

# REPETIR CURSO: CONDICIONANTES DEL NIVEL PRIMARIO USANDO DATOS DE PANEL GEORREFERENCIADOS (\*)

Marcos HERRERA

CONICET-IELDE, Universidad Nacional de Salta (Argentina)

## Resumen

Utilizando un panel de datos de 220 escuelas primarias con información georreferenciada, este trabajo explora los determinantes de la tasa de repetición a nivel escuela, contexto y vecindario geográfico. Los resultados muestran que a nivel escuela únicamente la matrícula escolar es una variable significativa. Los factores contextuales como la tasa de desempleo y la proporción de mujeres en edad escolar se relacionan con la tasa de repetición primaria. Además, el desempleo y la proporción de mujeres de las áreas contiguas muestran efectos significativos. Estos resultados reflejan la importancia de incorporar la localización de las escuelas y su entorno social a la hora de establecer adecuadas políticas educativas que contemplen a la repetición escolar.

*Palabras clave:* tasa de repetición, datos espaciales de panel, dependencia espacio-temporal.

## Abstract

Using a panel data of 220 primary schools with georeferenced information, this research explores the determinants of repetition rate at school, contextual and geographical neighborhood levels. Results show that the matriculation is a significant variable at school level. Contextual factors such as unemployment rate and percentage of female youth of primary school age are related to the grade repetition rate in primary school. In addition, the unemployment and the women's proportion of the contiguous areas show significant effects. These results reflect the importance to consider the school localization and its social context if we want to establish suitable educational policies that address the grade repetition.

*Key words:* repetition rate, spatial panel data, spatial-temporal dependence.

*JEL classification:* C23, I21, I28

## I. INTRODUCCIÓN

La repetición o retención escolar es un tema que levanta acalorados debates en países de diferentes niveles de desarrollo (1). La repetición de curso es una práctica que ocurre cuando los estudiantes inician un nuevo año escolar en el mismo curso que el año previo (Brophy, 2006). El motivo principal de esta práctica es un bajo nivel académico, aunque también puede deberse a otras causas como absentismo, decisión de los padres, imposición de la escuela, entre otros. El objetivo básico de esta medida es establecer un parámetro de calidad educativa tal que todos los alumnos promocionados posean un nivel mínimo de conocimiento en el curso lectivo superior.

El impacto de repetir el año escolar ha sido extensamente investigado en diferentes niveles de agregación: a nivel individual, a nivel de escuela y en la sociedad. La literatura destaca que la repetición está asociada, significativamente, con el posterior abandono o fracaso escolar de dicho alumno (Allensworth, 2005; Ou y Reynolds, 2010). Grissom y Sheppard (1989) encuentran que el alumno repetidor tiene entre un 20 y 30 por 100 más de probabilidad de no completar la educación secundaria. Por otra parte, la gran mayoría de las investi-

gaciones reflejan un dato crucial: el nivel educativo del estudiante no mejora al repetir el año lectivo. Un metaanálisis realizado por Jimerson (2001) no encuentra evidencia de una mejora en el rendimiento académico posterior y, además, refleja que la repetición temprana es condicionante de numerosos factores de riesgo emocionales y de salud.

Existen otras consecuencias que involucran al funcionamiento del sistema educativo. Además del importante coste en términos monetarios que implica perder un año académico, la elevada repetición puede ocasionar un aumento en el tamaño de la clase dificultando la evaluación y supervisión efectiva del docente. Esto generará limitaciones en el uso de los recursos públicos y puede suceder que no permita el acceso a la educación de más niños (World Bank, 2012). Es decir, para la sociedad en su conjunto, es más eficiente que cada alumno sea promovido al siguiente año académico.

Las mencionadas consecuencias negativas son solo una cara de la moneda. En la otra cara, los defensores de la repetición sostienen que esta práctica genera importantes beneficios. Por un lado, los alumnos de bajo nivel educativo tienen otra oportunidad de adquirir el conocimiento necesario antes de exponerse a un nivel al que no están

preparados. Igualmente, la amenaza de repetir el curso empuja al estudiante a realizar un esfuerzo adicional en aprender los contenidos y los mejores alumnos ven recompensado su esmero. El resultado de estos beneficios se reflejará en una educación de excelencia en donde la promoción al siguiente curso es un resultado meritocrático.

En América Latina, y particularmente en Argentina, existe un importante debate sobre la conveniencia de la repetición y esta discusión se focaliza especialmente en los primeros años de la educación primaria. El debate se centra en brindar oportunidades educativas a los más desfavorecidos sin dejar de premiar el esfuerzo de aquellos que buscan progresar. El equilibrio entre estas dos posiciones evitará situaciones extremas como las investigadas por N'tchougan-Sonou (2001) para dos países africanos: Ghana reformó su sistema educativo hacia la promoción automática, mientras que Togo continuó una política selectiva mediante altas tasas de repetición. Ambos sistemas tienen sus problemas: i) Togo ha sido criticado por las escasas oportunidades educativas, con tasas muy bajas de escolaridad en donde el reducido número de graduados proviene de altos estratos sociales; ii) En Ghana la promoción automática trajo aparejado resultados decepcionantes en las evaluaciones de calidad educativa. Queda claro que una muy baja tasa de repetición es más bien un indicativo de baja calidad académica.

En este trabajo se llama la atención sobre la necesidad de reconocer cuáles son los factores que determinan una alta o baja tasa de repetición en vez de buscar un determinado nivel de la misma. Conocer con mayor detalle los condicionantes socioeconómicos de la tasa de repetición permitirá entender la complejidad del problema y también sus posibles soluciones.

Las causas de la repetición son numerosas y tienen en común el reconocimiento de factores múltiples a diferentes niveles o escalas de agregación. Dependiendo de las características de los datos disponibles, algunas investigaciones analizan la retención escolar a nivel individual relacionando tal evento con características propias del estudiante o de su entorno familiar (Byrd y Weitzman, 1994; Corman, 2003). Otros estudios incluyen variables a nivel de escuela como el tamaño de la clase y la razón estudiantes/profesor (Taniguchi, 2015). Algunas, aunque escasas, investigaciones incluyen factores regionales como en el trabajo de Bali *et al.* (2005) en donde se detecta que la retención

escolar es dirigida por características regionales locales, además de las características anteriormente mencionadas.

Nuestro trabajo reconoce la acción de factores múltiples a diferentes escalas de agregación. Sin embargo, nuestro punto de focalización está en las características de la escuela (en vez de factores específicos del estudiante). El objetivo principal es mostrar los condicionantes de la *tasa de repetición* incorporando factores de cada centro educativo así como de índole regional local, en forma similar al trabajo de Bali *et al.* (2005). Adicionalmente, se incluyen efectos espaciales a nivel regional que son omitidos por la gran mayoría de las investigaciones revisadas.

El trabajo continúa de la siguiente manera. En la segunda sección se muestra la evolución de la repetición en América Latina y la variabilidad intranacional de Argentina para contextualizar la etapa analizada entre los años 2008-2011 en la provincia de San Luis. En la tercera sección se detalla la base de datos y la estrategia empírica aplicada. Los resultados inferenciales son analizados en la sección cuarta. Por último, las conclusiones se desarrollan en la quinta sección.

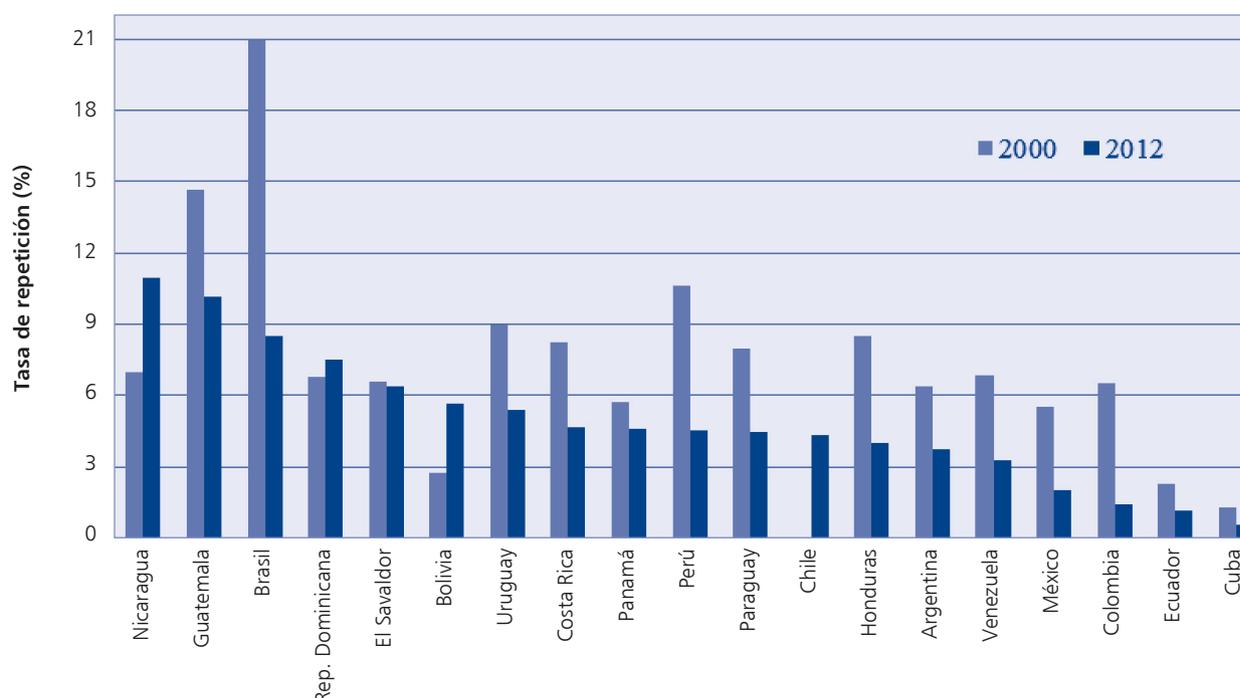
## II. EVOLUCIÓN DE LA TASA DE REPETICIÓN: UNA VISIÓN GENERAL

El gráfico 1 presenta las tasas de repetición promedio para los países de América Latina en los años 2000 y 2012. El nivel de las tasas para el año 2000 es notoriamente elevado, con algunos países superando el 12 por 100. A lo largo de la década puede observarse que la región ha mejorado el reporte de repetición, pasando de un promedio global de 7,25 por 100 en el año 2000 a un valor de 4,95 por 100 en el año 2012. En la mayoría de los países las tasas disminuyeron, aunque siguen observándose altos valores.

