

EFFECTOS DE LA INTERACCIÓN SOCIAL EN LAS DECISIONES SOBRE EDUCACIÓN DE LOS ADOLESCENTES EN COSTA RICA

Luis J. Hall¹

ÍNDICE

1. Introducción.....	324
2. Marco Político y Educacional.....	325
3. Formas de Interacción Social.....	325
4. Modelo e Implementación Empírica.....	327
4.1. Modelo sin Mercado.....	327
4.2. Estrategia de Estimación y Resultados Econométricos.....	328
4.2.1. Análisis de Componentes de Varianza.....	329
4.2.2. Análisis de Regresión.....	331
5. Conclusión.....	333
6. Bibliografía.....	334

RESUMEN

El objetivo principal de este trabajo es evaluar el grado de interacción social en las decisiones de escolaridad de los adolescentes en Costa Rica. El artículo muestra el alto grado de interacción social en este tipo de decisiones en los datos de Costa Rica. Los resultados obtenidos son robustos en la inclusión de una serie de variables de control que capturan un conjunto de factores no observables que podrían confundirse con las interacciones sociales. En particular, éstas variables de control incluyen características de las familias, de los colegios y características geográficas.

PALABRAS CLAVE: EDUCACIÓN, INTERACCIÓN SOCIAL, ECONOMETRÍA

1 El autor es profesor de la Escuela de Economía de la Universidad de Costa Rica. Correspondencia a luis.hall@ucr.ac.cr.

ABSTRACT

The main goal of this paper is to evaluate the degree of social interaction in teenagers' decision choice of schooling in Costa Rica. The paper supports a high degree of social interaction for this schooling choice. The results hold upon the inclusion of a series of variables to control for sources of non observable variables in the model. In particular, special attention has been paid to sources of variability at the family, school and geographical level across districts.

KEY WORDS: EDUCATION, SOCIAL INTERACTION, ECONOMETRICS

1. INTRODUCCIÓN

En las decisiones relacionadas con su educación, los adolescentes no solamente toman en consideración sus características personales o las de sus familias, sino también las decisiones de educación de otros adolescentes cercanos a ellos. La forma en que estas interacciones sociales se llevan cabo puede ser muy diversas. Por ejemplo, éstas pueden ser directas, en las cuales los adolescentes se ayudan entre ellos a aprender en la escuela; o podrían ser indirectas en las cuales los adolescentes responden a las normas y estigmas desarrolladas por los miembros del grupo. La forma misma que adopta ésta interacción es sumamente importante, no solo porque delimita la forma en que se podrían estimar estos efectos, sino también porque delinea las implicaciones de políticas económica derivadas de las mismas.

El objetivo principal de este trabajo es estimar el grado de interacción social entre los adolescentes costarricenses al llevar a cabo sus decisiones de educación, es decir, en qué medida las decisiones de educación de los vecinos afectan la decisión individual de educación. El documento provee evidencia que apoya estos efectos sociales y complementa aquellos estudios de educación que muestran cómo los factores individuales, familiares y las características de las escuelas son muy relevantes en la decisión de educación.

En el documento se utilizan dos metodologías complementarias de estimación de las interacciones sociales. La primera estrategia utiliza la relación entre la varianza agregada y la varianza individual en las decisiones de educación. En la segunda estrategia se utiliza la regresión lineal

utilizando el método de variables instrumentales. Esta comparación permite evaluar explícitamente las fortalezas de ambas metodologías y ver las implicaciones de sus conclusiones. Para capturar los efectos de interacción local, los datos permiten conocer el distrito donde viven los adolescentes así como su escolaridad alcanzada. Sin embargo, se desconoce la forma explícita de la interacción en este tipo de datos. Se captura por tanto interacciones que ocurren dentro de estos vecindarios definidos como distritos. Además, y dado el reducido número de colegios dentro del distrito, la interacción está altamente afectada por la dinámica que se da en estas instituciones.

En esta investigación se incluye un conjunto de variables de control que no han sido anteriormente utilizados en este tipo de estudios. Particularmente, se introduce un conjunto de variables geográficas que controlan la variabilidad geográfica y climática de las diferentes regiones. Adicionalmente, en base a la literatura empírica sobre educación, se utilizan una serie de variables a nivel de las escuelas que permite controlar la variabilidad en la calidad de las escuelas, en particular, el desempeño promedio de las escuelas, la calidad de sus educadores y de las juntas de padres que las representan. El rol de la familia es muy importante y en particular, el nivel de educación del jefe de la familia (ver Ashenfelter y Zimmerman, 1997) nos permite identificar las interacciones sociales en el ejercicio de variables instrumentales.

Este trabajo además, abre algunas interrogantes para futuras investigaciones. La literatura teórica muestra que las interacciones sociales se pueden obtener en distintos niveles, por ejemplo, los individuos podrían

estar influenciados por interacciones propias de su localidad y por las interacciones globales. En este estudio, específicamente, se sugiere la importancia de analizar el problema de deserción, es decir, la decisión de asistir o no a la escuela, dentro de un esquema de diferentes niveles de interacción.

