
Análisis de datos de panel en ciencia política: ventajas y aplicaciones en estudios electorales

*Panel Data Analysis in Political Science:
Advantages and applications in electoral studies*

ADRIÁN PIGNATARO

Universidad de Costa Rica

Cómo citar/Citation

Pignataro, A. (2018). Análisis de datos de panel en ciencia política: ventajas y aplicaciones en estudios electorales. *Revista Española de Ciencia Política*, 46, 259-283. Doi: <https://doi.org/10.21308/recp.46.11>

Resumen

La ciencia política ha recurrido a datos de panel o longitudinales de forma incremental en las últimas décadas. En comparación con datos transversales, las inferencias que se pueden obtener a través de datos de panel son considerablemente más reveladoras; sin embargo, su análisis no debe limitarse al modelo clásico lineal de regresión de mínimos cuadrados ordinarios. Con datos electorales reales y simulaciones, esta nota de investigación demuestra la mayor robustez de los modelos de efectos fijos y efectos aleatorios, así como los problemas con el estimador Arellano-Bond en el caso de paneles cortos en el tiempo, comunes en estudios electorales.

Palabras clave: panel, datos longitudinales, elecciones, simulación Monte Carlo.

Abstract

Political science has relied upon panel or longitudinal data incrementally in the last decades. In comparison with cross-section data, the inferences attainable through panel data are considerably more enlightening; however, its analysis should not be limited to the classical linear model of ordinary least-squares regression. Based on real electoral data and simulations, this research note demonstrates the greater robustness of the fixed and random effects models and the problems associated to the Arellano-Bond estimator, in the case of short panels, which are common in electoral studies.

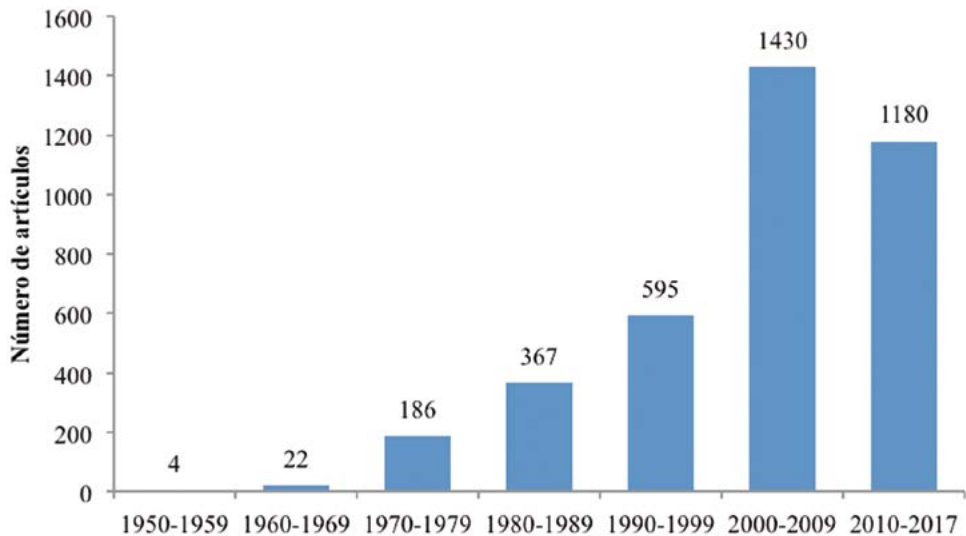
Keywords: panel, longitudinal data, elections, Monte Carlo simulation.

INTRODUCCIÓN

Los datos de panel, también llamados longitudinales, son definidos como cualquier conjunto de datos con observaciones repetidas a lo largo del tiempo (Arellano, 2003: 1). En otras palabras, a diferencia de los datos de corte transversal, los cuales son recopilados en un único momento, los longitudinales contienen unidades estadísticas (N) que son medidas más de una vez en el tiempo (T)¹. Estos datos han sido comunes en la investigación científica y la ciencia política no ha sido excepción. Así, una búsqueda simple de artículos académicos en JSTOR con las frases «panel data» y «longitudinal data» refleja que su uso en la disciplina política ha crecido de forma sostenida (figura 1).

FIGURA 1.

NÚMERO DE ARTÍCULOS CIENTÍFICOS CON DATOS DE PANEL O DATOS LONGITUDINALES EN CIENCIA POLÍTICA



Fuente: elaboración propia con base en búsquedas de www.jstor.org.

Diversos estudios electorales han recurrido a encuestas longitudinales, formadas por muestras repetidas de electores, por ejemplo, en el *American National Election*

1. Este artículo no distingue entre panel y datos longitudinales, a diferencia de Gerring (2012), para quien los primeros corresponden a datos con más de una medición en el tiempo mientras los segundos a datos sin variación transversal entre sí, pero que muestran variación temporal. Esta demarcación parece apuntar más bien hacia una característica de los datos y no a un diseño en la recolección. Por otro lado, la diferencia entre panel y longitudinal está relacionada con el uso desde distintas disciplinas (*v. g.*, en econometría se ha preferido el término «panel»).

Study (ANES) de Estados Unidos, en el *British Election Study*, en el Centro de Investigaciones Sociológicas (CIS) de España, en el *German Longitudinal Election Study* y en el *Italian National Election Studies* (ITANES). Sin embargo, no todos los datos de panel se basan en encuestas por muestreo a personas, denominados «micropaneles» por la literatura en inglés (Baltagi, 2005). Las unidades de análisis también pueden ser países (como en Persson y Tabellini, 2003), regiones subnacionales (Sátyro, 2013), municipalidades (De Benedetto y De Paola, 2016), distritos electorales (Seijas, 2014) y gabinetes presidenciales (Escobar-Lemmon y Taylor-Robinson, 2005); en general, reconocidos como «macropaneles». Además, existen diseños sofisticados que combinan encuestas, anidando individuos en distintos países y en el tiempo, y que modelan más de un nivel de análisis—micro y macro—de forma longitudinal (Fairbrother, 2014).

Dada la difusión de los datos de panel, los metodólogos políticos han discutido las herramientas para su análisis, advirtiendo cómo la mayoría de veces las técnicas aplicadas se escogen sin justificación o reflexión apropiada (Wilson y Butler, 2007). Incluso en ocasiones, se publican estudios con valiosas aportaciones sustantivas que, sin embargo, no explicitan el método de estimación para el panel que se aplicó (por ejemplo, Blais y Dobrzyńska, 1998).

Respecto a cómo analizar estos datos, existen reconocidos aportaciones para paneles donde el número de mediciones en el tiempo es relativamente largo respecto a las observaciones espaciales ($N < T$), que algunos denominan *time series cross-section data* (Beck, 2008; Beck y Katz, 1995, 2011; Fortin-Rittberger, 2015). No obstante, en estudios electorales es común encontrar la situación inversa: muchos países con un número reducido de elecciones ($N > T$), distinción que muchas veces se obvia. Esto ocurre, principalmente, porque las elecciones se celebran entre amplios intervalos de años, a diferencia de las mediciones económicas—indicadores como el desempleo, el crecimiento y la inflación—, que se registran de forma anual o incluso mensualmente. Dicha situación, denominada panel corto, además se acentúa en el caso de democracias jóvenes, donde se han celebrado pocas elecciones libres y competitivas.

En consecuencia, este artículo brinda mayor atención a los principales modelos estadísticos para el análisis de datos de panel corto, aquellos con mayor número de casos respecto a periodos de tiempo, comunes—pero no exclusivos—de los estudios electorales. Para ello, en la siguiente sección se resaltan las deficiencias derivadas de trabajar con datos transversales y cómo los datos longitudinales podrían superarlas. Seguidamente se exponen los principales modelos teóricos: efectos fijos, efectos aleatorios y estimador Arellano-Bond. Luego se ejemplifica el análisis con un panel corto de elecciones presidenciales en América Latina y se profundiza en algunos problemas presentes en los datos empíricos a través de simulaciones. Las conclusiones sostienen cómo la selección del modelo es relevante, en el caso de los datos longitudinales, pues afecta a las respuestas que se obtienen sobre temas sustantivos de ciencia política, coincidiendo con la literatura previa sobre la importancia de las decisiones técnicas en la estimación de modelos de panel (Beck, 2007; Beck y Katz, 1995; Plümper *et al.*, 2005; Stimson, 1985; Wilson y Butler, 2007).

PROBLEMAS CON LOS DATOS TRANSVERSALES Y FORTALEZAS EN LOS PANELES

Los datos de tipo transversal abundan en distintas áreas del conocimiento. Pese a ello, presentan ciertas vulnerabilidades metodológicas que los datos de panel ayudan a resolver.

La primera –y quizás la más relevante– deficiencia de los datos transversales es su incapacidad para contemplar la dinámica del fenómeno. Por su misma definición, estos datos son un corte en el tiempo, un momento particular en un contexto que muy probablemente sea cambiante. En el análisis macro –por ejemplo– de los niveles de identificación partidaria en distintos países, el notable declive de simpatizantes en la década de los noventa implica que un corte transversal en dichos años llevaría a conclusiones diversas de las que se encontraría en los ochenta o antes (Dalton, 2000). Asimismo, promediar los niveles de simpatía en un largo periodo de años para obtener una medición transversal no hace sino ocultar la variabilidad en el tiempo. En una perspectiva micro, como en el estudio de la intención de voto, un conjunto de encuestas transversales medidas a lo largo de una campaña electoral puede ilustrar en cada punto del tiempo un porcentaje de personas decididas a votar que, aunque podría aparentar ser estable, esconde movimientos de las personas que pasan de decisión a indecisión –y viceversa– entre las olas de medición².