Otro resultado destacable del gráfico 1 es la disparidad de las tasas entre países. Por ejemplo, Brasil muestra una clara mejoría entre los dos años; sin embargo, la tasa del 2012 duplica la reportada por ocho países de la región (Chile, Honduras, Argentina, Venezuela, México, Colombia, Ecuador y Cuba).

Dentro de la región, Argentina ha sido históricamente uno de los países con mejores indicadores educativos. Por ejemplo, datos de la Unesco ubican a la tasa neta de escolarización primaria por encima del 93 por 100 para el período 1999-2005, niveles comparables a países de alto desarrollo. En el

GRÁFICO 1  
REPETICIÓN PRIMARIA EN AMÉRICA LATINA, AÑOS 2000 Y 2012



Fuente: Elaboración propia basada en datos del Instituto de Estadísticas de la Unesco (UIS).

caso de la repetición sucede algo similar, con una tendencia decreciente entre los años 2000 y 2012, pasando de un 6,4 por 100 a un 3,73 por 100. Estos resultados ubican al país entre los seis países latinoamericanos con mejor comportamiento. Sin embargo, estas tasas nacionales ocultan una gran heterogeneidad intranacional con valores muy diferentes entre jurisdicciones argentinas.

El cuadro n.º 1 es un ejemplo de la disparidad jurisdiccional argentina en los años 2009-2010. La información se encuentra ordenada de mayor a menor repetición para las 23 jurisdicciones provinciales más Ciudad Autónoma de Buenos Aires (CABA). Las diferencias entre jurisdicciones son claras, CABA muestra una repetición siete veces inferior al mayor valor observado, Corrientes. Incluso sin considerar ambos extremos, la situación es desigual.

Una explicación de las diferencias jurisdiccionales proviene, además de factores composicionales socioeconómicos, del ordenamiento territorial del

país. Argentina es un país federal en donde las provincias son una unidad elemental de organización del poder público y esto incluye al sector educativo. Leiras (2013) realiza un detallado análisis histórico sobre la capacidad autonómica de las provincias para establecer normativas en todo su territorio. En el plano educativo, desde la década de los noventa se vienen sucediendo reformulaciones en el funcionamiento general del sistema. A inicios de los noventa se finaliza el proceso de descentralización, formalizado por la Ley Federal de Educación 24.195 (LFE, 1993), en donde se confiere a las provincias competencias previamente nacionales y se reformula la estructura de la educación primaria y secundaria (Tedesco y Tenti Fanfani, 2001). La puesta en funcionamiento de la LFE generó numerosos problemas y fue finalmente derogada dando lugar a la Ley Nacional de Educación 26.206 (LEN), sancionada el año 2006 y actualmente vigente. Esta ley reorganizó los años educativos al formato previo a la LFE, creando, además, el Consejo Federal de Educación encargado de la concertación, acuerdo

y coordinación de la política educativa nacional, aunque sosteniendo la autonomía de las provincias dentro de su territorio. Esto posibilita la puesta en marcha de proyectos nacionales así como proyectos provinciales propios. También es posible que proyectos internacionales se lleven a cabo en algunas provincias, sin involucrar a la totalidad del país. Por ejemplo, el proyecto «Todos pueden aprender» (2), apoyado por Unicef, ha sido implementado desde el año 2003 en cinco provincias argentinas con el objetivo de disminuir la repetición, la sobriedad y la deserción escolar.

En el caso de la provincia de San Luis, se han impulsado diferentes proyectos que incluyen entre sus objetivos la disminución de la repetición y el abandono escolar. Entre estas propuestas, San Luis impulsó por primera vez la puesta en marcha de «Escuelas autogestionadas», en el año 2000, con la idea de satisfacer una demanda heterogénea y dar al sector privado la posibilidad de contribuir a la formación educativa mediante proyectos sociales. Más recientemente puede mencionarse el proyecto de «Estampillas escolares: ahorro para mi futuro», iniciado en el año 2011, en donde se recompensa monetariamente al alumno que finaliza cada curso lectivo mediante estampillas que al final del ciclo secundario pueden canjearse por dinero en efectivo.

CUADRO N.º 1

**PROMEDIO DE TASAS DE REPETICIÓN POR PROVINCIA. ESCUELAS PÚBLICAS Y PRIVADAS. NIVEL PRIMARIO. AÑOS 2009-2010**

PROVINCIA	REPETICIÓN	PROVINCIA	PROVINCIA
Corrientes	13,3	Catamarca	4,0
Formosa	8,9	San Juan	3,9
Stgo. del Estero	8,8	Tierra del Fuego	3,9
Misiones	7,7	La Pampa	3,9
San Luis	7,7	Santa Fe	3,8
Santa Cruz	7,0	Buenos Aires	3,7
Salta	6,7	Córdoba	3,4
Chaco	6,0	Tucumán	3,3
Entre Ríos	5,6	Chubut	3,2
La Rioja	5,3	Río Negro	3,2
Neuquén	4,9	Jujuy	3,1
Mendoza	4,8	Ciudad de Buenos Aires	1,9

Fuente: INDEC, en base a datos del Ministerio de Educación de la Nación y DiNIECE. Relevamientos anuales, 2009 y 2010.

GRÁFICO 2  
**EVOLUCIÓN DE LA REPETICIÓN PRIMARIA POR ÁMBITO. TOTAL DE ESCUELAS PRIMARIAS. PROVINCIA DE SAN LUIS. PERÍODO 2000-2011**



Fuente: Elaboración propia en base a informes provinciales del Ministerio de Educación de la Provincia de San Luis.

Estas y otras propuestas en el ámbito provincial generan particularidades en el sistema educativo que se reflejan en las diferentes medidas de eficiencia.

Las acciones conjuntas a nivel provincial y nacional han afectado la evolución de la repetición a lo largo del período 2000-2011. En el gráfico 2 se visualiza el descenso de la repetición primaria según el ámbito educativo: urbano, rural aglomerado y rural disperso. El gráfico muestra el promedio general de todas las escuelas sin distinguir el tipo de gestión, aunque son las escuelas públicas las que tienen mayor peso en términos de matriculación y también en donde se encuentran las mayores tasas de repetición.

La evolución temporal de las escuelas rurales presenta una clara mejoría a partir del año 2006. Las escuelas en zonas de campo abierto (rural disperso) incluso muestran tasas inferiores a las observadas en regiones urbanizadas. En el caso de las escuelas en localidades de menos de 2.000 habitantes (rural aglomerado), la repetición ha descendido año tras año desde el 2006, aunque sus niveles se mantienen por encima de dos puntos porcentuales respecto a los otros ámbitos.

En resumen, tanto a nivel de América Latina como nacional y provincial, la repetición ha sido un tema de especial atención y puede observarse el resultado de las políticas educativas. Lo llamativo del caso es la prácticamente nula investigación académica realizada sobre este tema. Entre los pocos trabajos recientes que aproximan el problema de la repetición puede mencionarse el aporte de Gómez Vera (2014) en donde se realiza un análisis comparativo entre cuatro países latinoamericanos. La información utilizada proviene del programa PISA (*Programme for International Student Assessment*), que es una evaluación de calidad educativa estandarizada de la OCDE (Organización para la Cooperación y el Desarrollo Económico) para alumnos de 15 años de edad. La evidencia empírica presentada por Gómez Vera destaca la condición de repetidor del alumno como una característica limitante en la capacidad lectora. En dicho trabajo Argentina es comparada a Uruguay, Brasil y Chile, determinando qué aspectos, tanto individuales como escolares, se ven afectados por la repetición, y cuál es exactamente su influencia sobre la calidad (3). Nuestro trabajo busca contribuir en este vacío académico mediante nueva evidencia proveniente de un panel de datos de escuelas primarias.

### III. DATOS Y ESTRATEGIA EMPÍRICA

La información disponible permite una amplia cobertura (prácticamente total) de las escuelas primarias de gestión pública para la provincia de San Luis. Como ventaja adicional, el conjunto de datos se encuentra georreferenciado y se describe a continuación.

#### 1. Base de datos

La base de datos contiene 246 escuelas públicas que impartían el nivel primario en el año 2012. Para cada una de estas escuelas, el Ministerio de Educación de la provincia de San Luis recolectó información sobre la tasa de repetición y tasa de sobreedad en un único registro anual cubriendo el período desde el año 2000 hasta el año 2011.