El artículo está organizado de acuerdo al siguiente esquema: la primera parte describe los datos empleados y caracteriza las formas de interacción social potencial. La segunda parte describe en detalle el marco institucional del estudio. En particular, se da una descripción del sistema educativo. La tercera parte presenta un modelo de optimización de las interacciones sociales que sirven de marco para las implementaciones empíricas de la siguiente sección. La última parte presenta los resultados empíricos del estudio y finalmente las conclusiones.

2. MARCO POLÍTICO Y EDUCACIONAL

Costa Rica es una país democrático dividido en 7 provincias, 81 cantones y 443 distritos. Los distritos forman parte de los cantones, y los cantones son parte de las provincias. El presidente es electo a nivel nacional por mayoría simple; los diputados son electos a nivel provincial y los gobiernos locales son elegidos a nivel de cantón.

CUADRO 1
DESCRIPCIÓN DE LOS NIVELES DE EDUCACIÓN
EN COSTA RICA

Nombres	Niveles de Educación	Grados
Pre-escolar		
Primaria	I ciclo	1-3
	II ciclo	4-6
Secundaria	III ciclo	7-9
	IV ciclo	10-11(12)

El sistema político democrático en Costa Rica ha sido estable desde 1948 y es muy transparente. Costa Rica mantiene altos estándares de desarrollo social y humano y está ubicado en el lugar 31 a nivel mundial de acuerdo con la UNESCO (UNESCO, 2002) y 6to a nivel

latinoamericano. De acuerdo con sus estándares sociales, Costa Rica está calificada como un país subdesarrollado con un ingreso per cápita de \$8,213 por año en el 2002.

La educación primaria va de primero a sexto grado y está dividida en primero y segundo ciclo. La educación secundaria va de séptimo a undécimo año y conforman el tercero y cuarto ciclo. Los estudiantes de la educación secundaria tienen la opción además de escoger un colegio técnico vocacional en vez de las alternativas académicas. Los colegios vocacionales van de decimo a duodécimo año. Es mandatorio ante la ley asistir a la escuela primaria y secundaria aunque no es una medida que necesariamente se hace cumplir. Los niños inician la escuela primaria a la edad de 6 años y 6 meses² y de esta manera los estudiantes inician la educación secundaria con 12 años y 6 meses cumplidos.

La tasa de alfabetización en los adultos en Costa Rica fue del 95.8% en promedio período 2001-2004. Este alto estándar contrasta con la alta tasa de deserción escolar que alcanzó un 16.5% en 1987. Después de implementar una serie de medidas correctivas para este problema, esta tasa fue reducida a un 11.9% en el año 2000. Este número es aún muy alto en comparación con otros países. Por ejemplo, en el caso de los Estados Unidos, esta tasa es de un 4.5% en el 2002 (UNESCO 2006, US Census Bureau (2002)).

3. FORMAS DE INTERACCIÓN SOCIAL

El objetivo de este trabajo es estudiar las decisiones de educación de los adolescentes en Costa Rica y en particular, en qué medida la interacción social entre los adolescentes afecta su decisión individual. Por decisión de educación se entiende la elección del número de años de educación alcanzados por los adolescentes. Es importante estudiar las formas potenciales de interacción social que se podrían presentar. Para este documento y dada la naturaleza de la información utilizada, las escuelas se vuelven

2 La ley anterior establecía 6 años y 3 meses cumplidos.

un medio primordial donde ocurren las interacciones. Se va a describir primero los individuos bajo estudio. Posteriormente se discutirán las formas de interacción social que podrían presentarse.

La muestra se obtuvo de la Encuesta de Hogares del año 2003. La edad y la localización de los agentes determinan los grupos de los individuos estudiados. Se consideran adolescentes aquellos individuos que aparecen clasificados en la encuesta como hijos o hijas dentro de la familia, cuya edad oscila entre los trece y los diecisiete años. El vecindario está definido por los distritos donde ellos viven al momento de la encuesta³. Como medida de escolaridad se utiliza el número de años que el individuo habría completado al momento de la entrevista.

Los datos muestran que de los 4377 adolescentes incluidos en el estudio, el 54% asisten a colegios públicos, el 7% asisten a colegios privados, el 9% asisten aun a escuelas primarias y el 26% restante no asiste a la escuela⁴ (ver cuadro 2). En el Cuadro 3 se presenta el nivel de educación de la muestra de acuerdo con la edad y el grado escolar alcanzado. El promedio de años alcanzado por los adolescentes en la muestra es de 6.77 años y por cada año adicional de incremento en la edad, su educación en promedio aumenta en 0.5 años. Adicionalmente, se observa la alta tasa de deserción de la educación secundaria, como antes se había mencionado. De un 91% de participación en las aulas a la edad de 13 años, esta se reduce a un 59% a la edad de 17 años. Existe por lo tanto, una alta tasa de deserción al momento de iniciar la educación secundaria. Note como el grupo con sexto grado completo se mantiene

una importancia relativa y estable conforme el grupo como un todo aumenta en edad.