Un estudio de panel resolvería ambas situaciones ya que este diseño captura la variación en el tiempo. En el análisis macro, sería capaz de observar no solo la variabilidad entre países, sino en cada uno de ellos en el tiempo, revelando –en el ejemplo dado– cada tendencia de desidentificación partidaria por país. En el caso de la intención de voto, una encuesta de panel en la que se realicen entrevistas a las mismas personas en distintas olas permitiría estimar el porcentaje de cuántos dejan de ser indecisos y escogen el voto –y al revés–, midiendo además con mayor precisión el momento de decisión en comparación con las preguntas basadas en el recuerdo (*recall-based measures*) (Steinbrecher y Shoen, 2013). Además, los estudios de panel permiten estudiar los efectos de las campañas electorales que, por su naturaleza, ocurren en una secuencia temporal (Dilliplane, 2014; Moreno, 2010) a diferencia de predisposiciones políticas que suelen ser estables como la identificación ideológica y partidaria.

En segundo lugar, pero vinculado al primer punto, los datos transversales son susceptibles a problemas de causalidad inversa, es decir, cuando la variable dependiente podría estar afectando una o más variables independientes, siendo esta una fuente de endogeneidad en un modelo estadístico (Wooldridge, 2010: 54-55). Este fenómeno, común en política comparada (Franzese, 2007), se ejemplifica con el caso del desarrollo y la democracia: mientras algunos argumentan que el mayor desarrollo económico privilegia la democratización (Lipset, 1959), otros sostienen que la democracia facilita el crecimiento económico (Przeworski *et al.*, 2000). En la investigación micro, por su

2. Este es un ejemplo de la llamada «falacia ecológica» (Robinson, 1950).

parte, se sospecha de causalidad inversa al analizar percepción de candidatos y simpatía partidaria (Garzia, 2013), percepciones económicas e intención de voto (Lewis-Beck *et al.*, 2008) y partidismo y preferencias electorales (Lupu, 2015), entre otras posibles relaciones endógenas. Con datos transversales resultaría difícil respetar la precedencia temporal de las causas a los efectos ya que las variables se miden una única vez en el tiempo (Brady, 2008; Hsiao, 2003: 312).

Por el contrario, los datos de panel, por la forma en que son recogidos, permiten distinguir con más precisión entre el antes (causa) y el después (efecto), logrando una mejor comprensión causal del fenómeno. Así, en una investigación electoral es posible modelar la variable dependiente tomando en cuenta la medición realizada en una ola posterior, mientras que las independientes se obtienen de las primeras recogidas. Vale aclarar que, aunque los datos de panel son metodológicamente más robustos que los transversales para determinar causalidad, resultan inferiores ante los diseños experimentales que incorporan la asignación aleatoria de tratamientos (Frees, 2004; Przeworski, 2007). Sin embargo, para contextos observacionales, constituyen la mejor opción.

En tercera instancia, con datos transversales pueden encontrarse muestras reducidas que impiden un análisis estadístico; por ejemplo, en estudios de área ($N = 4$ en Europa del sur; $N = 6$ en América Central). Incluso si la muestra no es escasa, los parámetros por estimar pueden ser muchos en relación con las observaciones, lo cual —como se sabe— es poco conveniente (Western y Jackman, 1994). Ante ello, con datos de panel se amplía el número total de casos, pues se multiplica el tamaño de muestra N por el número de mediciones en el tiempo, incrementándose los grados de libertad. (King *et al.*, 1994: 219). Así, cinco países medidos diez años generan un número total de cincuenta observaciones, aceptable para el análisis cuantitativo. Con ello, además, se reducen problemas de «multicolinealidad», es decir, alta correlación entre variables explicativas, perjudicial en las estimaciones (Hsiao, 2003: 311).

Por último, cuando terceras variables (*confounders*) que afectan la relación causal entre la variable independiente y la dependiente no se pueden medir directamente, se genera un sesgo de variable omitida (King *et al.*, 1994: 169). Por ejemplo, en el estudio de la relación entre instituciones políticas y desempeño económico (Przeworski *et al.*, 2000) es difícil —si no imposible— obtener indicadores sobre variables culturales e históricas que se supone podrían incidir en la variable dependiente, *desempeño económico*. Con datos transversales no resulta sencillo incluir estos efectos inobservados. Aunque la opción más simple para el caso de un modelo de regresión consiste en capturar efectos de las unidades de análisis (como países) a través de variables *dummy*, esto tiene la desventaja de aumentar el número de parámetros por estimar, lo cual es especialmente problemático si se combina con el problema de los pocos casos.

Al respecto, los datos de panel son ventajosos ya que existen modelos que toman en cuenta la heterogeneidad o las características propias de las unidades de análisis, incluyendo variables inobservables, bajo la condición de que estas cambien poco o nada en el tiempo. Por ejemplo, en el caso de patrones culturales e históricos de un país, o bien de instituciones informales en los parlamentos, para los cuales es difícil obtener indicadores, un modelo de efectos fijos controlaría esta heterogeneidad no

observada asumiendo que estos aspectos son invariables en el tiempo. En la investigación micro del comportamiento político, se contemplarían actitudes y orientaciones latentes que permanecen (relativamente) constantes a lo largo de la vida, sin incurrir al sesgo de variable omitida por ausencia de indicadores.

Los cuatro beneficios potenciales de los datos de panel se resumen en la tabla 1.

TABLA 1.
DATOS TRANSVERSALES VS. DE PANEL

Problemas presentes con datos transversales	Potencial mejora con datos de panel
Ocultan variaciones temporales	Revelan la variabilidad temporal
Causalidad recíproca (endogeneidad)	Especificación del antes y el después en el modelo
Tamaño de muestra pequeño	Incremento de los grados de libertad a través de mediciones en el tiempo
Sesgo de variable omitida	Control de variables no observadas

Fuente: elaboración propia.

Aunque los datos de panel presentan palpables ventajas metodológicas, hay algunas dificultades en su aplicación (Bartels, 1999). La más frecuente corresponde a la pérdida de observaciones en el tiempo (*attrition*), la cual puede darse por el abandono voluntario (personas encuestadas que pierden interés de participar en subsiguientes entrevistas), por desaparición (partidos políticos o países que dejan de existir) o por «nacimiento» tardío (países que surgen en el tiempo). Por otro lado, en el caso de las encuestas, se produce un condicionamiento psicológico, debido a las múltiples olas de entrevistas (Lynn, 2009). Adicionalmente, la ganancia metodológica de los datos longitudinales requiere una mayor sofisticación en los modelos de análisis y su estimación. En otras palabras, hay que ir más allá del tradicional modelo lineal de mínimos cuadrados ordinarios, como se verá en la siguiente sección.

PRINCIPALES MODELOS DE ANÁLISIS

Un modelo de regresión lineal clásico para datos transversales de N observaciones sigue la siguiente forma:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, N. \quad (\text{Ecuación 1})$$

En esta ecuación y_i es la variable dependiente, x_{ki} son las variables independientes, β_k los coeficientes de regresión, α es un intercepto común y u_i es un término de error. Este modelo puede generalizarse para datos de panel como:

$$y_{it} = \alpha + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \dots + \beta_k x_{k,it} + u_{it}, \quad \begin{array}{l} i = 1, \dots, N; \\ t = 1, \dots, T_i \end{array} \quad (\text{Ecuación 2})$$

En pocas palabras, se expande para incorporar variación no solo entre casos de N , sino también en el tiempo T . Para la discusión que se plantea, se asumen propiedades asintóticas para N (v. g., que tiende a infinito) mientras que T es fijo y, por ende, $N > T$, el llamado panel corto.

Los parámetros de la ecuación 1 se pueden estimar haciendo uso del procedimiento de mínimos cuadrados ordinarios, el cual permite obtener los mejores estimadores lineales no sesgados bajo una serie de supuestos, según el teorema Gauss-Markov. Sin embargo, este método es inadecuado para la ecuación 2 ya que en un panel comúnmente existe heterogeneidad entre los casos (ver la sección anterior). Al agrupar observaciones sin contemplar esta heterogeneidad, que se evidencia por contener un intercepto único (α), los coeficientes (β_k) estimados resultan sesgados e inconsistentes (Gujarati y Porter, 2010: 594). Para estimar adecuadamente modelos con datos de panel se han desarrollado dos modelos básicos: el modelo de efectos fijos y el modelo de efectos aleatorios³.

El modelo de efectos fijos (*fixed effects model*), también conocido como modelo de mínimos cuadrados con variable indicadora (*least squares dummy variable*), se escribe de la siguiente forma:

$$y_{it} = \alpha_i + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \dots + \beta_k x_{k,it} + u_{it}. \quad (\text{Ecuación 3})$$

Puede notarse que la única diferencia entre la ecuación 3 y la ecuación 2 corresponde al término del intercepto, pues en el caso del modelo de efectos fijos este varía entre unidades (α_i). Esta especificación es la que permite contemplar la heterogeneidad entre unidades individuales como personas, parlamentos, partidos, países, etc. Para este modelo y los subsiguientes, es posible además especificar un término λ_t que capture la heterogeneidad entre periodos de tiempo. Sin embargo, en paneles con $N > T$, tiene más sentido incluir solamente el componente α_i . Los parámetros del modelo de efectos fijos se estiman por medio de mínimos cuadrados ordinarios.