La información sobre cada establecimiento educativo proviene de una planilla anual cumplimentada por personal de cada escuela. La información es entregada al Ministerio a lo largo de cada año lectivo. Este formulario, denominado Planilla de Organización, contiene cuatro secciones generales: *Datos de establecimiento*, *Nivel inicial y primario*,

*Nivel secundario*, *Personal en otra repartición y Datos del personal*. Dicha estructura es única para todos los establecimientos educativos (públicos, privados o autogestionados) en todos los niveles, y especialidades, con una cobertura temporal para los años 2008, 2009, 2010 y 2011. Esto permitió construir un conjunto de variables que describen la composición del personal docente y no docente, la edad promedio de los docentes, cantidad de matriculados en cada año lectivo, cantidad de habitaciones de la escuela y cantidad de habitaciones consideradas aulas.

La ubicación geográfica de cada escuela proviene del sitio web *Mapa Educativo* (<http://www.mapaeducativo.edu.ar>) que releva información georreferenciada de cada unidad educativa cubriendo todo el territorio de la República Argentina. Cada escuela es identificada con un código único de establecimiento (CUE) y el sitio web muestra un legajo en donde pueden consultarse los datos institucionales, la evolución de la matrícula por nivel y grado, así como su ubicación geográfica (latitud y longitud).

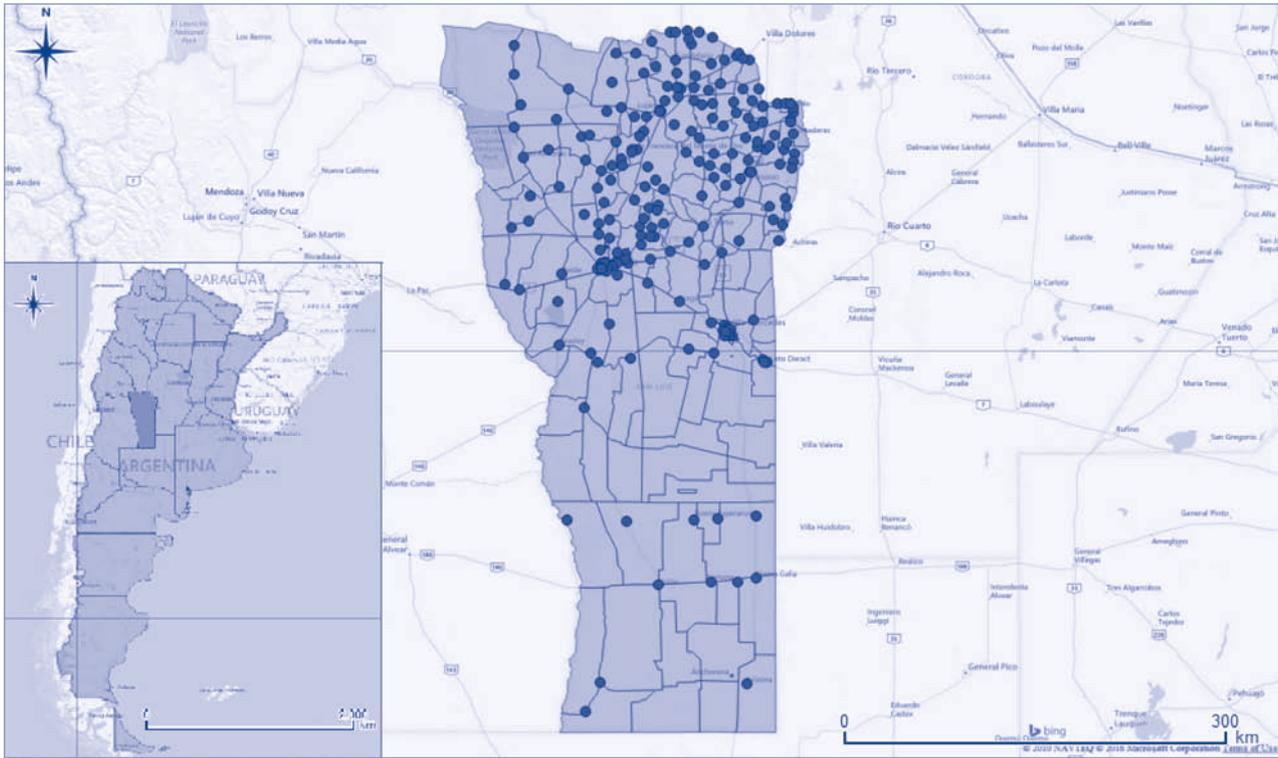
Una fuente de información adicional ha sido la información de la encuesta básica del Censo Nacional del año 2010 (4), disponible en la página web del Instituto Nacional de Estadísticas y Censos (INDEC). Los datos censales contienen información georreferenciada para un conjunto de variables socioeconómicas. La unidad de menor desagregación es denominada *radio censal* y posee una completa cobertura del territorio nacional. La definición y delimitación geográfica del *radio censal* depende de la cantidad de viviendas; en el caso de la provincia de San Luis existen 656 radios censales que contienen un promedio de 215 viviendas. En el gráfico 3 puede observarse conjuntamente los radios censales y la ubicación de las 246 escuelas primarias que forman parte de la base de datos original.

La conjunción de las diferentes bases de datos permite analizar la repetición y sus factores explicativos a nivel de escuelas primarias en un período temporal que abarca desde el año 2008 hasta el año 2011. Las escuelas primarias incluidas son únicamente de gestión pública, no considerando escuelas primarias de gestión privada o autogestionadas.

#### 2. Variables del estudio

Las 246 escuelas presentan información de repetición en al menos un año en el período 2008-

GRÁFICO 3  
DISTRIBUCIÓN GEOGRÁFICA DE LAS ESCUELAS PÚBLICAS CON NIVEL PRIMARIO



Nota: Representación de 246 escuelas primarias.

Fuente: Elaboración propia en base a información INDEC, Mapa educativo y Ministerio de Educación Provincial de San Luis.

2011, sin embargo, hay escuelas que no presentaron planillas de organización en todo el período y las mismas fueron eliminadas de la base final. Otro conjunto de 22 escuelas presentan planillas de organización en un único año y, siguiendo un análisis conservador, se optó por excluir de la base final a dichas instituciones (5). Con estas decisiones, la base de datos final incluye únicamente información de 220 escuelas públicas que complementaron al menos dos veces la planilla de organización. Esto implica un panel de datos desigual con 55 escuelas que presentan información en dos años, 75 escuelas que presentan información en tres años, y 90 escuelas que tienen información completa para los cuatro años del panel de datos. El panel posee brechas temporales (*gaps*) que son un tipo particular de desgaste muestral (*attrition bias*): hay escuelas sin información en los primeros años, algunas poseen datos faltantes en años intermedios y otras en el último período.

Las variables utilizadas pueden ordenarse según el grado de agregación considerando los niveles: escuela y radio censal. Las variables a nivel escuela son medidas en diferentes períodos temporales. Las variables contextuales son construidas únicamente con información censal de un único período y provienen de una encuesta básica. A continuación, se presentan las definiciones utilizadas para construir la variable dependiente y el conjunto de variables explicativas a utilizar:

**Nivel escuela:**

1. *Repetición*: variable dependiente medida como el porcentaje de alumnos repetidores respecto a la matrícula total de la escuela.
2. *Tamaño*: número total de matriculados en el año lectivo.

3. *Alum/prof*: porcentaje de la razón entre alumnos matriculados y número de profesores.

4. *Edadprof*: edad promedio de los profesores de la escuela, medida en años.

5. *Pct\_aula*: porcentaje de la cantidad de habitaciones utilizadas como aulas respecto al total de habitaciones de la escuela.

6. *Pct\_prof*: porcentaje de profesores respecto al total del personal vinculado a la escuela.

7. *Rural*: variable ficticia si la escuela pertenece a una zona rural aglomerada, es decir, en localidades con menos de dos mil habitantes.

#### **Nivel radio censal:**

8. *Desempleo*: relación porcentual entre las personas mayores que no están trabajando pero buscan trabajo activamente (ignora a los desalentados) en relación a la población adulta mayor económicamente activa.

9. *Hacina*: porcentaje de hogares en hacinamiento extremo en donde hay más de tres personas por habitación.

10. *Analf*: porcentaje de personas que no saben leer ni escribir respecto al total de la población en el radio censal.

11. *Pct\_fem014*: porcentaje de mujeres respecto al total de la población del grupo de edad de 0 a 14 años en el radio censal.

12. *Pct\_fem15+*: porcentaje de mujeres respecto al total de la población del grupo de edad de 15 o más años en el radio censal.

Las variables a nivel de radio censal fueron utilizadas para construir variables espaciales que capturan el valor promedio de los radios vecinos de desempleo, hacinamiento, analfabetismo e índices de feminidad:  $W \times Desempleo$ ,  $W \times Hacina$ ,  $W \times Analf$ ,  $W \times Pct\_fem014$ , y  $W \times Pct\_fem015+$ , donde  $W$  representa a la matriz de pesos espaciales.

La matriz  $W$  es una matriz de pesos espaciales que establece «quién es vecino de quién» para cada radio censal. Existen diferentes criterios de vecindad según el problema a analizar; en nuestro caso la construcción de esta matriz se realizó por criterio de contigüidad. Bajo este criterio se considera como

vecindario al conjunto de radios censales que comparten un límite común. Este criterio tiende a ser el más adecuado debido a que la educación pública se caracteriza, en términos generales, por estudiantes que provienen de los estratos sociales más desfavorecidos, con bajas posibilidades de elección y limitados a un entorno geográfico cercano (para mayor evidencia de estas características véase Narodowski y Nores 2000; Narodowski y Moschetti 2015). Bajo este criterio de vecindad, existe un promedio de seis vecinos para cada radio censal, implicando un vecindario promedio compuesto por 1.300 viviendas aproximadamente. En las principales zonas urbanas el vecindario se amplía a once o más radios vecinos y en zonas rurales la cantidad de radios vecinos disminuye, pero se incrementa la superficie de cobertura de dichos radios censales (véase el gráfico 3). El cuadro n.º 2 presenta las estadísticas descriptivas agregadas de las variables utilizadas en el análisis.