CUADRO 2
ADOLESCENTES SEGÚN DECISIÓN DE TRABAJO
Y DE EDUCACIÓN

Condición	Trabaja	No Trabaja	Total
Estudia	207	3052	3259
No estudia	654	464	1118
Total	861	3516	4377

Datos tomados de la Encuesta de Hogares 2003.

La educación secundaria es mandatoria por ley aunque no se hace cumplir, como se mencionó anteriormente. Los estudiantes deben ir a las escuelas públicas que están dentro de su distrito, si hay cupo disponible. De otra forma, deben ir a la más cercana de acuerdo con su residencia⁵. Las escuelas privadas no tienen este tipo de restricción. El tamaño promedio de un distrito es de 115.4 kilómetros cuadrados, con una mediana de 33.8 kilómetros cuadrados. De los 443 distritos que han en el país, 262 tienen al menos un colegio de educación secundaria y 162 tienen exactamente un colegio público. Hay 616 colegios públicos en el país y 427 son públicos. El número máximo de colegios públicos por distrito es de 7 instituciones.

El distrito como una unidad geográfica de análisis de este estudio resulta ser un escenario ideal en la medida en que los adolescentes deben ir a los colegios públicos de su distrito y como se mencionó anteriormente, un 40% de los distritos tienen únicamente un colegio. Dada la forma en la que se provee la educación, una fuente importante de variabilidad no observable en los datos se puede encontrar a nivel de las escuelas. En este sentido, la calidad de los maestros, la administración de la institución o aun la organización de juntas de padres de familia podrían ser factores muy importantes que explican el mayor o menor nivel de escolaridad de los estudiantes

3 Estos adolescentes deberían estar iniciando el séptimo año de su educación, es decir, el primer año de educación secundaria. Hay 4377 adolescentes en la muestra en 3240 hogares. Se consideran únicamente los hijos e hijas debido a que de otro modo no se tendría a disposición la educación de los padres. Si se incluyeran todos los adolescentes, se tendrían 4905 en la muestra.

4 La encuesta no especifica si el estudiante asiste a escuela pública o privada, la estimación mostrada fue obtenida de los reportes de matrícula por centro educativo.

5 Se han presentado recursos legales que buscan flexibilizar esta restricción geográfica de asistencia para las escuelas públicas. Sin embargo, la Sala Constitucional ha rechazado varios recursos de amparo en este sentido.

en estas localidades. De esta manera, controlar por factores propios de los colegios públicos es muy importante dada la enorme vinculación de estas instituciones a las localidades⁶.

Se consideraran tres tipos de variables que nos permitan controlar por estos efectos no observables. En primer lugar, se incluye un conjunto de variables que permiten con-

6 Para esta muestra, se desconoce el lugar de donde proceden los estudiantes de los colegios públicos, por lo que no es posible determinar cuáles estudiantes viven en distritos donde no hay colegios públicos.

trolar por la variabilidad en la calidad de los maestros entre los colegios. En segundo lugar, se controla también por los factores de representación de los padres en los colegios que podría explicar las diferencias observadas en la educación de sus hijos. Por ejemplo, parte del financiamiento de los colegios en Costa Rica se debe a las actividades de recaudación de fondos de las asociaciones de padres. En el estudio se incluyen algunas variables que permitan capturar estas diferencias. Finalmente, se incluyen algunas variables de desempeño promedio de los colegios para controlar por algunos otros factores no observables a nivel de los colegios.

CUADRO 3
AÑOS DE ESCOLARIDAD SEGÚN GRUPOS DE EDAD EN COSTA RICA

Años de Escolaridad	Edad					Total
	13	14	15	16	17	
0	6	7	12	6	8	39
1	6	2	5	1	4	18
2	14	4	7	11	10	46
3	29	23	18	12	20	102
4	51	32	19	16	26	144
5	136	75	42	28	27	308
6	472	322	259	216	243	1512
7	159	310	157	110	74	810
8	5	121	225	121	106	578
9	1	8	116	196	117	438
10	1		8	87	182	277
11			1	3	75	80
12					1	1
Obs. Incompletas	4	5	5	1	9	24
Total Adolescentes	884	909	874	808	902	4377
Promedio (años)	5.69	6.33	6.84	7.39	7.67	6.77
Asiste Colegio (%)	91%	81%	73%	68%	59%	74%

Datos tomados de la Encuesta de Hogares 2003.