Por su parte, el modelo de efectos aleatorios (*random effects model*), denominado por otros como modelo de componentes de variancia, es –en apariencia– similar al modelo de efectos fijos en tanto que incorpora un intercepto que varía entre unidades transversales; la diferencia radica en que el modelo de efectos aleatorio asume que el intercepto sigue una distribución de probabilidad, es decir, es una variable aleatoria. Formalmente se define así:

3. Una estrategia similar para abordar la heterogeneidad entre observaciones es el modelaje jerárquico o multinivel (Gelman y Hill, 2006). La principal diferencia entre este enfoque y el presente en el artículo consiste en que en el segundo se asume una secuencia temporal. Sin embargo, ambos se pueden combinar en complejos modelos multinivel de medidas repetidas.

$$y_{it} = \beta_1 x_{1,it} + \beta_2 x_{2,it} + \dots + \beta_k x_{k,it} + \eta_{it} \quad (\text{Ecuación 4})$$

donde η_{it} se descompone en la siguiente forma:

$$\eta_{it} = \alpha_i + u_{it},$$

en el cual el intercepto α_i está idénticamente e independientemente distribuido con media cero y variancia constante y es independiente del error idiosincrático u_{it} .

Ahora bien, si los parámetros del modelo de efectos aleatorios se estiman por mínimos cuadrados ordinarios serán insesgados pero no eficientes, por lo que se recurre al método de mínimos cuadrados generalizados para alcanzar eficiencia (Arellano, 2003: 35).

¿Qué se debe preferir entre los modelos de efectos fijos y efectos aleatorios? Aunque con paneles de T grande no se producen diferencias entre modelos fijos y aleatorios (Hsiao, 2003: 41), en paneles cortos se decide según criterios varios, resumidos en la tabla 2.

TABLA 2.
CRITERIOS DE SELECCIÓN DEL MODELO

	Efectos fijos	Efectos aleatorios
Origen de datos	Poblacionales o no estocásticos	Muestras aleatorias
Tipo de muestreo	Estratificado	Conglomerados
Coefficientes invariantes en el tiempo	No los estima	Sí los estima
Correlación entre variables independientes y efectos no observados	Admite correlación	No admite correlación
Prueba de Hausman	Significativa	No significativa

Fuente: elaboración propia con base en Frees (2004) y Wooldridge (2010).

En primer lugar, si los datos provienen de una muestra aleatoria, como ocurre con las encuestas probabilísticas, se debe preferir los efectos aleatorios. Los efectos fijos puede aplicarse en casos de datos poblacionales (también denominados no estocásticos por Western y Jackman, 1994), donde no se extrae una muestra, sino que se estudia el universo, como todas las provincias, los distritos o las regiones de un país.

En situaciones de muestra complejas o multietápicas, Frees (2004: 75) indica que el modelo de efectos aleatorios debe privilegiarse cuando se realiza muestreo de conglomerados; si el muestreo es estratificado, entonces debe utilizar el modelo de efectos fijos. En esto hay una coincidencia con el primer criterio, pues al muestrear conglomerados se está haciendo una selección aleatoria, mientras que la estratificación recurre a un territorio completo.

En cuanto a los resultados de la estimación, debe atenderse que los modelos de efectos fijos usualmente no estiman coeficientes invariables en el tiempo, mientras que los efectos aleatorios sí. Por ejemplo, si el sistema electoral no cambia en ningún país incluido en el panel, entonces no se podrá estimar su coeficiente bajo el modelo de efectos fijos. A propósito de esta deficiencia, Plümper y Troeger (2007) propusieron una corrección para obtener coeficientes de variables que cambian poco o nada en el tiempo.

Por último, Wooldridge (2010: 286) sostiene que el aspecto más relevante para decidir entre efectos fijos y aleatorios es si existe correlación entre las variables independientes observadas ($x_{k,it}$) y lo efectos no observados (α_i). Los modelos de efectos aleatorios asumen que no existe correlación entre ambos, mientras que los efectos fijos sí la permiten. Para ilustrarlo, una variable explicativa como *sistema electoral* puede estar correlacionada con una variable no observada como la *historia colonial* ya que las colonias británicas suelen tener sistemas mayoritarios (Persson y Tabellini, 2003). Para este caso, sería conveniente asumir que existe correlación y, por ende, utilizar el modelo de efectos fijos.

Para comprobar la correlación entre variables explicativas y los efectos no observados usualmente se ejecuta la prueba de Hausman (1978), la cual compara los estimadores de los coeficientes de regresión de efectos fijos con los coeficientes de efectos aleatorios, bajo la hipótesis nula de que no existen variables omitidas constantes en el tiempo. Si no se rechaza la hipótesis nula bajo algún nivel de significancia, entonces el estimador de efectos aleatorios es más eficiente ya que no existen variables omitidas o efectos no observados. Por el contrario, si se rechaza la hipótesis nula, entonces el estimador de efectos fijos es preferible ya que se existen variables omitidas correlacionadas con las observadas.

Tanto los modelos de efectos fijos como los de efectos aleatorios son apropiados cuando se tratan paneles estáticos. Cuando existen efectos temporales, es decir, donde se quiere modelar la historia o la secuencia temporal, se habla de panel dinámico. Un caso frecuente –pero no el único– de modelo dinámico lo constituye aquel que incluye la variable dependiente rezagada (*lagged dependent variable*) o medida en un punto anterior en el tiempo ($y_{i,t-1}$):

$$y_{it} = \gamma y_{i,t-1} + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \dots + \beta_k x_{k,it} + \alpha_i + u_{it}. \quad (\text{Ecuación 4})$$

El efecto de este rezago se estima por el coeficiente γ , para el cual se asume $|\gamma| < 1$, denominado el «supuesto de estacionariedad».

Este tipo de especificación es útil para considerar factores que tienen un impacto creciente en el tiempo, o bien impactos que decaen (Wawro, 2002). De esta manera, Franklin (2004), en el estudio de la participación electoral, incluyó los efectos de la llamada «inercia del electorado», es decir, los valores previos de participación, justificado por el hecho de que existe una porción del electorado que vota o se abstiene de forma habitual (Plutzer, 2002). Persson y Tabellini (2003) también recurren a los

modelos de variable dependiente rezagada para explicar resultados de políticas públicas y económicas tales como el superávit presupuestario que depende no solo de variables explicativas institucionales y geográficas, sino también de los propios niveles pasados del balance presupuestario. En general, es importante incluir la variable dependiente rezagada únicamente cuando se justifica teóricamente, es decir, cuando existe un proceso dinámico en el que los valores pasados influyen en los presentes (Achen, 2000; Keele y Kelly, 2006).

Para estimar modelos dinámicos de variable dependiente rezagada, se ha probado que tanto los efectos fijos como los efectos aleatorios resultan sesgados debido al carácter endógeno de la variable $y_{i,t-1}$ (Nickell, 1981; Baltagi, 2005): este término –según se puede demostrar– se correlaciona con los errores, lo cual genera coeficientes que en promedio se alejan de los valores reales de los parámetros (sesgo). Por ello, se ha recurrido como alternativa a las variables instrumentales. Una variable instrumental –o un instrumento– es aquella que está correlacionada parcialmente con una variable endógena (en este caso, $y_{i,t-1}$), pero no con el término del error (Wooldridge, 2010: 90). Bajo esta lógica, Anderson y Hsiao (1982) propusieron utilizar $y_{i,t-2}$, es decir, la variable dependiente con dos periodos hacia atrás, así como la diferencia entre el segundo periodo y el primero ($y_{i,t-1} - y_{i,t-2}$) para estimar la variable dependiente diferenciada ($\Delta y_{it} = y_{it} - y_{i,t-1}$).

Por su cuenta, Arellano y Bond (1991) encontraron que los estimadores Anderson-Hsiao son ineficientes (*i. e.*, generan amplios errores estándar), por lo que promueven la estimación utilizando no solo $y_{i,t-2}$, sino todos los rezagos disponibles (excepto $y_{i,t-1}$ que es endógena). Este estimador, que se enmarca dentro del método de los momentos generalizados (*generalized method of moments*), se ha convertido en uno de los más populares en la econometría (Wansbeek, 2012), aunque también existen revisiones y críticas (Arellano y Bover, 1995; Blundell y Bond, 1998; Kiviet, 1995; Roodman, 2009). En ciencia política, el estimador Arellano-Bond ha sido utilizado para predecir el nivel de interés político de las personas en función de su interés en el año anterior (Prior, 2010), para analizar el vínculo entre nivel de democracia e inversión extranjera directa (Doces, 2010), para estudiar la relación entre fragmentación política y déficit presupuestario (Elgie y McMenamin, 2008) y para constatar los determinantes de la corrupción política (Castañeda Rodríguez, 2016), por citar algunos ejemplos.

La tabla 3 ilustra, de forma simplificada, cómo funciona una matriz de instrumentos del modelo Arellano-Bond para un panel de cinco periodos de tiempo. Cuando se está en la medición temporal $t = 1$ no existen rezagos y no se estima la ecuación. En $t = 2$, tampoco se estima la ecuación pues solo se posee un rezago que es endógeno. En $t = 3$ se puede incluir como instrumento $y_{i,t-2}$ únicamente. Ya en $t = 4$ se incluye dos instrumentos ($y_{i,t-2}$ y $y_{i,t-3}$) y en $t = 5$ tres instrumentos ($y_{i,t-2}$, $y_{i,t-3}$ y $y_{i,t-4}$).