Las variables a nivel escuela fueron construidas siguiendo la revisión de la literatura. Por ejemplo, Taniguchi (2015) utiliza el número de matriculados como tamaño escolar; Duflo *et al.* (2015) utiliza la ratio alumnos/profesor y la experiencia medidas en años, entre numerosas investigaciones que utilizan dichas variables. Las variables «proporción de profesores respecto al total del personal docente» y «la proporción de aulas» han sido construidas para describir la heterogeneidad de las escuelas (algunas poseen una importante proporción de personal administrativo y de maestría, así como de habitaciones que no son utilizadas como aulas).

El factor contextual del hacinamiento ha sido estudiado como un elemento causal del rendimiento escolar, mostrando que aumenta la probabilidad de repetición (Goux y Maurin, 2005). Otros estudios cercanos que analizan el fracaso escolar han incorporado efectos contextuales similares a los aquí utilizados. De acuerdo a la Comisión Europea (2015) y a De Witte *et al.* (2013) el abandono temprano se relaciona con condiciones laborales desfavorables (desempleo), exclusión social y pobreza. En nuestro trabajo estos efectos contextuales intentan ser capturados por la tasa de desempleo y el analfabetismo promedio en el radio censal. El analfabetismo actúa como factor de bajo capital humano en la región comúnmente relacionado a pobreza y exclusión social. Respecto al porcentaje de mujeres, diversos estudios utilizan la ratio de género para caracterizar al grado escolar y al profesorado (Gomez-Neto y Hanushek, 1994; Duflo *et al.*, 2015), existiendo también investigaciones sobre el impacto diferencial de la presencia materna en el rendimiento

CUADRO N.º 2

## RESUMEN DE LAS VARIABLES. AÑOS COMBINADOS

VARIABLE	OBS.	MEDIA	S.D.	MIN.	MEDIANA	MAX.
<i>Repetición</i>	695	8,52	10,49	0,00	6,00	100,00
<i>Tamaño</i>	695	184,21	215,67	1,00	65,00	792,00
<i>Alum/prof</i>	695	13,07	5,56	1,00	13,00	31,00
<i>Edadprof</i>	695	44,16	5,30	25,00	44,00	65,00
<i>Pct_prof</i>	695	80,33	16,00	25,00	79,49	100,00
<i>Pct_aula</i>	695	57,09	20,54	10,00	58,33	100,00
<i>Rural</i>	695	0,12	0,32	0,00	1,00	1,00
<i>Desempleo</i>	695	4,60	3,50	0,00	4,50	13,98
<i>Hacina</i>	695	4,93	4,57	0,00	4,00	23,91
<i>Analf</i>	695	8,80	3,84	2,86	8,15	25,00
<i>Pct_fem014</i>	695	48,01	7,01	0,00	49,12	64,71
<i>Pct_fem15+</i>	695	46,49	6,14	26,79	47,57	60,48
<i>W×Desempleo</i>	695	4,40	2,69	0,00	4,26	13,19
<i>W×Hacina</i>	695	4,60	2,60	0,00	4,61	14,57
<i>W×Analf</i>	695	8,83	2,84	0,00	8,52	21,82
<i>W×Pct_fem014</i>	695	47,88	4,89	16,67	48,47	66,67
<i>W×Pct_fem15+</i>	695	45,25	6,00	24,42	45,24	57,62

Nota: La información descriptiva desagregada por año se presenta en el cuadro A del Apéndice.

Fuente: Elaboración propia según base de datos.

académico de los alumnos de primaria (Downey, 1994). Debido a la falta de información más desagregada, el trabajo introduce este índice de género a nivel contextual y espacial como variables proxies de los efectos de género enunciados.

### 3. Estrategia empírica de estimación

El diagrama del gráfico 4 muestra la estructura de datos a ser analizada. Pueden distinguirse tres niveles de anidación o jerárquicos: i) nivel 1: período temporal; ii) nivel 2: escuela primaria; y iii) nivel 3: radio censal. Para cada nivel hay variables que son de relevancia; la variable dependiente es medida al nivel 1; hay características de las escuelas del nivel 2; y todas las variables socioeconómicas provienen del nivel 3.

La estructura habitual de un panel de datos solo considera el nivel 1 y nivel 2, tal como puede observarse en el esquema del gráfico 4.

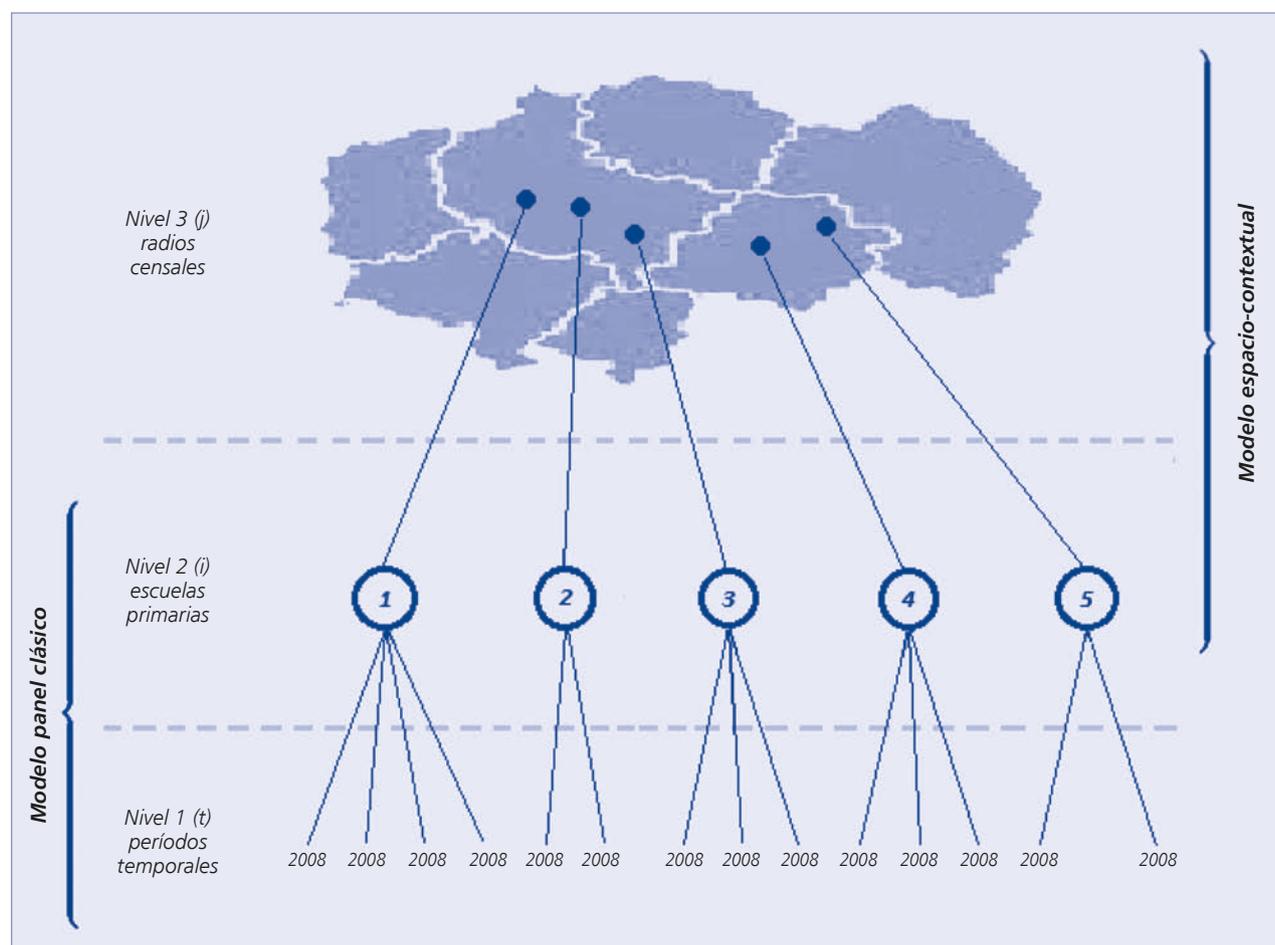
Debido a que no todas las escuelas son observadas en los cuatro períodos se tiene un panel de

datos desequilibrado: la escuela 1 posee información completa para todos los años considerados. Las escuelas 3 y 4 únicamente tienen información para tres años, pero no son los mismos para ambas escuelas. Las escuelas 2 y 5 muestran información de dos años con la particularidad que para la escuela 2 los años son consecutivos y para la escuela 5 la información disponible es para el primer y último período.

Además, en el gráfico 4 puede observarse un nivel 3 de agregación superior que es representado por los radios censales con información atemporal (único año censal, 2010). La información de este nivel fue utilizada para contextualizar el entorno de las escuelas.