4. MODELO E IMPLEMENTACIÓN EMPÍRICA

4.1. Modelo sin Mercado

Es útil describir el modelo de interacción social explícitamente de forma tal que se

puedan formalizar nuestros resultados, trabajar otros factores de interés, y especificar la estrategia de identificación y estimación a ser implementada. Con este propósito se utiliza el modelo propuesto por Glaeser y Scheinkman (2001, 2002) para modelar la decisión individual y la estrategia de estimación e identificación.

Supongamos que el índice $i = 1, \dots, n$ representa a cada adolescente en nuestra muestra y asíociémele A_i a cada adolescente i . La variable A_i representa entonces la acción promedio de todos los otros agentes en mi vecindario, es decir el promedio de años de educación del distrito particular. Cada agente está sujeto a un choque en las preferencias Θ_i , una variable aleatoria con soporte en Θ que nos permite introducir posteriormente características observables del individuo tales como el ingreso de la familia, la estructura familiar, o inclusive características observables del distrito. Esto induce el tipo de agente dentro de la economía. Cada adolescente para nuestro caso escogerá una acción $a_i \in \Lambda$. Para nuestro estudio posterior se tomara la variable a_i como el número de años de educación alcanzado por el adolescente. La utilidad de los adolescentes i está en función tanto de la acción del agente mismo, a_i , como de la acción escogida promedio por todos los agentes de su vecindario A_i y por su choque de preferencias Θ_i . Formalmente decimos que,

$$U^i = U^i(a_i, A_i, \Theta_i) \quad (1)$$

Donde,

$$A_i = \sum_j^{n_j} \gamma_{ij} a_j \quad (2)$$

Con $r_{ij} \geq 0$, si $j \in A_i$, $\sum_j^k \gamma_{ij}^k = 1$. El índice n_j representa el número de individuos en el vecindario f , el distrito particular, y se asume el mismo entre los vecindarios para simplicidad de la notación, es decir, $n_j = n$.

Un equilibrio se define de la siguiente forma; para vectores dados $\Theta = (\Theta_1, \dots, \Theta_n) \in \Theta^n$ un equilibrio Θ es un vector tal que, para cada i ,

$$a_i \in \operatorname{argmax} \{U^i(a_i, \sum_{j=1}^n \gamma_{ij} a_j, \Theta_i)\}. \quad (3)$$

Para que este equilibrio exista, el agente i debe observar a A_i , el estadístico resumen de las acciones de los otros individuos que afectan su utilidad en el sentido del equilibrio de Nash.

Scheinkman and Glaeser (2002) prueban que tal equilibrio existe bajo un escenario muy general. Es posible mostrar la unicidad de tal equilibrio después de introducir condiciones más fuertes. La cuestión de la multiplicidad de los equilibrios se vuelve un problema más significativo cuando se trabaja con un modelo

de interacción global en vez de uno local, como es el caso de este trabajo. Para los propósitos de esta investigación nos concentramos en los efectos locales y por lo tanto la unicidad del equilibrio está garantizada.

Para obtener nuestra ecuación de estimación, se restringe más la función de utilidad y se introducen preferencias cuadráticas (Glaeser and Scheinkman, 2001) como sigue:

$$U^i = U^i(a_i, A_i, \Theta_i) = \Theta_i a_i - \frac{1-\alpha}{2} a_i^2 - \frac{\alpha}{2} (a_i - A_i)^2. \quad (4)$$

Aquí $0 \leq \alpha \leq 1$ mide la preferencia por conformidad. Para este caso, la solución óptima del modelo está dada por

$$a_i = [\alpha A_i + \Theta_i] \quad (5)$$

Esta es la ecuación que nosotros evaluaremos en la siguiente sección. La solución óptima de los agentes, es decir de los adolescentes, responde por lo tanto a la decisión de educación promedio de los adolescentes en el vecindario, a los choques en los gustos y al parámetro.

4.2. Estrategia de Estimación y Resultados Econométricos

El estudio sigue dos estrategias complementarias para estimar los efectos de interacción local. El primero es el análisis de componentes de la varianza, el cual se enfoca en las desviaciones de la varianza agregada de la varianza individual para estimar de esta forma el grado de interacción social entre los agentes. Una de las ventajas más importantes de este método de identificación es que se evita el problema de simultaneidad, conocido también como el problema de reflexión (vea Manski (1993)) y que estaría presente si seguimos una estrategia de identificación ecuación por ecuación.

El segundo método empleado es el análisis de regresión en donde las decisiones individuales son regresadas a los promedios de las decisiones de educación del grupo de individuos del vecindario. Un problema crucial presente

en esta estimación es el problema de identificación causada por la simultaneidad introducida, como se describió anteriormente. Con el fin de corregir este problema, se sigue en esta estrategia de identificación el método de variable instrumental. Ambas estrategias presentan aun el problema latente de autoselección en los vecindarios por parte de los agentes y algunos otros problemas de variables no observables. Estos problemas potenciales en los datos son discutidos posteriormente.

4.2.1. Análisis de Componentes de la Varianza

Scheinkman y Glaeser (1996, 2001) proponen una metodología que estima las interacciones sociales a partir de las diferencias entre la varianza agregada y la varianza individual entre los vecindarios.