El hecho de que para los tiempos 1 y 2 no se puede incluir instrumentos implica que todas las observaciones medidas en dichos periodos se descartan al utilizar el estimador Arellano-Bond. Esto reduce la muestra de manera considerable, pues se pierde el doble del número de unidades individuales ($2 * N$), de forma que el enfoque requiere de paneles con tres o más mediciones temporales.

TABLA 3.

ILUSTRACIÓN DE UNA MATRIZ DE INSTRUMENTOS

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$
y_{t-1}	No existe	Endógena	Endógena	Endógena	Endógena
y_{t-2}	No existe	No existe	Instrumento	Instrumento	Instrumento
y_{t-3}	No existe	No existe	No existe	Instrumento	Instrumento
y_{t-4}	No existe	No existe	No existe	No existe	Instrumento
y_{t-5}	No existe	No existe	No existe	No existe	No existe

Fuente: elaboración propia.

CASO DE APLICACIÓN DE PARTICIPACIÓN ELECTORAL COMPARADA

El estudio de la participación electoral se ha convertido en un tema clásico de la ciencia política, y acumula una vasta cantidad de trabajos (Cancela y Geys, 2016; Geys, 2006; Smets y van Ham, 2013; Stockemer, 2017). Gran parte de ellos se basa en la comparación de los porcentajes de asistencia a las urnas entre varios países o unidades subnacionales en el tiempo, con lo que forma así paneles de elecciones (Cancela y Geys recogen 185 estudios; Stockemer 130). Por tanto, no sorprende que en dicho estilo de análisis agregado se recurra a modelos de regresión para datos de panel como los expuestos anteriormente. Sin embargo, se han detectado errores metodológicos en el tratamiento de estos datos: no siempre se incluye una discusión o justificación sobre el modelo apropiado para los datos; en ocasiones no se indica el método de estimación; se utilizan efectos fijos para modelos dinámicos (contrario a lo que la teoría sugiere), o se rehúye de examinar la robustez de las estimaciones al no comparar diversos modelos.

Dada la popularidad de los datos de panel en el estudio de la participación electoral agregada, así como las omisiones en sus análisis, el siguiente ejemplo utiliza una base de datos original de 68 elecciones presidenciales de primera vuelta en América Latina (1995-2011) para contrastar los resultados de cuatro métodos (Pignataro 2014b).

El macropanel electoral escogido presenta algunas propiedades interesantes. En primer lugar, mantiene la característica de tener mayor número de casos espaciales ($N = 18$) que mediciones temporales (T_i mínimo = 2; T_i máximo = 5), como es común en estudios electorales. Claramente es un panel corto en el tiempo que involucra retos en su análisis. Segundo, los periodos constitucionales varían entre países (por ejemplo, cuatro años en Costa Rica, cinco en Panamá y seis en México) y existen interrupciones en los periodos de presidentes (de la Rúa en Argentina, Bucaram en Ecuador, Fujimori en Perú y Lucio Gutiérrez en Bolivia), por lo que no todas las unidades de análisis cuentan con la misma cantidad de mediciones temporales. Se trata, pues, de un panel desbalanceado.

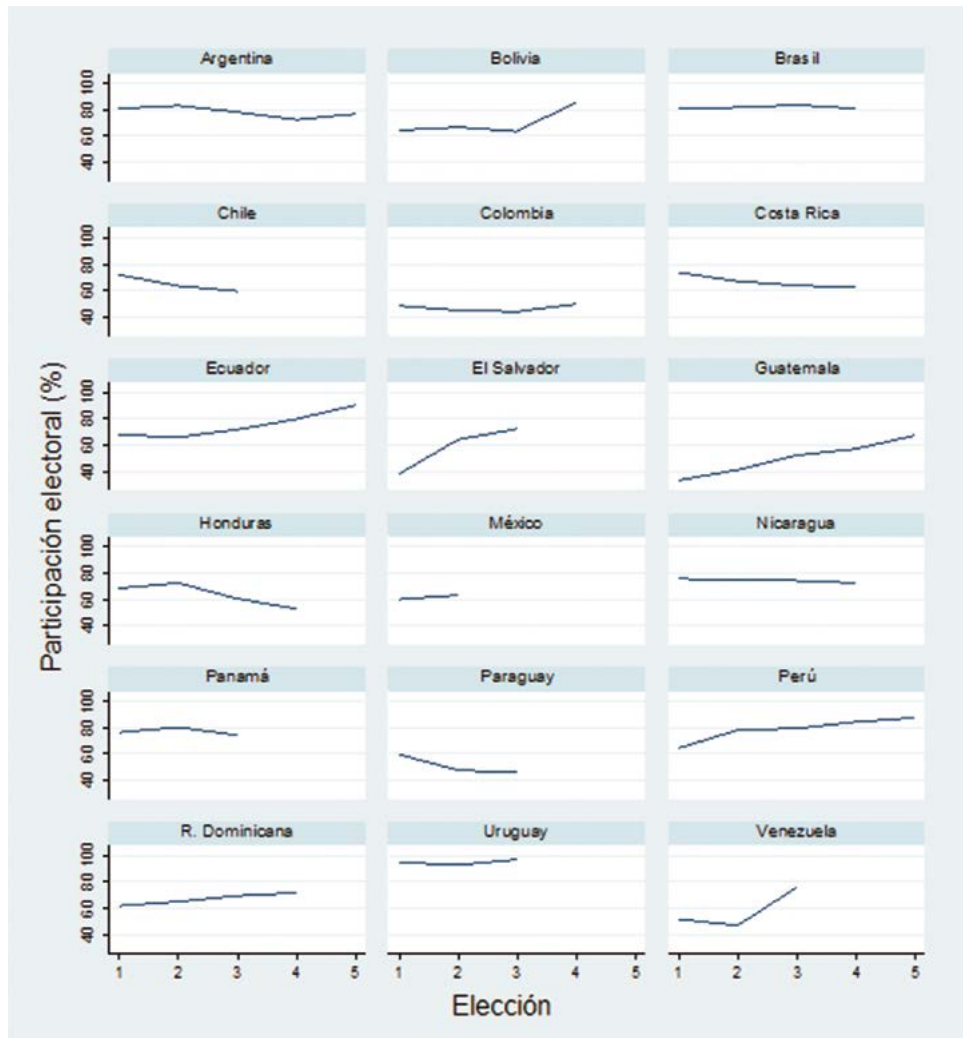
En este ejemplo se busca predecir el nivel de participación electoral, medido como porcentaje de votos entre la población en edad de votar (figura 1)⁴, con base en variables del marco teórico del institucionalismo de elección racional (Blais, 2000; Dettrey y Schwindt-Bayer, 2009; Downs, 1957; Franklin, 2004; Norris, 2004; Pérez-Liñán, 2001). El argumento general consiste en que características institucionales y políticas incentivan la participación, de forma que la variabilidad entre los porcentajes de asistencia a las urnas está explicada por la inercia del electorado, la competencia electoral, el poder presidencial institucional, la mayoría parlamentaria, la obligatoriedad del voto y el modo de empadronamiento⁵. No se incluyen variables socioeconómicas, ya que han demostrado tener escasa validez en la región latinoamericana (Fornos *et al.*, 2004; Lavez-zolo, 2008) por lo que resulta más pertinente analizar los citados factores institucionales y políticos.

En primera instancia, la inercia del electorado, medida como la participación previa ($t - 1$), la cual captura el efecto del hábito del voto (Alfaro-Redondo, 2014; Franklin, 2004; Pluzter, 2002), debería incrementar la participación. Este análisis se convierte, entonces, en uno de tipo dinámico al incluir variable dependiente rezagada.

Luego, se espera encontrar una relación negativa entre competencia electoral y participación, ya que cuanto mayor sea la diferencia entre los votos por un candidato y otro, menor es la intensidad de competición, por lo que la posibilidad de influir en dicho resultado es menor y se desincentiva la participación (Blais, 2000; Downs, 1957; Franklin, 2004). Por otro lado, cuanto mayor sea la capacidad del presidente para generar políticas (beneficios), mayor será la motivación para votar. Por ello, el poder presidencial institucional (Dettrey y Schwindt-Bayer, 2009) y la mayoría parlamentaria deberían relacionarse de forma positiva con la variable dependiente (Franklin, 2004). También se prevé un signo positivo para el coeficiente de la obligatoriedad del voto y del modo de registro automático (Blais y Dobrzynksa, 1998; Fornos *et al.*, 2004; Franklin, 2004; Jackman, 1987; Norris, 2004; Powell, 1986; Schraufnagel y Sgouraki, 2005). El sistema electoral presidencial se incluye como variable de control⁶.