Dada la estructura de los datos, es posible utilizar diferentes estrategias de especificación. Una posibilidad es imputar a las escuelas la información superior, proveniente de los radios censales, y trabajar el panel como un panel de datos espaciales con efectos de vecindario proveniente de la proximidad «entre escuelas». De esta forma, se trabajaría únicamente con escuelas utilizando la información

GRÁFICO 4  
ESTRUCTURA JERÁRQUICA Y ESPACIAL DE LOS DATOS



Fuente: Elaboración propia siguiendo la estructura jerárquica de la base de datos.

georreferenciada para incluir efectos espaciales. Un problema de esta estrategia es de tipo operativo ya que los paneles espaciales trabajan con datos balanceados y solo podrían usarse 90 escuelas que tienen información completa o deberían imputarse datos a 130 escuelas con los riesgos inferenciales que ello conllevaría.

Otra especificación alternativa es considerar un modelo jerárquico de dos niveles, escuelas y radios censales, agregando la información temporal como un promedio para cada escuela y trabajar los datos como si fuesen de un corte transversal único. Esta estrategia es identificada en el gráfico 4 como Modelo espacio-contextual, ya que pueden usarse los efectos contextuales de cada radio censal y los

efectos espaciales considerando los radios censales vecinos. Este tipo de especificación posee varias limitaciones, pero la más notoria es la pérdida de variabilidad temporal intraescuela.

Nuestra propuesta sigue la línea de investigación de modelos jerárquicos bajo estructuras complejas (Rabe-Hesketh y Skrondal, 2012, cap. 7-8). La estrategia adoptada permite utilizar toda la información disponible sin agregar información temporal o descartar escuelas.

La especificación inicial contiene tres niveles de agregación (tiempo, escuela y radio censal). A nivel superior se aprovecha la información socioeconómica y la distribución geográfica para construir las

variables del vecindario espacial. Además, se utiliza toda la información del panel desbalanceado aprovechando la información temporal para cada escuela.

Las variables medidas a nivel de radio censal proveen de los llamados efectos contextuales que afectan a las escuelas situadas en dicho radio censal. Por ejemplo, en el gráfico 4 se imputa un valor de desempleo del radio censal a las escuelas 1, 2 y 3; y otro valor de desempleo para las escuelas 4 y 5, suponiendo que la tasa cambia entre estos radios censales.

Además, se incorpora el efecto espacial a nivel de radio censal, que es el valor promedio de los radios censales vecinos tal que toda la información de los radios censales es utilizada para medir estos efectos, no solo de los radios que tienen escuelas. En el gráfico 4, la variable espacial  $W \times \text{Desempleo}$  para el radio censal con tres escuelas será construida como el valor promedio del desempleo de los radios limítrofes (incluirá el desempleo del radio vecino con 2 escuelas y el valor del desempleo de los radios censales limítrofes en donde no hay escuelas).

Una particularidad de estos determinantes contextuales y espaciales es que no poseen variabilidad intraescuela ocasionando que algunos métodos de estimación no puedan utilizarse para dar respuesta a la pregunta de investigación. En particular, no puede utilizarse la estimación por efectos fijos y esto decanta en el uso de métodos de estimación de efectos aleatorios. Una excelente discusión sobre las ventajas del modelos de efectos aleatorios puede revisarse en Bell y Jones (2015).

La especificación inicial incluyó tres niveles jerárquicos y se contrastó la significancia de cada uno de ellos comenzando desde el superior, radio censal. La descomposición ANOVA del modelo vacío, sin incluir variables explicativas, mostró componentes de varianza conjuntamente significativos a nivel de radio censal y de escuela. Sin embargo, al incluir variables explicativas a nivel escuela, la variabilidad entre radios censales se torna no significativa concluyendo que el efecto aleatorio a nivel 3, de radio censal, no es necesario en la especificación (los resultados pueden consultarse en el cuadro B del Apéndice). Con estos resultados, se simplifica la especificación hacia un modelo jerárquico de dos niveles siendo el superior la escuela y el inferior el tiempo (similar a un panel desbalanceado de efectos aleatorios). La información de los radios censales (contextual y espacial) puede ser imputada a cada escuela sin mayores inconvenientes.

#### IV. ANÁLISIS EMPÍRICO

La elección del modelo más adecuado fue realizada secuencialmente utilizando tres modelos alternativos de especificación. En primer lugar, se estimó un modelo con variables propias a cada escuela, modelo Escuela de efectos aleatorios:

$$y_{it} = x'_{it} \beta + u_i + \varepsilon_{it}, \quad i = 1, \dots, 220; \quad t = 2008, \dots, 2011, \\ u_i \sim N(0, \sigma_u^2), \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2), \quad [1]$$

donde  $y_{it}$  representa la tasa de repetición de la escuela  $i$ -ésima en el año  $t$ -ésimo, siendo  $x_{it}$  un vector de variables explicativas y  $\beta$  es un vector de parámetros desconocidos. El término de error está compuesto por  $u_i$ , efecto aleatorio para la escuela  $i$ -ésima, y una perturbación aleatoria,  $\varepsilon_{it}$ , representando el error de especificación bajo normalidad.

El segundo modelo incorporó variables contextuales: desempleo, hacinamiento, analfabetismo y participación femenina. Denominado como modelo Escuela-Contexto, su especificación es la siguiente:

$$y_{it} = x'_{it} \beta + Z'_j \theta + u_i + \varepsilon_{it}, \quad i = 1, \dots, 220; \quad t = 2008, \dots, 2011; \\ j = 1, \dots, 186, \quad u_i \sim N(0, \sigma_u^2), \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2), \quad [2]$$

donde  $Z_j$  representa a las variables contextuales obtenidas a nivel de radio censal. Este modelo solo utiliza información de los radios censales en donde hay al menos dos escuelas observadas en el mismo período o una escuela observada en al menos dos períodos temporales.

Finalmente, se incorporaron los efectos espaciales al modelo (2) obteniéndose el modelo más completo posible, modelo Escuela-Contexto-Espacial, con una estructura de especificación:

$$y_{it} = x'_{it} \beta + Z'_j \theta + WZ'_j \gamma + u_i + \varepsilon_{it}, \quad i = \\ = 1, \dots, 220; \quad t = 2008, \dots, 2011; \quad [3]$$

$$j = 1, \dots, 186, \quad u_i \sim N(0, \sigma_u^2), \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2).$$

Este último modelo incorpora variables espaciales utilizando información del vecindario parcialmente excluido del modelo Escuela-Contexto. Los resultados de la estimación de los tres modelos se presentan en el cuadro n.º 3.

Para cada modelo, además de los términos lineales se incluyeron efectos cuadráticos e interac-

CUADRO N.º 3

## RESULTADOS DE ESTIMACIÓN DE MODELOS ALTERNATIVOS

MODELO	ESCUELA		ESCUELA-CONTEXTUAL		ESCUELA-CONTEXTUAL-ESPACIAL	
Variables	Beta Est.	Test T	Beta Est.	Test T	Beta Est.	Test T
<b>nivel escuela</b>						
Tamaño	0,028***	3,66	0,021**	2,46	0,019**	2,03
Tamaño <sup>2</sup>	-0,000***	-3,72	-0,000***	-2,76	-0,000**	-2,48
Alum_prof	-0,110	-1,01	-0,120	-1,16	-0,112	-1,09
Edadprof	-0,074	-0,81	-0,100	-1,25	-0,059	-0,75
Pct_prof	-0,023	-0,50	-0,022	-0,50	-0,024	-0,55
Pct_aula	-0,011	-0,31	-0,015	-0,44	-0,009	-0,28
Rural	11,411	0,82	17,836	1,21	12,231	0,79
Rural×Tamaño	-0,126*	-1,71	-0,142	-1,50	-0,175*	-1,82
Rural×Tamaño <sup>2</sup>	0,000*	1,71	0,000	1,34	0,000*	1,69
Rural×Alum_prof	-0,201	-0,66	-0,087	-0,27	0,121	0,36
Rural×Edad_prof	0,107	0,32	0,018	0,05	0,083	0,23
Rural×Pct_prof	0,077	1,11	0,069	0,97	0,086	1,22
Rural×Pct_aula	-0,127	-0,88	-0,162	-1,07	-0,149	-1,03
<b>contextuales</b>						
Desempleo			1,402***	2,69	1,523***	2,87
Desempleo <sup>2</sup>			-0,080**	-2,10	-0,091**	-2,34
Hacina			0,612*	1,89	0,598*	1,90
Hacina <sup>2</sup>			-0,018	-1,19	-0,019	-1,29
Analf			-0,366	-0,60	-0,741	-1,30
Analf <sup>2</sup>			0,022	0,89	0,034	1,57
Pct_fem014			-0,494***	-2,74	-0,442**	-2,29
Pct_fem014 <sup>2</sup>			0,006***	2,58	0,005**	2,22
Pct_fem15+			-2,038	-1,15	-2,744	-1,54
Pct_fem15+ <sup>2</sup>			0,021	1,11	0,031	1,57
<b>espaciales</b>						
W×Desempleo					0,445*	1,90
W×Hacina					0,371	1,62
W×Analf					0,107	0,46
W×Pct_fem014					-0,069	-0,79
W×Pct_fem15+					-0,322**	-2,33
<b>comp. de varianza</b>						
$\hat{\sigma}_u^2$	19,458***		14,729***		12,034***	
$\hat{\sigma}_a^2$	86,355***		86,042***		86,133***	
<b>contrastes residuales</b>						
Coef. AR(1) resid.	0,185*		0,220*		0,214*	
z(l Moran global)	1,349**		0,768		0,506	
AIC	5216,521		5209,215		5204,789	

Nota: \*signif. al 10 %; \*\*signif. al 5 %; y \*\*\*signif. al 1 %. El test de Moran global se construyó usando una matriz inversa de distancia a nivel escuela con dimensiones diferentes para cada año, considerando los efectos aleatorios (r.e.) predichos por cada modelo. Los componentes de varianza fueron obtenidos previamente a corrección residual. Todas las estimaciones incluyen corrección robusta de los errores. AR(1): autorregresivo de orden 1.

ciones quedando seleccionados aquellos términos más adecuados según su significancia conjunta y utilizando el criterio de información de Akaike (AIC). Los efectos cuadráticos predominan en el caso de las variables contextuales. Bajo ninguna especificación se encontraron efectos significativos de las interacciones entre variables explicativas continuas.