Para establecer esta estrategia, supongamos que tomamos el modelo de la utilidad cuadrática discutido en la sección anterior en la ecuación (5). Adicionalmente, para incluir las características observables de los individuos, definamos $\Theta_i = \theta_i + f(X_i)$ donde θ_i tiene media cero y varianza σ_θ^2 y X_i representa el conjunto de variables observables del individuo. Estas variables observables también incluyen algunas características a nivel del distrito.

El modelo de decisión es entonces

$$a_i = [\alpha A_i + \theta_i + f(X_i)]. \quad (6)$$

Si escribimos $\overline{f(X)}$ para representar el nivel promedio de $f(\cdot)$, $\overline{f(X)}_j$ para el nivel medio de la función $f(\cdot)$ en el distrito j , y \bar{a}_j para la acción o decisión promedio de los agentes en el distrito f entonces a acción individual cuando sustraemos la acción promedio del distrito o vecindario y obtenemos

$$a_i - \bar{a}_j = \theta_i + f(X_i) - \overline{f(X)}_j + \alpha(A_i - \bar{a}_j). \quad (7)$$

Si condicionamos en el distrito j , $f(X_i)$ es entonces independiente de A_i debido al supuesto de que no existe autoselección. De esta manera, si se escribe $Var_j^{f(X)}$ como la varianza de las

características $f(X)$ en el distrito f , se puede mostrar que

$$Var_i = Var(a_i - \bar{a}_j) = \frac{\sigma_\theta^2 + Var_j^{f(X)}}{1 - \alpha^2}.$$

Se puede también calcular la varianza agregada de las acciones o decisiones en el distrito, sumando las acciones individuales y tomando el límite de N . La varianza agregada, se puede mostrar que converge a

$$Var_a = Var\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \left(a_i - \frac{\overline{f(X)}_j}{1 - \alpha}\right)\right) \rightarrow \frac{\sigma_\theta^2 + Var_j^{f(X)}}{(1 - \alpha)^2}.$$

De esta manera, si podemos observar ambas varianzas, la varianza agregada y la varianza individual, entonces podemos tomar la razón de ambas varianzas y resolver la ecuación para obtener el coeficiente de interacción social, α_i

$$\alpha = \frac{Var_a - Var_i}{Var_a + Var_i}$$

En el Cuadro 4 se presentan los resultados de estimar esta ecuación. Se estiman varios modelos. El modelo denominado Sencillo no incluye ninguna variable de control y básicamente calcula la ecuación anterior comparando la varianza individual con la varianza agregada. En este caso, el coeficiente de interacción α es 0.772. Este valor es relevante pues permite una primera aproximación a la medición del grado de interacción en las decisiones de educación entre los adolescentes y porque nos sirve como valor de comparación para las siguientes estimaciones donde se incluye un mayor control por otros factores potenciales (vea Hall (2007)).

En el segundo renglón del Cuadro 4 se muestra la estimación del coeficiente de interacción después de controlar por tres factores adicionales; los efectos de la familia, los colegios y la geografía del distrito. El primer conjunto de variables de control se enfoca en el grado de educación de los padres y algunas características de las familias, tales como, el número de niños en la familia, la presencia de ambos padres, presencia de hermanas en la familia, y la educación del jefe de familia.

CUADRO 4
INTERACCIONES SOCIALES EN LA DECISIÓN DE EDUCACIÓN EN COSTA RICA, 2003

Modelos	Número de Distritos	Varianza Individual	Varianza Agregada	Alfa α ,
Sencillo	296	1.098	8.530	0.772
Colegios como variable de control	171	0.818	4.061	0.665
Colegios como variables ficticia	296	0.818	3.947	0.657
Variable de control y efectos fijos	171	0.818	4.432	0.688

En el segundo grupo y como se discutió en la sección anterior, se incluye una serie de variables que buscan controlar por la variabilidad no observada en la calidad de educación entre los distritos, la cual puede explicar las diferencias observadas en la educación alcanzada por los adolescentes. En particular, se incluyen controles del grado de desempeño de los colegios medido, por ejemplo, por el porcentaje de estudiantes que aprueban el curso, la calidad de los maestros, y las diferencias en la calidad de las Juntas de Padres representados en los colegios.

El grupo final de variables de control en este ejercicio incluye una serie de variables geográficas. En particular, se emplea la caracterización por el sistema de zonas de vida Holdridge, el cual divide el país en 12 zonas ecológicas basado en el tipo de vegetación, condiciones del suelo y climáticas. Las aéreas más rentables para la agricultura en Costa Rica son las zonas de vida denominadas húmedas y las zonas con menor valoración son las zonas de vida denominadas zonas pluviales y las zonas tropicales secas. Para cada zona, se incluyen variables ficticias para identificar qué tipo de zona está presente en el distrito. Se esperaría que suelos más fértiles estén correlacionados con ciertas zonas más ricas en el país y por tanto con mayor nivel de educación. La presencia de estas variables geográficas busca así controlar por factores propios del distrito y no observables directamente.