-
4. Los porcentajes de participación electoral se obtuvieron principalmente de los organismos electorales de cada país así como de las bases de datos disponibles en *International IDEA* (<https://www.idea.int/>) y *Political Database of the Americas* (<http://pdba.georgetown.edu/>).
 5. Como bien advirtió un revisor de este manuscrito, el análisis agregado permite realizar inferencias únicamente para la elección y no para individuos, pues de lo contrario se estaría incurriendo en una falacia ecológica. Sería, pues, recomendable utilizar un modelaje multinivel que capture tanto efectos en niveles agregados (elección) e individuales (electores). Sin embargo, ello significaría adoptar técnicas distintas de las presentadas en la sección anterior y el artículo se distanciaría del objetivo central de comparar las herramientas para el análisis de datos de panel.
 6. Competencia electoral se mide como la diferencia en puntos porcentuales entre el primer y el segundo candidato con más votos en la primera vuelta. El poder presidencial institucional consiste en el índice elaborado por García (2009), el cual oscila entre 0 (menor poder) y 1 (mayor

FIGURA 2.
PARTICIPACIÓN ELECTORAL SEGÚN LA POBLACIÓN EN EDAD DE VOTAR



Fuente: elaboración propia con base en Pignataro (2014b).

poder). El estatus mayoritario corresponde al total de votos obtenido en la Cámara Baja por el partido (o coalición) del candidato presidencial ganador de la primera vuelta. El voto obligatorio se codifica con 0 si el voto no es obligatorio, 0.5 si es obligatorio sin sanciones y 1 si es obligatorio con sanciones. El modo de registro electoral asume el valor 0 si es obligatorio o voluntario y 1 si es automático. El sistema de elección presidencial indica 0 si el presidente se elige por mayoría simple y 1 por mayoría absoluta o con umbral reducido. Para mayor información sobre las variables y las fuentes, consúltese Pignataro (2014a).

Para probar estas hipótesis se formulan los modelos de regresión para datos de panel que se expusieron previamente: efectos fijos, efectos aleatorios y estimador Arellano-Bond. El modelo de efectos fijos estima la siguiente ecuación:

$$\begin{aligned} Participación_{it} = & \beta_1 Inercia_{it} + \beta_2 Competencia_{it} + \beta_3 Poder_{it} + & \text{(Ecuación 5)} \\ & + \beta_4 Estatus_{it} + \beta_5 Obligatorio_{it} + \beta_6 Registro_{it} + \beta_7 Sistema_{it} + \alpha_i + u_{it}; \end{aligned}$$

Por otro lado, el modelo de efectos aleatorios la estima como:

$$\begin{aligned} Participación_{it} = & \beta_1 Inercia_{it} + \beta_2 Competencia_{it} + \beta_3 Poder_{it} + & \text{(Ecuación 6)} \\ & + \beta_4 Estatus_{it} + \beta_5 Obligatorio_{it} + \beta_6 Registro_{it} + \beta_7 Sistema_{it} + \eta_{it}, \end{aligned}$$

con $\eta_{it} = \alpha_i + u_{it}$.

Estos modelos se suponen adecuados para tratar la heterogeneidad del panel, aunque no la característica dinámica que implica la variable de inercia. Por ello se comparan con el método Arellano-Bond que se supone sería conveniente. La ecuación ahora reescribe la variable de inercia con la notación habitual de rezagos⁷:

$$\begin{aligned} Participación_{it} = & \gamma Participación_{i,t-1} + \beta_1 Competencia_{it} + \beta_2 Poder_{it} + & \text{(Ecuación 7)} \\ & + \beta_3 Estatus_{it} + \beta_4 Obligatorio_{it} + \beta_5 Registro_{it} + \beta_6 Sistema_{it} + \alpha_i + u_{it}. \end{aligned}$$

Adicionalmente, los anteriores modelos se comparan con la estimación de un modelo de intercepto constante por mínimos cuadrados ordinarios, que obvia las características de panel. Al incorporar estos resultados se podrá constatar cómo un modelo que ignora la heterogeneidad se diferencia de aquellos que sí la contemplan:

$$\begin{aligned} Participación_{it} = & \alpha + \beta_1 Inercia_{it} + \beta_2 Competencia_{it} + \beta_3 Poder_{it} + & \text{(Ecuación 8)} \\ & + \beta_4 Estatus_{it} + \beta_5 Obligatorio_{it} + \beta_6 Registro_{it} + \beta_7 Sistema_{it} + u_{it}. \end{aligned}$$

La estimación de los cuatro modelos, realizada con el paquete Stata⁸, se incluye en la tabla 4⁹.

-
7. El estimador Arellano-Bond se puede calcular en uno o dos pasos; se recomienda hacerlo en un paso (*one-step*) por su mayor eficiencia (Arellano y Bond, 1991).
 8. Con el mismo fin es posible utilizar el paquete «plm» en R.
 9. Adviértase que en los primeros tres modelos el número de observaciones desciende a 50 ya que el cálculo de la participación previa implica la pérdida de observaciones en la primera elección ($t = 1$). En el caso del estimador Arellano-Bond se pierde el doble de observaciones porque utiliza un mínimo de dos rezagos como instrumento, disminuyendo a 32.

TABLA 4.

MODELOS ESTIMADOS PARA PREDECIR LA PARTICIPACIÓN ELECTORAL EN AMÉRICA LATINA

	Intercepto constante		Efectos fijos		Efectos aleatorios		Arellano-Bond	
	Coef. (e.e.)	Sig.	Coef. (e.e.)	Sig.	Coef. (e.e.)	Sig.	Coef. (e.e.)	Sig.
Inercia (Participación _{t-1})	0,706 (0,097)	0,000	0,541 (0,176)	0,005	0,659 (0,116)	0,000	1,286 (0,781)	0,099
Competencia electoral	-0,064 (0,149)	0,670	-0,045 (0,155)	0,775	-0,049 (0,141)	0,731	0,125 (0,282)	0,657
Poder presidencial institucional	-0,032 (0,106)	0,765	0,463 (0,436)	0,297	-0,038 (0,141)	0,787	0,329 (0,735)	0,654
Estatus mayoritario	0,194 (0,095)	0,048	0,212 (0,105)	0,053	0,218 (0,093)	0,019	0,125 (0,160)	0,432
Obligatoriedad del voto	4,769 (3,694)	0,204	30,149 (15,001)	0,055	6,171 (4,942)	0,212	31,727 (24,756)	0,200
Modo de registro	-0,390 (2,301)	0,866	(omitido)		0,004 (3,460)	0,999	(omitido)	
Sistema electoral presidencial	6,812 (3,006)	0,029	(omitido)		6,638 (3,964)	0,094	(omitido)	
Observaciones	50		50		50		32	
R ²	0,730		0,475		0,728		0,606	

Fuente: elaboración propia.

El primer modelo, llamado de intercepto constante, encuentra significativas ($p < 0,05$) las variables *estatus mayoritario* y *sistema electoral presidencial*, ambas con el signo esperado. También la variable *inercia* es significativa ($p < 0,01$) y positiva, lo que significa que cuanto mayor haya sido la participación previa, mayor será la participación promedio. Las demás variables no son significativas bajo los umbrales tradicionales. Es este modelo el que resulta más eficiente (con menores errores estándar); sin embargo, esta estimación se dice poco confiable pues un modelo de intercepto común con datos de panel sobreestima o subestima las variancias de los coeficientes (Beck y Katz, 1995).

En el modelo de efectos fijos son significativas *inercia* ($p < 0,01$), *estatus mayoritario* y *obligatoriedad del voto* ($p < 0,10$), con los signos positivos previstos. Obsérvese que este modelo omite la estimación de los coeficientes modo de registro y sistema electoral, ya que son constantes en el tiempo. Bajo el modelo de efectos aleatorios resultan significativas la *inercia* ($p < 0,01$), *estatus mayoritario* ($p < 0,05$) y *sistema electoral* ($p < 0,10$); los signos coinciden con lo esperado.

Finalmente, el estimador Arellano-Bond es el de menor eficiencia (mayores errores estándar), por lo que no encuentra ninguna variable significativa estadísticamente exceptuando la *participación previa* ($p < 0,10$) con un coeficiente mayor a 1, muy distinto a los correspondientes para los otros modelos. Se nota entonces que los modelos de efectos fijos y efectos aleatorios presentan mayor eficiencia que el Arellano-Bond, pese a que estos primeros se suponen inapropiados para paneles dinámicos con la variable dependiente rezagada.

En general, las estimaciones son poco estables, pues los coeficientes varían entre sí, exceptuando estatus mayoritario. Particularmente llama la atención que el coeficiente de la variable dependiente rezagada sea similar entre los modelos de intercepto constante, efectos fijos y efectos aleatorios, mas no con Arellano-Bond.

En relación con la bondad de ajuste, los coeficientes de determinación (R^2) indican que los modelos con mayor variancia explicada son el intercepto constante y el de efectos aleatorios: 73 % cada uno. El Arellano-Bond explica un 61 % y el de efectos fijos un 48 %. La prueba de Hausman, que compara efectos fijos con aleatorios, no permite rechazar la hipótesis nula de que los coeficientes son iguales ($\chi^2_{(5)} = 5,69$; $p = 0,347$), lo cual se puede interpretar que no existen variables omitidas; en consecuencia, es preferible el modelo de efectos aleatorios.

De escogerse un modelo entre los cuatro estimados puede descartarse, en primer lugar, el de intercepto constante, pues las previsiones teóricas no permiten confiar en sus estimaciones, aunque vale destacar que los coeficientes no resultan excesivamente diferentes del resto. Por su parte, los modelos de efectos fijos y aleatorios presentan buena eficiencia. Si se sigue el criterio del tipo de datos, puesto que las elecciones de cada país no se seleccionan aleatoriamente, sino que corresponden a una población, entonces tendría más sentido el modelo de efectos fijos. No obstante, en términos de porcentaje de variancia explicada y según la prueba de Hausman, los efectos aleatorios aparecen como ganadores.

