.321 (0.370)
o.frack	-	-	-
o.treatment	-	-	-
lnwells	0.141 (0.0894)	0.175** (0.0750)	-0.0527 (0.0828)
2008.year	0.718*** (0.0727)	0.312*** (0.0610)	0.481*** (0.0673)
2010.year	0.187** (0.0730)	0.142** (0.0613)	0.0392 (0.0677)
2012.year	0.815*** (0.0754)	0.438*** (0.0633)	0.290*** (0.0699)
2014.year	-0.0238 (0.0782)	-0.168** (0.0656)	-0.205*** (0.0724)
Constant	15.57*** (0.595)	13.98*** (0.499)	15.07*** (0.551)
Observations	240	240	240
R-squared	0.572	0.410	0.412
Number of state_id	48	48	48

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Cuadro 15: Diferencias en Diferencias por Estado

	(1) DID total	(2) DID Republicanos	(3) DID Demócratas
did	0.227*** (0.0601)	0.0793 (0.0525)	0.321*** (0.0801)
o.frack	-	-	-
o.treatment	-	-	-
lnwells	0.141 (0.0942)	0.175*** (0.0551)	-0.0527 (0.0586)
2008.year	0.718*** (0.0755)	0.312*** (0.0722)	0.481*** (0.0967)
2010.year	0.187*** (0.0606)	0.142*** (0.0419)	0.0392 (0.0423)
2012.year	0.815*** (0.0599)	0.438*** (0.0668)	0.290*** (0.0603)
2014.year	-0.0238 (0.0788)	-0.168*** (0.0581)	-0.205*** (0.0609)
Constant	15.57*** (0.641)	13.98*** (0.766)	15.07*** (0.694)
Observations	240	240	240
R-squared	0.572	0.410	0.412
Number of state_id	48	48	48

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

porque las mismas se mantienen en el período pre y post implementación de la AUH. De hecho, la asistencia escolar ya difería antes de la implementación de la AUH entre ambos grupos. Se presume que estas diferencias en la asistencia escolar están explicadas por las diferencias en las características.

- D. Además de tendencias paralelas, los autores mencionan otra parte del supuesto de identificación del método de diferencias en diferencias. Explicar cuál es y qué evidencia y/o argumento presentan al respecto.

Los autores explican que aunque los niveles de asistencia son distintos antes de la implementación del programa, los comportamientos temporales son similares. Comprueban este efecto con un test de tendencia común para el período previo a la implementación del programa. Luego de la implementación del programa el grupo de tratamiento achica su diferencia con el grupo de control.

- E. ¿Por qué la Tabla 5 es importante? ¿En qué se parece este tipo de ejercicio a los resultados de la Columna 4 de la Tabla 8 y de la Columna 4 de la Tabla 11?

Los ejercicios placebo son importantes para validar la estrategia de identificación. Se evidencia gracias a esta tabla que solo a partir de 2009 la asistencia del grupo de tratamiento comienza a acercarse a la del grupo de control. Esto es similar a lo que ocurre con el grupo de los mayores de 18 años. Como no son elegibles para la AUH no deberían presentar efectos post 2009. Si los presentaran significaría que podría haber algún otro factor que afecta su asistencia a partir de 2009. Algo similar puede decirse de los hijos de padres con educación superior. La idea de lo que se intenta probar es que una asignación monetaria a personas de bajos recursos los ayuda a sostener la posibilidad de mantener al hijo o la hija en la escuela. Se supone que los padres con educación superior no tendrían este tipo de problemas y por tanto deberían verse poco afectados por ser incluidos en el grupo de tratamiento.

- F. ¿Tenemos evidencia sobre mecanismos? ¿Cuáles se les ocurre que podrían ser y cómo podríamos testearlos?

El mecanismo por el cuál la asignación favorece la permanencia en la escuela no queda claro en el trabajo, y se propone como línea de investigación posterior. Por un lado, podría tratarse de que la ayuda monetaria posibilita que los hijos permanezcan en el colegio en lugar de tener que salir a buscar trabajo. Como se trata de trabajo ilegal y por lo tanto en negro, resulta muy difícil obtener datos y probar la hipótesis. Una estrategia posible sería utilizar encuestas al nivel de las personas que abandonan la escuela o al nivel de los colegios para preguntarles por estas personas. El otro mecanismo posible sería que los alumnos permanezcan en la escuela solo por el hecho de querer recibir el dinero a cambio. Por supuesto, ambos mecanismos podrían estar funcionando al unísono.

## Referencias

- Angrist, J., y Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*.
- Cunningham, S. (2021). *Causal inference: The mixtape*. London: Yale University Press.



- Dal Bó, E., y Rossi, M. A. (2011). Term Length and the Effort of Politicians. *Review of Economic Studies*, 78(4), 1237-1263. doi: 10.1093/rdr010
- Edo, M., Marchionni, M., y Garganta, S. (2017, Jul.). Compulsory education laws or incentives from cct programs? explaining the rise in secondary school attendance rate in argentina. *Education Policy Analysis Archives*, 25, 76. doi: 10.14507/epaa.25.2596
- Hansen, B. (2014). *Econometrics*. University of Wisconsin, Department of Economics.
- Olken, B. (2007, 02). Monitoring corruption: Evidence from a field experiment in indonesia. *Journal of Political Economy*, 115, 200-249. doi: 10.1086/517935
- Sances, M. W., y You, H. Y. (2022). Voters and donors: The unequal political consequences of fracking. *The Journal of Politics*, 84, 1667 - 1682. doi: 10.1086/717084