Después de controlar por estos tres tipos de factores descritos, el coeficiente de interacción social α es 0.665. El grado de interacción social es aún considerable e importante. La

varianza agregada es considerablemente reducida como se puede observar en los datos. De variabilidad agregada de 8.530 años de educación, la dispersión agregada promedio se reduce a 3.947 años. Se muestra como la inclusión de este conjunto de variables de control reduce la dispersión agregada, sin embargo también se obtiene una reducción en la variabilidad individual producto de la inclusión de las variables de control de la familia. El resultado es una reducción en el tamaño del coeficiente de interacción.

Puesto que no todos los distritos tienen colegios en la muestra como se explicó anteriormente, el siguiente renglón en el Cuadro 4 repite el ejercicio después de incluir una variable ficticia que controla por los distritos que tienen un colegio, sustituyendo las variables de control a nivel del colegio. De esta forma, el número de distrito en el ejercicio no varía de 296 a 171 como fue anteriormente el caso. El coeficiente de interacción no varía considerablemente del estimado anteriormente. La variable ficticia aproxima bastante bien las características de los colegios anteriormente incluidas.

La división geográfica en Costa Rica por medio de cantones es sumamente importante desde el punto de vista político y económico (véase Hall, Arce y Monge (2000)). La agregación de distritos conforma los cantones. El último renglón del Cuadro 4 muestra la estimación de la interacción social cuando se incluyen efectos fijos que controlan por el Cantón para reducir cualquier efecto que sea producido a este nivel de agregación geopolítica. El valor de interacción se incrementa levemente pero los

resultados se mantienen. El grado de interacción social persiste en los datos en esta serie de ejercicios econométricos.

A partir de estos ejercicios econométricos realizados, se puede concluir por lo tanto, que el grado de interacción entre los adolescentes en sus decisiones de educación es considerable y explica en buena medida las decisiones de educación. Además, como se ha mostrado, los resultados son bastante robustos a la presencia de las variables de control utilizadas⁷.

4.2.2. Análisis de Regresión

Otra estrategia ampliamente utilizada en este tipo de estudios sobre la medición de la interacción social es mediante la regresión de las acciones individuales al promedio de decisión del vecindario o distrito.

Utilizando el modelo mostrado en (5) se tiene que $a_i = [\beta A_i + \Theta_i]$. Incluyendo las variables observables en el componente aleatorio, se puede escribir $\Theta_i = \theta_i + f(X_i) + \lambda_j$ donde θ_i y X_i han sido previamente definidas y λ_j es una variable fija del distrito que afecta por igual a todos los individuos de ese distrito.

El coeficiente de regresión de la decisión individual sobre la decisión promedio de los individuos en el vecindario es

$$\beta_{OLS} = \frac{Cov(a_i, \sum_{j \neq i} a_j / (m-1))}{Var(\sum_{j \neq i} a_j / (m-1))} \quad (8)$$

El coeficiente β_{OLS} es sin embargo sesgado. Si se asume que $m = n$ y $\sigma_j = 0$, y se resuelve la ecuación anterior explícitamente, se obtiene que

$$\beta_{OLS} = 2\beta \frac{(n-1)^2}{n(n-1+\beta)} - \beta^2 \frac{(n-1)^2}{n(n-1+\beta)^2} \quad (9)$$

Tomando el límite, se obtiene que el coeficiente OLS converge a $2\beta - \beta^2$ y por lo tanto la estimación por mínimos cuadrados es sesgada.

Este problema fue inicialmente observado por Jones (1990) y más tarde por Case and Katz (1991). El problema es que se la forma en que se escribe el modelo introduce un problema de simultaneidad. Esto se puede explicar del siguiente modo: no es únicamente mi vecino quien explica mi decisión de educación, sino que también mi decisión la que explica su decisión.

Para resolver este problema de simultaneidad, podemos incluir una variable instrumental. Considere el conjunto de variables de control de la familia previamente introducido y denotado ahora X . Supongamos que esta variable puede ser observada por el individuo y es parte del error individual, es decir, $\theta_i = \gamma X_i + \mu_i$. De esta manera, se descompone el término de error en una parte idiosincrática no observable y un término directamente observable. Suponga adicionalmente que estos elementos de θ_i son ortogonales a λ_j y entre ellos. De esta forma, se puede mostrar que el coeficiente de variables instrumentales es

$$\beta_{IV} = \frac{Cov(a_i, \sum_{j \neq i} a_j / (m-1))}{Cov(\sum_{j \neq i} a_j / (m-1), \sum_{j \neq i} X_j / (m-1))} = \frac{\beta}{\beta + (1-\beta) \frac{(n-1)}{(m-1)}} \quad (10)$$

Cuando $m=n$, el estimador β_{IV} se reduce a β y el estimador es insesgado.