El modelo Arellano-Bond, aunque pretende resolver los problemas de endogeneidad de la variable rezagada, no se ajusta bien a los datos: su porcentaje de explicación es bajo, los errores estándar son amplios y la estimación del coeficiente de rezago es inconsistente con el supuesto de estacionariedad, pues resulta mayor a 1.

Este último resultado es llamativo y se pueden conjeturar algunas explicaciones al respecto. Primero, puede objetarse que el número de variables independientes sea muy grande respecto al número de casos, lo cual perjudica la eficiencia en la estimación. Sin embargo, al estimar modelos con subconjuntos de variables, el coeficiente de inercia permanece mayor a 1 (resultados no se muestran).

Segundo, la estimación pudo haber fallado porque el panel es muy corto en el tiempo y el instrumento resulta «pobre» (véase Arellano, 2003: 96). En otras palabras, se disponen de pocos rezagos para estimar el parámetro (ver tabla 3). Por ejemplo, siete países tienen cuatro elecciones, es decir, únicamente dos instrumentos; seis países tienen tres elecciones, por lo tanto, solo un instrumento; México, con dos elecciones, queda fuera de la estimación. Con panel más largo en el tiempo o completamente balanceado la estimación podría resultar más precisa.

Para probar esta interpretación, se realizaron simulaciones Monte Carlo con las cuales se puede examinar si en un proceso de generación de datos los modelos responden de forma diferenciada ante condiciones impuestas (Mooney, 1997); en concreto, si la estimación se ve afectada por el desbalance o por el número de variables independientes en relación con los casos.

Para ello se compara un panel completamente balanceado de $N = 16$ con cinco mediciones temporales (ochenta observaciones) con otro de $N = 16$ casos, pero desbalanceado (la mitad de casos con cinco periodos de tiempo y la otra mitad con tres, para un total de 64 observaciones), formulando un modelo de pocas variables independientes (una métrica, una categórica y la variable dependiente rezagada). Para cada escenario se ejecutan mil iteraciones.

El proceso de generación de datos sigue el siguiente modelo:

$$y_{it} = \gamma y_{i,t-1} + \beta_1 x_{it} + \beta_2 q_{it} + \alpha_i + u_{it} \quad (\text{Ecuación 9})$$

con $i = 1, \dots, N$; $t = 1, \dots, T$, donde $y_{i,t-1}$ es la variable dependiente con un rezago, x_{it} es una variable métrica definida como $x_{it} = 0,8 * x_{i,t-1} + \varphi_{it}$, donde $\varphi_{it} \sim N(0; 20)$, q_{it} es una variable categórica con distribución binomial ($p = 0,7$), $u_{it} \sim N(0; 1)$ y $\alpha_i \sim N(0; 1)$. Además, se asignan los siguientes valores a los parámetros: $\gamma = 0,8$; $\beta_1 = 0,2$; y $\beta_2 = 0,5$ (el diseño se basa en Arellano y Bond, 1991; Beck y Katz, 2004; Flannery y Watson Hankins, 2013; Kiviet, 1995).

La tabla 5 muestra las medias de las mil réplicas según los cuatro modelos en los dos escenarios propuestos. La mayor precisión se obtiene para los coeficientes de las variables β_1 y β_2 . Para el parámetro de la variable dependiente rezagada γ presentan menor sesgo, en orden descendente, el modelo de efectos aleatorios, el estimador Arellano-Bond, el modelo de efectos fijos y el modelo de intercepto constante. Nótese que mientras los modelos de efectos aleatorios y de intercepto constante sobreestiman el valor del parámetro, los efectos fijos y el Arellano-Bond lo subestiman. Estos hallazgos son invariables entre el panel balanceado y el no balanceado. Esto implica que la presencia de casos con pocos periodos de tiempo y escasos instrumentos –como en el ejemplo de participación electoral– no perjudica necesariamente la estimación mediante el método de Arellano-Bond, por lo que el resultado anómalo obedece a otros factores no explorados.

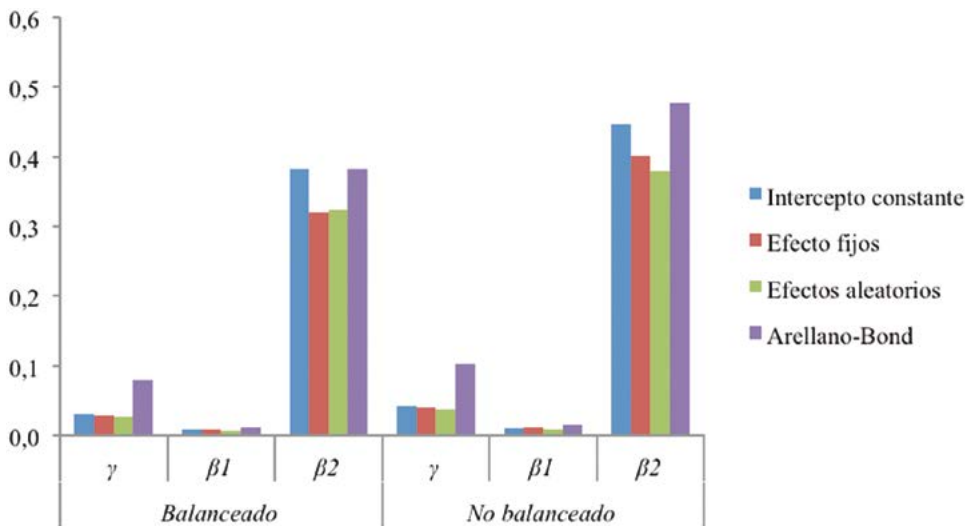
Las simulaciones coinciden con el panel electoral en detectar que el estimador Arellano-Bond es menos eficiente pues estima mayores errores estándar promedio que los modelos de efectos fijos y efectos aleatorios (figura 3). El desbalance, por su parte, disminuye la eficiencia en todos los modelos ya que implica un menor número de observaciones. Sin embargo, en comparación con los modelos de participación electoral, las simulaciones generan estimaciones bastante más estables, incluso en el escenario de desbalance y no se predice un parámetro de la variable dependiente rezagada mayor a 1.

TABLA 5.
MEDIAS DE LOS COEFICIENTES ESTIMADOS EN LAS SIMULACIONES

		Intercepto constante	Efecto fijos	Efectos aleatorios	Arellano-Bond
Balanceado	$\gamma = 0,8$	0,835	0,782	0,807	0,787
	$\beta_1 = 0,2$	0,195	0,200	0,199	0,199
	$\beta_2 = 0,5$	0,501	0,491	0,495	0,499
No balanceado	$\gamma = 0,8$	0,835	0,777	0,805	0,782
	$\beta_1 = 0,2$	0,196	0,200	0,200	0,199
	$\beta_2 = 0,5$	0,502	0,487	0,497	0,494

Fuente: elaboración propia con base en simulaciones (1000 réplicas).

FIGURA 3.
MEDIAS DE LOS ERRORES ESTÁNDAR DE LOS COEFICIENTES ESTIMADOS EN LAS SIMULACIONES



Fuente: elaboración propia con base en simulaciones (1000 réplicas).

CONCLUSIONES

Parafraseando a Barbara Geddes, este artículo expone «cómo los datos y el modelo que uno escoge afectan las respuestas que uno obtiene» (2003). Por muchas razones, los datos de panel superan limitaciones de los transversales: contemplan la variabilidad

temporal; miden con mayor precisión el antes y el después; aumentan los grados de libertad, y contemplan la heterogeneidad entre los casos. Con ello se obtiene mayor potencial para contrastar hipótesis y teorías; no sorprende, en consecuencia, que haya incrementado su uso en la ciencia política.

Esta ganancia conlleva la necesidad de modelos especiales, más allá de la tradicional regresión lineal estimada por mínimos cuadrados ordinarios. Los modelos de efectos fijos y efectos aleatorios incluyen la heterogeneidad entre casos que conforman un panel, lo cual no se logra con un modelo de intercepto constante. Los modelos dinámicos incorporan efectos temporales, a través de rezagos en la variable dependiente, aunque ello induce endogeneidad. Por ende, se han planteado soluciones desde el enfoque de variables instrumentales, como el reconocido estimador Arellano-Bond.

Con el ejemplo de participación electoral se ha intentado ilustrar algunos retos al analizar datos de panel. Por un lado, el hecho de que las elecciones se lleven a cabo en intervalos de más de un año –en oposición a mediciones más frecuentes como las económicas– reduce el número de observaciones temporales; por otro, puesto que los comicios no están distanciados por periodos de tiempo iguales, se genera desbalance.

Los modelos de efectos fijos y efectos aleatorios –aunque teóricamente se decían inapropiados cuando existe una variable dependiente rezagada la cual resulta endógena– estimaron con precisión y poco error los coeficientes; no así el Arellano-Bond. Las simulaciones, realizadas bajo condiciones de pocas variables y que compararon un panel balanceado con uno desbalanceado similar al panel electoral, confirmaron la robustez de los modelos de efectos fijos y efectos aleatorios y la ineficiencia del Arellano-Bond. A la vez, la estimación de intercepto constante, es decir, el modelo clásico de mínimos cuadrados ordinarios, que ignora lo característico del panel, muestra comparativamente el mayor sesgo. Por lo tanto, si el Arellano-Bond resulta ineficiente y el modelo de intercepto constante sesgado, no hay razón para aplicarlos en el análisis de modelos dinámicos de panel.