Una vez obtenida la especificación polinómica más adecuada se realizaron chequeos sobre la estructura de los residuos para cada modelo. Según la tendencia temporal del gráfico 2, los últimos años muestran series más estables con baja persistencia. Pero dicho gráfico incluye a todas las escuelas primarias, no solo las públicas, y puede suceder que exista cierta correlación serial para el grupo de las escuelas que conforman el panel. Bajo el modelo simple de Escuela pudo utilizarse la estrategia propuesta por Wooldridge (2002) para contrastar la presencia de un proceso AR(1) en los residuos. El valor del contraste es no significativo al 5 por 100 y este resultado fue considerado como un primer indicio. Sin embargo, hay problemas relacionados al contraste: utiliza efectos fijos y por ende el resultado es el mismo cuando se aplica a los otros modelos, ya sea Escuela-Contexto o Escuela-Contexto-Espacial. Ante esta dificultad se optó por estimar el coeficiente autorregresivo residual considerándolo como un parámetro de molestia. El valor estimado del coeficiente y su significancia se presenta para cada modelo al final del cuadro n.º 3, *Contrastes residuales*.

Adicionalmente, se contrastó la presencia de autocorrelación espacial a nivel escuela. Para ello se construyó una matriz de contactos general con dimensión variable para cada año del panel. El criterio de vecindad elegido fue la inversa de distancia con un parámetro de alisado igual a 2 (6). El test de Moran univariante sobre la variable dependiente arrojó una dependencia espacial significativa; pero cuando se incorporan las variables condicionantes no se encuentra evidencia de tal dependencia. Estos resultados del test I de Moran se presentan al final del cuadro n.º 3, sección Contrastes residuales. Para cada uno de los modelos se muestra evidencia de no-autocorrelación espacial. En todos los casos se requirió la estimación de las predicciones de los efectos aleatorios para cada escuela.

Finalmente, se realizaron contrastes de Wald de significancia conjunta de las variables contextuales en el segundo modelo y de las variables espaciales en el tercer modelo, rechazando la hipótesis nula en ambos casos. La evidencia estadística permite establecer al modelo Escuela-Contexto-

Espacial como el más completo para explicar la tasa de repetición.

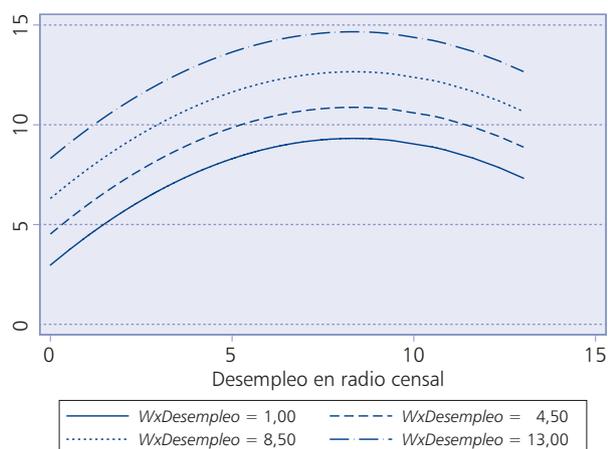
Los resultados muestran que los procesos contextuales y espaciales son factores condicionantes sobre la tasa de repetición escolar. Además, la significancia de variables espacio-contextuales se produce simultáneamente con efectos aleatorios importantes a nivel escuela. Estos resultados evidencian la presencia de procesos que actúan a diferente nivel de agregación y que deberían ser especificados explícitamente sin perder información en ningún nivel.

Respecto a los condicionantes, la única variable importante en términos estadísticos a nivel escuela es el tamaño, medido por el número de matriculados, siendo significativo en todos los modelos. Este resultado está en línea con evidencia similar encontrada para especificaciones más desagregadas en otros países (véase Barrow y Rouse, 2005). La no significancia del cociente entre alumnos y profesor ha sido reportada por diferentes estudios sobre variables educativas, por ejemplo el trabajo de Dearden *et al.* (2002) usando datos de escuelas públicas británicas.

En el modelo Escuela-Contexto, el desempleo del radio censal a la que pertenece la escuela juega un rol importante al explicar la repetición escolar agregada. Este resultado está alineado con los obtenidos por De Witte *et al.* (2013), donde el desempleo regional se relaciona con el fracaso escolar. El hacinamiento también aparece como una variable de relevancia, aunque solo es significativa al 10 por 100. La proporción de niñas en la población estudiantil muestra una relación negativa, este signo era esperado considerando que es una variable proxy de la mayor presencia femenina en las escuelas ubicadas en el radio censal. Esta relación es similar a la encontrada en estudios con otro nivel de desagregación como los presentados por Agasisti y Cordero (2015).

Los principales resultados del estudio provienen del tercer modelo que mantiene la importancia de los factores previamente mencionados e incorpora el impacto del vecindario. El modelo Escuela-Contexto-Espacial muestra que los efectos espaciales son importantes condicionantes sobre la tasa de repetición, en particular la proporción femenina adulta y en menor medida el desempleo. Para una mejor interpretación de los resultados de este modelo se presentan algunas gráficas de efectos marginales considerando el rango de variabilidad adecuada para cada factor explicativo.

GRÁFICO 5  
EFECTOS MARGINALES DEL DESEMPLEO SOBRE  
LA TASA DE REPETICIÓN

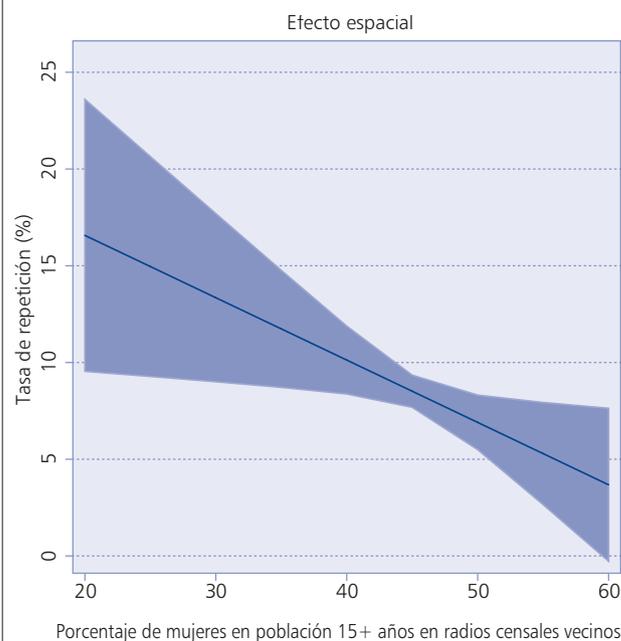
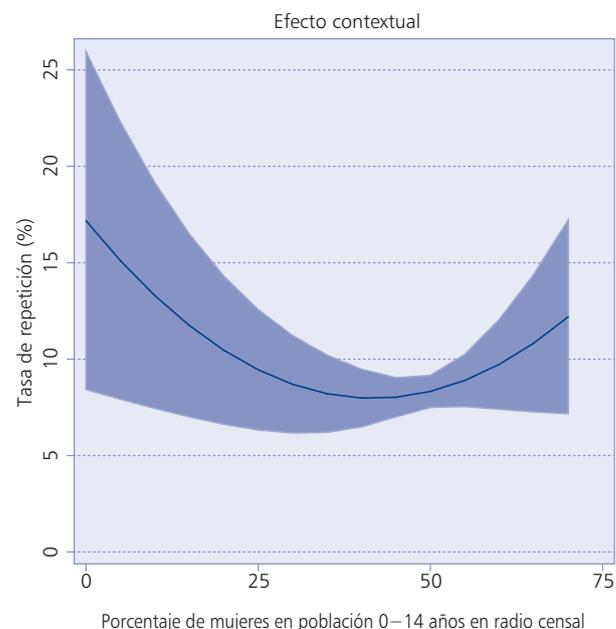


Fuente: Elaboración propia según estimación modelo Escuela-Contexto-Espacial.

En el gráfico 5 se muestra el efecto marginal de la tasa de desempleo sobre la repetición, *ceteris paribus*, para los percentiles del desempleo del radio censal entre el 5 por 100 y 95 por 100. Además se condiciona el efecto contextual por diferentes valores del desempleo del vecindario (efecto espacial). Los efectos no lineales son capturados a nivel contextual predominando el efecto lineal hasta alcanzar un desempleo del 8 por 100. Este impacto contextual es obtenido suponiendo que el vecindario no se ve afectado ante un cambio local del desempleo. Un análisis más detallado de los datos muestra que existe una correlación positiva significativa entre el desempleo local y el vecindario de 0,42 (correlación parcial de 0,24) implicando que un cambio en el desempleo local tiende a ser acompañado por un cambio de igual signo en el vecindario. De esta forma, una situación más adecuada es contemplar cambios simultáneos en el desempleo, contextual y espacial, que trasladan la curva de forma paralela tal como se bosqueja en el gráfico 5. A mayor desempleo espacial se espera un traslado adicional hacia arriba de la curva.