La variable escolaridad del jefe de familia es utilizada como el instrumento para identificar los efectos de interacción social (vea Ashenfelter y Zimmerman (1997)). Para cada uno de las decisiones individuales, se utiliza el promedio de educación de los padres para instrumentar el promedio de educación escogido de los adolescentes en el distrito (vea Jones (1990) y Case y Katz (1991)). Este tipo de instrumento satisface las características deseadas, presumiendo que no hay un serio problema de autoselección en las localidades como se ha mencionado anteriormente. En este caso particular, la educación promedio de los padres de los adolescentes en mi vecindario es un buen instrumento de la educación de sus hijos en mi vecindario en el tanto satisface las dos condiciones deseadas. Por un lado, la educación promedio de los padres esta correlacionado con

7 Vease Hall (2007) donde se incluyen más detalle de las estimaciones realizadas y algunos anexos con descripciones más precisas de los datos y metodologías empleadas.

la educación promedio de los hijos e hijas en mi vecindario. Por otro lado, dicha educación promedio de los padres es ortogonal a los términos de error en mi ecuación individual, es decir, no se correlaciona con otros factores que explican mi desempeño individual y que han sido omitidos de mi regresión y forman por tanto parte del término de error.

Los resultados de la estimación se resumen en el Cuadro 5. El Cuadro se compone de dos conjuntos de resultados. La parte superior del Cuadro muestra los resultados cuando se incluyen las variables de control. La estimación emplea el mismo grupo de variables de control anteriormente utilizado, familias, colegios y geografía. La parte inferior sustituye estas variables de control con una variable ficticia que denota la presencia o no de un colegio público en el distrito, con el fin de mantener el tamaño de la muestra.

El primer renglón del Cuadro reporta el valor de la interacción cuando no se emplea ninguna variable de control y se emplea como valor de referencia. En el segundo renglón se muestra el coeficiente de interacción de 0.31 cuando se controla por los factores anteriormente descritos. Este valor significa que por cada año promedio de educación de los adolescentes en el distrito, el individuo aumenta su

nivel de educación en un tercio de año adicional. El coeficiente es además significativo después de controlar por estos factores.

En el resto del Cuadro se muestra el valor del coeficiente cuando se incluyen subconjuntos de las variables de control para determinar que tan robusto es el resultado anteriormente obtenido. En particular, es relevante observar que el valor del coeficiente es casi idéntico en caso de que controlemos únicamente por los efectos del colegio o por los efectos geográficos. Sin embargo, el valor es considerablemente reducido cuando se incluye las variables de control al nivel de la familia. En este caso el valor del coeficiente pasa de 0.90 a 0.28. Los resultados se repiten en la parte inferior del Cuadro sustituyendo las variables de control al nivel de distrito por la variable ficticia de presencia o ausencia de un colegio público en el distrito. En este caso, el valor del coeficiente es de 0.22 y significativo.

Se puede concluir que hay una considerable reducción en el grado de interacción social cuando controlamos específicamente por factores de la familia en contraste con los efectos geográficos o de colegio. El coeficiente es, sin embargo, significativo e importante aun después de incluir dichos controles. Cada año de educación promedio de los adolescentes en

CUADRO 5
ANÁLISIS DE REGRESIÓN EN LA INTERACCIÓN ESCOLAR EN COSTA RICA, 2003

Coeficiente Interacción	t-estadístico	Variables de Control N=2936		
0.9025	(17.02)			
0.3156	(4.35)	Familia	Colegios	Geográfica
0.2881	(5.01)	Familia		
0.8370	(10.65)		Colegios	
0.8903	(12.82)			Geográfica
Coeficiente Interacción	t-estadístico	Variables de Control N=4344		
0.9024	(15.20)			
0.2234	(3.83)	Familia	Colegios	Geográfica
0.2707	(5.30)	Familia		
0.8689	(15.05)		Colegios	
0.8064	(13.30)			Geográfica

Resumen de ejercicio econométricos, ver detalles en Hall (2007).

el distrito tiende a producir un incremento de un tercio de año adicional de educación en los adolescentes del distrito. Combinando todas las variables de control, el efecto del vecindario tiende a crecer ligeramente.

Un problema latente en estos análisis es el problema de autoselección o *sorting*. Los padres por ejemplo con un mismo nivel de educación podrían haber elegido vivir en los mismos distritos por distintas razones de asociación. Esta decisión de localización podría entonces explicar el nivel de desempeño observado en el grupo el cual no responde propiamente a la interacción. La corrección adecuada de este problema requeriría de una información más detallada de la que la encuesta provee sobre las localizaciones. En general, el grado de movilidad dentro y fuera de los distritos es menor, por lo que la autoselección podría ser reducida⁸.