Por ello, compartimos una agenda metodológica que ha advertido la importancia de las especificaciones técnicas en el estudio de datos de panel (Beck y Katz, 1995; Beck, 2007; Wilson y Butler, 2007). De esta investigación, la conclusión más relevante es prescribir, en general, el uso de datos de panel por sus fortalezas metodológicas y recomendar para los análisis de paneles cortos en el tiempo, como los presentes en estudios electorales, los modelos de efectos fijos y efectos aleatorios por su robustez y eficiencia.

AGRADECIMIENTOS

El autor agradece las sugerencias y comentarios de los evaluadores anónimos y de los editores de esta revista, así como las observaciones de Gilbert Brenes, Juan Manuel Muñoz y Mariano Torcal a versiones previas de este manuscrito.

Referencias

- Achen, Christopher. 2000. «Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables», en *Annual Meeting of the Political Methodology Section. American Political Science Association*. University of California, Los Angeles.
- Alfaro-Redondo, Ronald. 2014. «Lifecycle changes and the activation of habitual voting: The case of Costa Rica», *Electoral Studies*, 35: 188-199. Disponible en: <https://doi.org/10.1016/j.electstud.2014.06.003>.
- Anderson, Theodore W. y Cheng Hsiao. 1982. «Formulation and estimation of dynamic models using panel data», *Journal of Econometrics*, 18: 47-82. Disponible en: [https://doi.org/10.1016/0304-4076\(82\)90095-1](https://doi.org/10.1016/0304-4076(82)90095-1).
- Arellano, Manuel. 2003. *Panel Data Econometrics*. New York: Oxford University Press. Disponible en: <https://doi.org/10.1093/0199245282.001.0001>.
- Arellano, Manuel y Stephen Bond. 1991. «Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations», *The Review of Economic Studies*, 58 (2): 277-297. Disponible en: <https://doi.org/10.2307/2297968>.
- Arellano, Manuel y Olympia Bover. 1995. «Another look at the instrumental variable estimation of error-components models», *Journal of Econometrics*, 68 (1): 29-51. Disponible en: [https://doi.org/10.1016/0304-4076\(94\)01642-D](https://doi.org/10.1016/0304-4076(94)01642-D).
- Baltagi, Badi H. 2005. *Econometric Analysis of Panel Data*. Chichester: John Wiley and Sons.
- Bartels, Larry M. 1999. «Panel Effects in the American National Election Studies», *Political Analysis*, 8 (1): 1-20. Disponible en: <https://doi.org/10.1093/oxfordjournals.pan.a029802>.
- Beck, Nathaniel. 2007. «From Statistical Nuisances to Serious Modeling: Changing How We Think About the Analysis of Time-Series–Cross-Section Data», *Political Analysis*, 15 (2): 97-100. Disponible en: <https://doi.org/10.1093/pan/mpm001>.
- Beck, Nathaniel. 2008. «Time-series Cross-section Methods», en Janet M. Box-Steffensmeier, Henry E. Brady y David Collier, (eds.), *The Oxford Handbook of Political Methodology*. New York: Oxford University Press.
- Beck, Nathaniel y Jonathan N. Katz. 1995. «What to do (and not to do) with Time-Series Cross-Section Data», *The American Political Science Review*, 89 (3): 634-647. Disponible en: <https://doi.org/10.2307/2082979>.
- Beck, Nathaniel y Jonathan N. Katz. 2004. «Time-Series-Cross-Section Issues: Dynamics», en *Annual Meeting of the Society for Political Methodology*. Stanford: Stanford University.
- Beck, Nathaniel y Jonathan N. Katz. 2011. «Modeling Dynamics in Time-Series–Cross-Section Political Economy Data», *Annual Review of Political Science*, 14: 331-52. Disponible en: <https://doi.org/10.1146/annurev-polisci-071510-103222>.

- Blais, André. 2000. *To Vote or Not to Vote. The Merits and Limits of Rational Choice Theory*. Pittsburgh: University of Pittsburgh Press. Disponible en: <https://doi.org/10.2307/j.ctt5hjrrf>.
- Blais, André y Agnieszka Dobrzynska. 1998. «Turnout in electoral democracies», *European Journal of Political Research*, 33: 239-261. Disponible en: <https://doi.org/10.1111/1475-6765.00382>.
- Blundell, Richard y Stephen Bond. 1998. «Initial conditions and moment restrictions in dynamic panel data models», *Journal of Econometrics*, 87: 115-143. Disponible en: [https://doi.org/10.1016/S0304-4076\(98\)00009-8](https://doi.org/10.1016/S0304-4076(98)00009-8).
- Brady, Henry. 2008. «Causation and Explanation in Social Science», en Janet M. Box-Steffensmeier, Henry E. Brady y David Collier (eds.), *The Oxford Handbook of Political Methodology*. New York: Oxford University Press.
- Cancela, João y Benny Geys. 2016. «Explaining voter turnout: A meta-analysis of national and subnational elections», *Electoral Studies*, 42: 264-275. Disponible en: <https://doi.org/10.1016/j.electstud.2016.03.005>.
- Castañeda Rodríguez, Víctor Mauricio. 2016. «Una investigación sobre la corrupción pública y sus determinantes», *Revista Mexicana de Ciencias Políticas y Sociales*, 61 (227): 103-136. Disponible en: [https://doi.org/10.1016/S0185-1918\(16\)30023-X](https://doi.org/10.1016/S0185-1918(16)30023-X).
- Dalton, Russell J. 2000. «The Decline of Party Identifications», en Russell J. Dalton y Martin P. Wattenberg (eds.), *Parties without Partisans. Political Change in Advanced Industrial Democracies*. New York: Oxford University Press.
- De Benedetto, Marco Alberto y Maria De Paola. 2016. «The impact of incumbency on turnout. Evidence from Italian municipalities», *Electoral Studies*, 44: 98-108. Disponible en: <https://doi.org/10.1016/j.electstud.2016.06.012>.
- Dettrey, Bryan J. y Leslie A. Schwindt-Bayer. 2009. «Voter Turnout in Presidential Democracies», *Comparative Political Studies*, 42 (10): 1317-1338. Disponible en: <https://doi.org/10.1177/0010414009332125>.
- Dilliplane Susanna. 2014. «Activation, Conversion, or Reinforcement? The Impact of Partisan News Exposure on Vote Choice?», *American Journal of Political Science*, 58 (1): 79-94. Disponible en: <https://doi.org/10.1111/ajps.12046>.
- Doces, John A. 2010. «The Dynamics of Democracy and Direct Investment: An Empirical Analysis», *Polity*, 42 (3): 329-351. Disponible en: <https://doi.org/10.1057/pol.2010.1>.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.
- Elgie, Robert y Iain McMenamin. 2008. «Political Fragmentation, Fiscal Deficits and Political Institutionalisation», *Public Choice*, 136 (3/4): 255-267. Disponible en: <https://doi.org/10.1007/s11127-008-9294-x>.
- Escobar-Lemmon, Maria y Michelle M. Taylor-Robinson. 2005. «Women Ministers in Latin American Government: When, Where, and Why?», *American Journal of Political Science*, 49 (4): 829-844. Disponible en: <https://doi.org/10.1111/j.1540-5907.2005.00158.x>.

- Fairbrother, Malcolm. 2014. «Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets», *Political Science Research and Methods*, 2 (1): 119-140. Disponible en: <https://doi.org/10.1017/psrm.2013.24>.
- Flannery, Mark J. y Kristine Watson Hankins. 2013. «Estimating dynamic panel models in corporate finance», *Journal of Corporate Finance*, 19: 1-19. Disponible en: <https://doi.org/10.1016/j.jcorpfin.2012.09.004>.
- Fornos, Carolina A., Timothy J. Power y James C. Garand. 2004. «Explaining Voter Turnout in Latin America, 1980 to 2000», *Comparative Political Studies*, 37 (8): 909-940. Disponible en: <https://doi.org/10.1177/0010414004267981>.
- Fortin-Rittberger, Jessica. 2015. «Time-series cross-section», en Henning Best y Christof Wolf (eds.), *The SAGE Handbook of Regression Analysis and Causal Inference*. London: SAGE.
- Franklin, Mark N. 2004. *Voter Turnout and the Dynamics of Electoral Competition in Established Democracies since 1945*. New York: Cambridge University Press. Disponible en: <https://doi.org/10.1017/CBO9780511616884>.
- Franzese, Robert J. 2007. «Multicausality, Context-Conditionality, and Endogeneity», en Carles Boix, y Susan C. Stokes (eds.), *The Oxford Handbook of Comparative Politics*. New York: Oxford University Press.
- Frees, Edward W. 2004. *Longitudinal and Panel Data. Analysis and Applications in the Social Sciences*. New York: Cambridge University Press. Disponible en: <https://doi.org/10.1017/CBO9780511790928>.
- García, Mercedes. 2009. *Presidentes y Parlamentos: ¿quién controla la actividad legislativa en América Latina?* Madrid: Centro de Investigaciones Sociológicas.
- Garzia, Diego. 2013. «Changing Parties, Changing Partisans: The Personalization of Partisan Attachments in Western Europe», *Political Psychology*, 34 (1): 67-89. Disponible en: <https://doi.org/10.1111/j.1467-9221.2012.00918.x>.
- Geddes, Barbara. 2003. *Paradigms and Sand Castles. Theory Building and Research Design in Comparative Politics*. Ann Arbor: The University of Michigan Press.
- Gelman, Andrew y Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. Disponible en: <https://doi.org/10.1017/CBO9780511790942>.
- Gerring, John. 2012. *Social Science Methodology. A Unified Framework*. New York: Cambridge University Press.
- Geys, Benny. 2006. «Explaining voter turnout: A review of aggregate-level research», *Electoral Studies*, 25: 637-663. Disponible en: <https://doi.org/10.1016/j.electstud.2005.09.002>.
- Gujarati, Damodar N. y Dawn C. Porter. 2010. *Econometría*. México: McGraw Hill.
- Hausman, Jerry A. 1978. «Specification Tests in Econometrics», *Econometrica*, 46 (6): 1251-1271. Disponible en: <https://doi.org/10.2307/1913827>.
- Hsiao, Cheng. 2003. *Analysis of Panel Data*. New York: Cambridge University Press. Disponible en: <https://doi.org/10.1017/CBO9780511754203>.