Otro condicionante de interés es el referido a la presencia de mujeres (contextuales y espaciales) representado en el gráfico 6. Los signos de estos efectos están en consonancia con los obtenidos por estudios que utilizan agregaciones diferentes (la literatura encuentra que la proporción de más

GRÁFICO 6  
EFECTOS MARGINALES DE LA PROPORCIÓN DE  
MUJERES SOBRE LA TASA DE REPETICIÓN



Fuente: Elaboración propia según estimación modelo Escuela-Contexto-Espacial.

niñas en la escuela o la presencia de las madres en la educación primaria disminuye la probabilidad de repetición). Un incremento de niñas en edad escolar en el radio censal disminuye la tasa de repetición

aunque el efecto no es lineal y la mejora es esperada cuando las niñas parten de una clara situación de minoría. Este efecto supone que la mayor presencia de mujeres en el radio censal se ve reflejado en una mayor presencia en la matrícula de las escuelas locales. Otro efecto agregado que se pudo detectar fue a nivel vecindario, donde una mayor cantidad de mujeres adultas condiciona la tasa de repetición a un menor valor. Estas variables no presentan una clara interacción entre sí, reflejado por correlaciones (total y parcial) no significativas para el conjunto de datos.

Por último, un resultado observable en el cuadro n.º 3 es el referido a la variabilidad del nivel escuela. La varianza entre escuelas ( $\sigma_{ij}^2$ ) se reduce a medida que el modelo propuesto se torna más complejo. El valor 19.485 del primer modelo se reduce a un valor de 14.729 para el modelo Escuela-Contexto y finalmente a 12.034 para el modelo Escuela-Contexto-Espacial. A pesar de su reducción, la significancia de este parámetro implica la presencia de procesos simultáneos a diferentes niveles que deben ser considerados explícitamente.

## V. CONCLUSIONES

La repetición es un problema en numerosos países y ha sido estudiada en diferentes niveles de agregación. Los estudios más habituales utilizan datos de alumnos mostrando que existen factores individuales importantes: falta de comprensión de texto, alta tasa de absentismo, escasa dedicación en el hogar, falta de materiales de estudio, entre otros. Farías *et al.* (2007) destaca que estos niños tienden a concentrarse en los hogares más pobres y por ello trabajar intensamente en la disminución de la repetición y el abandono escolar es equivalente a trabajar en la mejora de las oportunidades educativas de esos hogares.

Los resultados de esta investigación apoyan lo destacado por Farías *et al.* (2007) aunque en un nivel diferente al que generalmente se direccionan las políticas educativas. El estudio detecta que debe prestarse atención al contexto socio-espacial del lugar en donde se enseña. Las condiciones laborales desfavorables de la zona, así como el nivel de desempleo del vecindario que rodea a dicha zona afectan a la repetición observada en la escuela y los mismos deberían ser tomados en cuenta. Adicionalmente se encuentra evidencia de la importancia de la presencia femenina a nivel local y regional sobre la tasa de repetición en las escuelas.

La mayoría de las iniciativas educativas se focalizan en el alumno dejando de lado los factores aquí analizados. Los resultados de esta investigación plantean como desafío la realización de proyectos más integrales que consideren al área de influencia de la escuela y su posible impacto para el logro de los objetivos que involucran a la tasa de repetición.

El análisis desarrollado en esta investigación dista mucho de agotar completamente la complejidad del problema, aunque abre un conjunto de preguntas más generales que deberían ser consideradas en la agenda en investigación educativa:

- Basado en la comparación de investigaciones a nivel individual: *¿puede ser que los resultados presentados aquí sean efecto de la agregación a nivel escuela debido a la imposibilidad de contar con información a nivel alumno y familiar?*

- En base a los resultados obtenidos: *¿no será necesario incorporar información contextual y de localización a las encuestas que recopilan información del alumno para comprender más íntegramente el proceso de aprendizaje (ya sea repetición, calidad, etc.)?*

Ninguna de estas preguntas puede ser respondida de una forma adecuada sin una mejora en la información educativa. La primera pregunta, además, pone de relieve una limitación de los resultados presentados: no pueden generalizarse hacia el nivel individual/familiar. Nuestro estudio se focaliza sobre el comportamiento agregado a nivel escuela, y, por tanto, no implica que los resultados inferenciales sean válidos a nivel alumno.

La segunda pregunta es esencial, y un punto crítico reflejado a lo largo del trabajo, destacando la necesidad de contar con bases de datos más complejas que permitan analizar simultáneamente características individuales, familiares, escolares y contexto-espaciales. Prácticamente la totalidad de las encuestas educativas no incluyen información georreferenciada para contextualizar los resultados educativos y esto genera que los interrogantes de esta investigación no puedan profundizarse a menos que se procure una información integral del proceso educativo. Basta simplemente mencionar que muchas discusiones entabladas por especialistas y actores políticos consideran elecciones de políticas educativas como si el lugar y la geografía no fuesen de importancia. Realmente, *¿pueden entenderse los resultados educativos, como la tasa de repetición, sin considerar el contexto geográfico y social en donde se realiza la actividad?*

## NOTAS

(\*) El autor agradece los comentarios realizados por ENRIQUE LÓPEZ-BAZO en las Jornadas de discusión de Funcas: «Papeles de Economía Española: Redes de Interacción Social y Espacial: Aplicaciones a la economía española» (23-24 de junio, Madrid, 2016). Esta investigación ha sido parcialmente financiada mediante el proyecto de investigación del Ministerio de Economía y Competitividad de España (ECO2015-65758-P). E-mail: mherreragomez@gmail.com

(1) España no es una excepción a esta problemática tal como puede consultarse en una reciente entrada del blog *Nada es Gratis* (<http://nadaesgratis.es/admin/efectividad-de-la-repeticion-de-curso-en-el-sistema-educativo-espanol>).

(2) Enmarcado dentro del proyecto hemisférico «Elaboración de políticas y estrategias para la prevención del fracaso escolar» ([http://www.porlainclusion.educ.ar/proy\\_hemisferico/proy\\_hemisferico.htm](http://www.porlainclusion.educ.ar/proy_hemisferico/proy_hemisferico.htm)).

(3) Si bien la calidad educativa no es analizada en nuestro trabajo (debido a falta de información), claramente es un tema que se encuentra relacionado a la evolución de la repetición. Por ejemplo, Argentina participó en las ediciones PISA 2000, 2006, 2009 y 2012 con magros resultados en el *ranking* internacional. Los últimos resultados muestran que Argentina se encuentra entre los peores ocho países participantes y no ha logrado mejorar en ninguna asignatura luego de una década de participación (GANIMIAN, 2013).

(4) Respecto al Censo 2010, el INDEC presentó oficialmente la información a fines del año 2015. Como dato de relevancia, debe mencionarse la inclusión por primera vez de los mapas y códigos de radios y fracciones censales que permiten la georreferenciación.

(5) Para evaluar el impacto de estas decisiones sobre los resultados inferenciales finales se realizó un análisis de sesgo de selección. Los resultados obtenidos no muestran una pérdida de información significativa al utilizar una submuestra de 220 escuelas. La estimación de la segunda etapa de este análisis puede consultarse en el cuadro C del Apéndice.

(6) También se realizaron pruebas de robustez de la función de distancia utilizando un parámetro de alisado igual a 1. Los resultados no difieren de los presentados en el cuadro n.º 3.