Para ambas estrategias de estimación, se puede por lo tanto concluir que el efecto de interacción social entre los adolescentes es significativo e importante. Los resultados obtenidos por la estrategia de identificación por diferencia en las varianzas tienden a ser más estable a la presencia de las variables de control.

CONCLUSIÓN

El objetivo principal de este trabajo es evaluar el grado de interacción social en las decisiones de educación de los adolescentes en Costa Rica. El estudio provee evidencia que apoya la hipótesis de un alto grado de interacción entre los adolescentes al decidir sobre su nivel de escolaridad. Los resultados son robustos a la presencia de una serie de variables de control que buscan reducir el efecto de posibles variables no observables que pueden influenciar este tipo de decisión. En particular, se incluyen variables de control a nivel de la familia, del colegio y algunos efectos geográficos de la localidad.

8 Estimaciones preliminares sugieren que menos de un 10% de las familias reporta dicha movilidad entre cantones.

El ejercicio econométrico favorece la utilización de la estrategia de identificación basado en los componentes de la varianza, en la medida en que los valores son más robustos a la presencia de variables de control y además por su propia construcción, reduce los problemas de identificación derivados de la simultaneidad de las decisiones.

Como se muestra en el documento para el caso de Costa Rica y ha sido objeto de estudio en la literatura en general, la calidad de las escuelas, los factores familiares y los factores individuales son elementos relevantes y determinantes del nivel de educación de los adolescentes. El trabajo permite, sin embargo, valorar también la importancia de los factores de agrupación o de interacción social de los adolescentes. La decisión de educación de los otros estudiantes es sumamente importante, como se ha demostrado en el documento, en la decisión de educación individual del adolescente. En este documento se ha hecho especial énfasis, dada la naturaleza de los datos, en valorar dichas interacción en el nivel de educación alcanzado, particularmente el número de años de educación. En este sentido, es sumamente importante para efectos de diseño de política pública el distinguir adecuadamente los factores que inciden en las decisiones de educación de los adolescentes, puesto que la incidencia que dichas medidas tengan será mayor en el tanto las medidas adoptadas sean más precisas al conjunto de incentivos. El documento se limita a sugerir que dichos elementos de interacción social son sumamente importantes en las decisiones de deserción escolar la cual convendría estudiar más en detalle en investigaciones futuras.

El estudio deja abierta la posibilidad de problemas derivados de la autoselección de las familias en los distritos, es decir las decisiones de localización. La decisión de mudarse o no a un cierto distrito como resultado de los incentivos provistos en las distintas comunidades puede explicar el desempeño de los adolescentes. A pesar de que el grado de movilidad observado entre los distritos en Costa Rica es reducido, el estudio puede ser extendido para corregir apropiadamente por este problema potencial.

Finalmente, otro punto de interés para futuras investigaciones en este tema es el de incorporar explícitamente la decisión de deserción escolar y valorar su interacción con los efectos de interacción social local y global. Como se detalló en el texto, alrededor de un cuarto de la muestra desertó de la educación. El foco principal de atención en este trabajo es estudiar el número de años de educación de los adolescentes alcanzado al momento de la entrevista. Sin embargo muchos adolescentes ya han desertado. Esta decisión debería ser explícitamente modelada para determinar cuanta interacción existe en este tipo de decisión.

BIBLIOGRAFÍA

- Ashenfelter, O. and Zimmerman, D. (1997): "Estimates of the Return to Schooling from Sibling Data: Father, Sons and Brothers," *Econometrica*, 79, 1, 1-9.
- Borjas, G. (1995): "Ethnicity, Neighborhoods and Human-Capital Externalities," *American Economic Review*, 85, 3, 365-390.
- Card, David. (1999): "The causal effect of education on earnings," in Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics*, Amsterdam: Elsevier Science.
- Case, Anne C. and Lawrence F. Katz (1991): "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth," working paper 3705, NBER.
- Glaeser, E. and Jose A. Scheinkman (2001): "Measuring Social Interactions," in *Social Dynamics* edited by S. Durlauf and P. Young. Cambridge: MIT Press.
- Glaeser, E. and Jose A. Scheinkman (2002): "Non-Market Interactions," published in *Advances in Economics and Econometrics: Theory and Applications*, Eight World Congress, M. Dewatripont, L.P. Hansen, and S. Turnovsky (eds.), Cambridge University Press, 2002.
- Hall, Luis, Gilberto Arce and Alexander Monge (2002): *Bailouts in Costa Rica*, Inter-American Development Bank Working Paper R-475.
- Hall, Luis, *Social Interactions in Schooling Choice in Costa Rica* (2007), Chapter 2, Doctoral Dissertation, New York University, Department of Economics.
- Jones, Stephen R. G. (1990): "Worker Interdependence and Output: The Hawthorn Studies Reevaluated," *American Sociological Review* 55, 176-190.
- Manski, Charles (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531-542.
- UNESCO (2006): *EFA Global Monitoring Report 2006*.
- US Census Bureau, 2002.