- Jackman, Robert W. 1987. «Political Institutions and Voter Turnout in the Industrial Democracies», *The American Political Science Review*, 81 (2): 405-424. Disponible en: <https://doi.org/10.2307/1961959>.
- Keele, Luke y Nathan J. Kelly. 2006. «Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables», *Political Analysis*, 14: 186-205. Disponible en: <https://doi.org/10.1093/pan/mpj006>.
- King, Gary, Robert O. Keohane y Sidney Verba. 1994. *Designing Social Inquiry. Scientific Inference in Qualitative Research*. New Jersey: Princeton University Press.
- Kiviet, Jan F. 1995. «On bias, inconsistency, and efficiency of various estimators in dynamic panel data models», *Journal of Econometrics*, 68: 53-78. Disponible en: [https://doi.org/10.1016/0304-4076\(94\)01643-E](https://doi.org/10.1016/0304-4076(94)01643-E).
- Lavezzolo, Sebastián. 2008. «Adversidad económica y participación electoral en América Latina, 1980-2000», *Revista Española de Ciencia Política*, 18: 67-93.
- Lewis-Beck, Michael S., Richard Nadeau y Angelo Elias. 2008. «Economics, Party, and the Vote: Causality Issues and Panel Data», *American Journal of Political Science*, 52 (1): 84-95. Disponible en: <https://doi.org/10.1111/j.1540-5907.2007.00300.x>.
- Lipset, Seymour M. 1959. «Some Social Requisites of Democracy: Economic Development and Political Legitimacy», *The American Political Science Review*, 53 (1): 69-105. Disponible en: <https://doi.org/10.2307/1951731>.
- Lupu, Noam. 2015. «Partisanship in Latin America», en Ryan E. Carlin, Matthew M. Singer y Elizabeth J. Zechmeister (eds.), *The Latin American Voter. Pursuing Representation and Accountability in Challenging Contexts*. Ann Arbor: University of Michigan Press.
- Lynn, Peter. 2009. «Methods for Longitudinal Surveys», en Peter Lynn (ed.), *Methodology of Longitudinal Surveys*. Chichester: John Wiley and Sons. Disponible en: <https://doi.org/10.1002/9780470743874.ch1>.
- Mooney, Christopher Z. 1997. *Monte Carlo Simulation*. California: Sage. Disponible en: <https://doi.org/10.4135/9781412985116>.
- Moreno Martínez, Cristina. 2010. «El efecto de la campaña para las elecciones generales españolas de 2008 sobre la información política y la participación electoral de los votantes: ¿se puede hablar de una función de legitimación de las campañas electorales?», *Revista Española de Ciencia Política*, 24: 53-81.
- Norris, Pippa. 2004. *Electoral Engineering. Voting Rules and Political Behavior*. New York: Cambridge University Press. Disponible en: <https://doi.org/10.1017/CBO9780511790980>.
- Nickell, Stephen. 1981. «Biases in Dynamic Models with Fixed Effects», *Econometrica*, 49 (6): 1417-1426. Disponible en: <https://doi.org/10.2307/1911408>.
- Pérez-Liñán, Aníbal. 2001. «Neoinstitutional accounts of voter turnout: moving beyond industrial democracies», *Electoral Studies*, 20: 281-297. Disponible en: [https://doi.org/10.1016/S0261-3794\(00\)00019-6](https://doi.org/10.1016/S0261-3794(00)00019-6).
- Persson, Torsten y Guido Tabellini. 2003. *The Economic Effects of Constitutions*. Cambridge: MIT Press.

- Pignataro, Adrián. 2014a. «Participación electoral comparada en América Latina: un modelo desde la teoría de elección racional», *Revista Derecho Electoral*, 17: 154-184.
- Pignataro, Adrián. 2014b. Base de datos de elecciones en América Latina (1995-2011) [archivo Excel].
- Plümper, Thomas y Vera E. Troeger. 2007. «Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects», *Political Analysis*, 15 (2): 124-139. Disponible en: <https://doi.org/10.1093/pan/mpm002>.
- Plümper, Thomas, Vera E. Troeger y Philip Manow. 2005. «Panel data analysis in comparative politics: Linking method to theory», *European Journal of Political Research*, 44: 327-354. Disponible en: <https://doi.org/10.1111/j.1475-6765.2005.00230.x>.
- Plutzer, Eric. 2002. «Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood», *The American Political Science Review*, 96 (1): 41-56. Disponible en: <https://doi.org/10.1017/S0003055402004227>.
- Powell, G. Bingham. 1986. «American Voter Turnout in Comparative Perspective», *The American Political Science Review*, 80 (1): 17-43. Disponible en: <https://doi.org/10.2307/1957082>.
- Prior, Markus. 2010. «You've Either Got It or You Don't? The Stability of Political Interest over the Life Cycle», *The Journal of Politics*, 72 (3): 747-766. Disponible en: <https://doi.org/10.1017/S0022381610000149>.
- Przeworski, Adam. 2007. «Is the Science of Comparative Politics Possible?», en Carles Boix y Susan C. Stokes (eds.), *The Oxford Handbook of Comparative Politics*. New York: Oxford University Press.
- Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub y Fernando Limongi. 2000. *Democracy and Development. Political Institutions and Well-Being in the World, 1950-1990*. New York: Cambridge University Press. Disponible en: <https://doi.org/10.1017/CBO9780511804946>.
- Robinson, William S. 1950. «Ecological Correlations and the Behavior of Individuals», *American Sociological Review*, 15 (3): 351-357. Disponible en: <https://doi.org/10.2307/2087176>.
- Roodman, David. 2009. «How to do xtabond2: An introduction to difference and system GMM in Stata», *The Stata Journal*, 9 (1): 86-136.
- Sátyro, Natália. 2013. «Institutional constraints, parties and political competition in Brazilian states», *Revista de Ciência Política*, 33 (3): 583-605.
- Schraufnagel, Scot y Barbara Sgouraki. 2005. «Voter Turnout in Central and South America», *The Latin Americanist*, 49: 39-69. Disponible en: <https://doi.org/10.1111/j.1557-203X.2005.tb00064.x>.
- Seijas Macías, J. Antonio. 2014. «Análisis del grado del 'Mal-apportionment' en los parlamentos autonómicos del Estado español», *Revista Española de Ciencia Política*, 34: 199-221.
- Smets, Kaat y Carolien van Ham. 2013. «The embarrassment of riches? A meta-analysis of individual-level research on voter turnout», *Electoral Studies*, 32: 344-359. Disponible en: <https://doi.org/10.1016/j.electstud.2012.12.006>.

- Steinbrecher, Markus y Harald Shoen. 2013. «Not all campaign panels are created equal: Exploring how the number and timing of panel waves affect findings concerning the time of voting decision», *Electoral Studies*, 32: 892-899. Disponible en: <https://doi.org/10.1016/j.electstud.2013.10.004>.
- Stimson, James A. 1985. «Regression in Space and Time: A Statistical Essay», *American Journal of Political Science*, 29 (4): 914-947. Disponible en: <https://doi.org/10.2307/2111187>.
- Stockemer, Daniel. 2017. «What Affects Voter Turnout? A Review Article/ Meta-Analysis of Aggregate Research», *Government and Opposition*, 52 (4): 698-722. Disponible en: <https://doi.org/10.1017/gov.2016.30>.
- Wansbeek, Tom. 2012. «On the remarkable success of the Arellano-Bond estimator», *AENORM*, 20 (77): 15-20.
- Wawro, Gregory. 2002. «Estimating Dynamic Panel Data Models in Political Science», *Political Analysis*, 10 (1): 25-48. Disponible en: <https://doi.org/10.1093/pan/10.1.25>.
- Western, Bruce y Simon Jackman. 1994. «Bayesian Inference for Comparative Research», *The American Political Science Review*, 88 (2): 412-423. Disponible en: <https://doi.org/10.2307/2944713>.
- Wilson, Sven E. y Daniel M. Butler. 2007. «A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications», *Political Analysis*, 15 (2): 101-123. Disponible en: <https://doi.org/10.1093/pan/15.2.101>.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: The MIT Press.

Presentado para evaluación: 12 de agosto de 2017.

Aceptado para publicación: 12 de febrero de 2018.

ADRIAN PIGNATARO

adrian.pignataro@gmail.com

Estudiante del doctorado en Ciencia Política de la Scuola Superiore Sant'Anna y la Università degli Studi di Siena. Docente de la Escuela de Ciencias Políticas de la Universidad de Costa Rica. Ha realizado investigaciones en las áreas de comportamiento electoral, opinión pública y métodos cuantitativos.