## BIBLIOGRAFÍA

- AGASISTE, T., y J. M. CORDERO (2015), *The determinants of repetition rates in European countries: insights from an empirical analysis using PIRLS 2011 data*, Instituto Nacional de Evaluación Educativa-INEE.
- ALLENSWORTH, E. (2005), «Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion», *Educational Evaluation and Policy Analysis*, 27: 341-364.
- BALI, V.; ANAGNOSTOPOULOS, D., y R. ROBERTS (2005), «Toward a political explanation of grade retention», *Educational Evaluation and Policy Analysis*, 27: 133-155.
- BARROW, L., y C. ROUSE (2005), *Causality, causality, causality: the view of education inputs and outputs from economics*, WP-05-15, Federal Reserve Bank of Chicago.
- BELL, A., y K. JONES (2015), «Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data», *Political Science Research and Methods*, 3: 133-153.
- BROPHY, J. (2006), *Grade repetition* (Vol. 6). París, Bruselas: International Institute for Educational Planning, International Academy of Education and UNESCO.
- BYRD, R., y M. WEITZMAN (1994), «Predictors of early grade retention among children in the United States», *Pediatrics*, 93: 481-487.
- COMISION EUROPEA (2016), *Early school leaving*. Consultado de: [http://ec.europa.eu/education/policy/school/early-school-leavers\\_en.htm](http://ec.europa.eu/education/policy/school/early-school-leavers_en.htm)
- CORMAN, H. (2003), «The effects of state policies, individual characteristics, family characteristics and neighbourhood characteristics on grade repetition in the United States», *Economics of Education Review*, 22: 409-420.
- DE WITTE, K.; NICAISE, I.; LAVRIJSEN, J.; VAN LANDEGHEM, G.; LAMOTHE, C., y J. VAN DAMME (2013), «The impact of institutional context, education and labour market policies on early school leaving: a comparative analysis of EU countries», *European Journal of Education*, 48: 331-345.
- DEARDEN, L.; FERRI, J., y C. MEGUIR (2002), «The effect of school quality on educational attainment and wages», *Review of Economics and Statistics*, 84: 1-20.
- DOWNNEY, D. (1994), «The school performance of children from single-mother and single-father families: Economic or interpersonal deprivation?», *Journal of Family Issues*, 15: 129-147.
- DUFLO, E.; PASCALINE, D., y M. KREMER (2015), «School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenya primary schools», *Journal of Public Economics*, 123: 92-110.
- FARIAS, M.; FIOL, D.; KIT, I., y S. MELGAR (2007), *Todos Pueden Aprender. Propuestas para superar el fracaso escolar*. Fondo de las Naciones Unidas para la infancia y Asociación Civil Educación para Todos, Buenos Aires.
- GANIMIAN, A. (2013), *No logramos mejorar. Informe sobre el desempeño de Argentina en el Programa para la Evaluación Internacional de Alumnos (PISA) 2012*. Programa Educar 2050, Buenos Aires.
- GOMES-NETO, J., y E. HANUSHEK (1994), «Causes and consequences of grade repetition: Evidence from Brazil», *Economic Development and Cultural Change*, 43(1): 117-148.
- GÓMEZ VERA, G. (2013), «Los efectos de la repitencia en tanto que política pública en cuatro países del Cono Sur: Argentina, Brasil, Chile y Uruguay. Un análisis en base a PISA 2009», *Revista Latinoamericana de Educación Comparada*, 4, SAECE: 59-70.
- GOUX, D., y E. MAURIN, «The effect of overcrowded housing on children's performance at school», *Journal of Public Economics*, 89: 797-819.
- GRISSON, J., y L. SHEPPARD (1989), «Repeating and dropping out of school», en SHEPPARD, L. y SMITH, L. (eds.): *Flunking grades: Research and policies on retention*, Falmer, Londres, pp. 34-63.
- JIMERSON, S. (2001), «Meta-analysis of grade retention research: Implications for practice in the 21st century», *School Psychology Review*, 30: 420-437.
- LEIRAS, M. (2013), «Las contradicciones aparentes del federalismo argentino y sus consecuencias políticas y sociales», en ACUÑA, C. (ed): *¿Cuánto importan las instituciones?. Gobierno, Estado y actores en la política argentina*, Siglo XXI.
- NARADOWSKI, M., y M. MOSCHETTI (2015), «The growth of private education in Argentina: evidence and explanations», *Compare: A Journal of Comparative and International Education*, 45: 47-69.
- NARADOWSKI, M., y M. NORES (2000), «¿Quiénes quedan y quiénes salen? Características socioeconómicas en la composición de la matrícula en las escuelas públicas y en las escuelas privadas en la Argentina», *Documento de Trabajo n.º 31*, Fundación Gobierno y Sociedad, Buenos Aires.
- N'TCHOUGAN-SONOU, C. H. (2001), «Automatic promotion or large-scale repetition - which path to quality?», *International Journal of Educational Development*, 21: 149-162.
- OU, S., y A. REYNOLDS (2010), «Grade retention, post-secondary education and public aid receipt», *Educational Evaluation and Policy Analysis*, 32: 118-139.
- RABE-HESKETH, S., y A. SKRONDAL (2012), *Multilevel and longitudinal modeling using Stata*, vol. 1, 3rd ed., College Station: Stata Press.
- TANIGUCHI, K. (2015), «Determinants of grade repetition in primary school of sub-Saharan Africa: An event history analysis for rural Malawi», *International Journal of Educational Development*, 45: 98-111.
- TEDESCO, J., y E. TENTI FANTANI (2001), *La reforma educativa en la Argentina. Semejanzas y particularidades*, Ministerios de Educación de Argentina, Chile y Uruguay, Grupo Asesor de la Universidad de Stanford/BID, Buenos Aires.
- WOOLDRIDGE, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.
- WORLD BANK (2012), *World Development Indicator*, Development Data Centre, Washington D. C.

## APÉNDICE

CUADRO A

## RESUMEN DE LAS VARIABLES. NIVEL ESCUELA POR AÑO

	VARIABLE	MEDIA	STD	RANGO	P25	MEDIANA	P75
Año 2008	Repetición	8,04	8,81	36,00	0,00	7,00	13,00
	Tamaño	168,35	212,73	791,00	14,00	47,00	297,00
	Alum/prof	13,07	5,73	30,00	9,00	12,65	16,97
	Edadprof	43,44	5,79	37,00	40,00	43,00	46,00
	Prop_prof	81,52	16,00	68,57	72,73	80,00	100,00
	Prop_aula	56,01	21,26	90,00	40,48	56,41	70,00
	Rural	0,12	0,33	1,00	0,00	1,00	1,00
<b>Escuelas N = 215</b>							
	VARIABLE	MEDIA	STD	RANGO	P25	MEDIANA	P75
Año 2009	Repetición	9,87	10,20	83,00	0,00	9,00	15,00
	Tamaño	230,28	227,22	783,00	34,00	123,00	433,00
	Alum/prof	14,05	5,80	27,58	10,00	14,00	17,93
	Edadprof	43,90	4,58	32,00	42,00	44,00	46,00
	Prop_prof	78,55	14,56	75,00	70,45	78,57	84,63
	Prop_aula	60,99	18,93	85,71	50,00	61,90	72,73
	Rural	0,15	0,36	1,00	0,00	1,00	1,00
<b>Escuelas N = 149</b>							
	VARIABLE	MEDIA	STD	RANGO	P25	MEDIANA	P75
Año 2010	Repetición	8,43	12,51	100,00	0,00	5,00	12,00
	Tamaño	173,74	209,98	773,00	13,00	54,00	316,00
	Alum/prof	12,66	5,45	26,68	8,00	13,00	16,63
	Edadprof	44,25	5,05	32,00	41,00	44,00	46,00
	Prop_prof	80,33	16,70	75,00	70,59	80,00	100,00
	Prop_aula	55,69	20,60	90,00	40,00	55,56	70,00
	Rural	0,12	0,33	1,00	0,00	1,00	1,00
<b>Escuelas N = 189</b>							
	VARIABLE	MEDIA	STD	RANGO	P25	MEDIANA	P75
Año 2011	Repetición	7,95	10,16	50,00	0,00	6,00	12,00
	Tamaño	173,81	210,71	704,00	13,00	51,50	319,00
	Alum/prof	12,65	5,62	25,50	8,33	12,67	16,40
	Edadprof	45,39	5,39	31,00	41,00	45,00	47,00
	Prop_prof	80,38	16,49	75,00	70,00	79,13	100,00
	Prop_aula	56,51	20,71	88,89	40,00	59,79	70,00
	Rural	0,13	0,33	1,00	0,00	1,00	1,00
<b>Escuelas N = 142</b>							

Fuente: Elaboración propia según base de datos.

CUADRO B

## ESTIMACIÓN DE MODELOS ALTERNATIVOS CON TRES NIVELES

MODELO	VACÍO	ESCUELA	ESCUELA-CONTEXTUAL	ESCUELA-CONTEXTUAL-ESPACIAL
<b>Nivel Escuela</b>	No	Sí	Sí	Sí
<b>Contextuales</b>	No	No	Sí	Sí
<b>Espaciales</b>	No	No	No	Sí
<b>Comp. de Varianza</b>				
$\hat{\sigma}_v^2$ (radio censal)	13,471	5,498	0,284	0,000
$\hat{\sigma}_u^2$ (escuela)	17,360	13,942	14,632	12,219
$\hat{\sigma}_\Delta^2$ (residual)	89,121	86,326	86,317	86,189
<b>Contraste LR</b>				
$H_0: \sigma_v^2 = 0$	4,92 (0,03)	00,90 (0,34)	0,00 (0,96)	0,00 (1,00)

Nota: p-valores entre paréntesis.

Fuente: Elaboración propia según base de datos.

CUADRO C

## SESGO DE SELECCIÓN. RESULTADOS DE 2º ETAPA

MODELO	ESCUELA-CONTEXTUAL-ESPACIAL	
VARIABLES	beta est.	test t
Tamaño	0,019**	2,00
Tamaño <sup>2</sup>	-0,000**	-2,47
Alum_prof	-0,113	-1,10
Edadprof	-0,058	-0,74
Pct_prof	-0,023	-0,53
Pct_aula	-0,010	-0,30
Rural	12,749	0,83
Rural×Tamaño	-0,174*	-1,79
Rural×Tamaño <sup>2</sup>	0,000*	1,67
Rural×Alum_prof	0,113	0,33
Rural×Edadprof	0,075	0,21
Rural×Pct_prof	0,086	1,22
Rural×Pct_aula	-0,150	-1,03
Desempleo	1,522***	2,87
Desempleo <sup>2</sup>	-0,091**	-2,34
Hacina	0,598*	1,89
Hacina <sup>2</sup>	-0,019	-1,29
Analf	-0,775	-1,34
Analf <sup>2</sup>	0,036	1,60
Pct_fem014	-0,438**	-2,28
Pct_fem014 <sup>2</sup>	0,005**	2,20
Pct_fem15+	-2,741	-1,53
Pct_fem15+ <sup>2</sup>	0,031	1,57
W×Desempleo	0,466*	2,00
W×Hacina	0,334	1,47
W×Analf	0,138	0,59
W×Pct_fem014	-0,073	-0,84
W×Pct_fem15+	-0,314**	-2,27
$\hat{\lambda}$ (inversa de la ratio de Mills)	<b>0,657</b>	<b>0,46</b>

Notas: \* signif. al 10 %; \*\* signif. al 5 %; y \*\*\* signif. al 1 %.