

Las colecciones de Documentos de Trabajo del CIDE representan un medio para difundir los avances de la labor de investigación, y para permitir que los autores reciban comentarios antes de su publicación definitiva. Se agradecerá que los comentarios se hagan llegar directamente al (los) autor(es). ♦ D.R. © 2001, Centro de Investigación y Docencia Económicas, A. C., carretera México-Toluca 3655 (km.16.5) Lomas de Santa Fe, 01210 México, D. F., tel. 727-9800, fax: 292-1304 y 570-4277. ♦ Producción a cargo del (los) autor(es), por lo que tanto el contenido como el estilo y la redacción son responsabilidad exclusiva suya.



NÚMERO 32

Andrés Zamudio y Teresa Bracho

**RENDIMIENTOS ECONÓMICOS DE LA ESCOLARIDAD III:
EL PROBLEMA DE SESGO POR ELECCIÓN**

Prólogo

El presente documento forma parte de una serie de tres, derivados del proyecto de investigación "Los rendimientos económicos de la escolaridad" a cargo de Teresa Bracho de la División de Estudios Políticos, y Andrés Zamudio, de la División de Economía, y forma parte del programa de investigaciones "Políticas y decisiones educativas" que se realiza en el CIDE, que busca vincular aspectos sociales, políticos y económicos de la estructura y procesos educacionales en México.

El presente documento discute uno de los probables sesgos que se presentan en las estimaciones simples a las tasas de retorno, esto es, el sesgo resultado de no considerar la variable de escolaridad como una variable de elección de los individuos.

Los análisis empíricos se basan en la muestra original de la Encuesta Nacional de Ingreso y Gasto de los Hogares 1989, del Instituto Nacional de Estadística, Geografía e Informática. El documento se inicia con una discusión sobre el problema del sesgo por elección, para después pasar a la estimación de los rendimientos con y sin ajuste por sesgo.

Estos documentos tienen como antecedentes trabajos de investigación previos de los dos autores, de los que destacamos los más recientes: *Gasto educativo privado* de T. Bracho (Documento interno, CIDE, 1992); *Tasas de retorno a la educación* de A. Zamudio (Documento interno, CIDE, 1993). *Tasas de retorno: el sesgo por elección*, de A. Zamudio (Conferencia, II Congreso Nacional de Investigación Educativa, octubre de 1993).

Ambos autores son igualmente responsables de la investigación y de sus resultados. El orden en que aparecen los autores en cada documento indica en quien recayó la responsabilidad de la redacción final de ese documento y de ninguna manera señala responsabilidades diferenciales en la investigación o en sus distintos segmentos.

A.Z. y T.B.

Introducción

A partir de los años sesenta se han producido una gran cantidad de investigaciones sobre los rendimientos económicos a la escolaridad. La mayoría de los estudios se han realizado para información sobre Estados Unidos o sobre países desarrollados; sin embargo, existe una literatura importante sobre estudios hechos sobre países en desarrollo y, en particular, sobre América Latina.

El método para llevar a cabo los cálculos ha variado a través del tiempo.¹ Primeramente el método de estimación consistió en el simple cálculo de la tasa interna de retorno a la inversión educativa. Posteriormente se utilizó la llamada ecuación minceriana en su forma simple. Esto dio origen a una primera generación de estimaciones las cuales dieron por resultado retornos altos.² Desde mediados de los setenta se empezaron a discutir algunos problemas de la estimación. De esta manera se introdujeron elementos como la heterogeneidad de los individuos, la omisión de variables relevantes, diferencias en la calidad de la educación, sesgo por elección de los individuos y otros problemas. Al llevar a cabo las estimaciones con los ajustes necesarios, se produjeron cambios significativos en la magnitud de los retornos. En la mayoría de los casos los ajustes produjeron retornos más bajos. Sin embargo, en algunos casos, el ajuste fue hacia arriba. Éste ha sido el caso, en términos generales, cuando se ajusta por la endogeneidad del nivel de educación, es decir, cuando se toma en cuenta que el nivel de escolaridad es una variable de elección del individuo.

En este documento se discute y revisa la evidencia empírica que existe para México sobre el posible sesgo por elección en las estimaciones simples de las tasas de retorno. Con este fin se calculan tasas de retorno a la educación con y sin ajuste para llevar a cabo la comparación. Para la estimación se utiliza información proveniente de la Encuesta Nacional de Ingresos y Gastos de los Hogares, 1989.³

El problema del sesgo por elección

En el primer documento de esta serie se habló de la posibilidad de que las estimaciones simples de los retornos estuvieran sesgadas. Una de las posibles fuentes de este sesgo proviene de la endogeneidad de la escolaridad, es decir, del hecho de que la escolaridad es una variable de elección para el individuo.

Este problema del sesgo por elección se debe a que en la estimación simple de la ecuación minceriana se están comparando de manera implícita los ingresos de indivi-

¹ Véase el primer documento de esta serie: *Discusión teórica y métodos de estimación*.

² Véase N. Gaston, y J. Tenjo (1992).

³ Para una descripción de la muestra de ENIGH-1989, de las variables centrales utilizadas en este cálculo, y los resultados de la estimación de tasas de retorno por diferentes métodos, véase el segundo documento paralelo de esta serie: *Estimaciones para el caso mexicano, 1989*.

duos que no son comparables, esto es, de individuos que son heterogéneos. Esta heterogeneidad tiene su origen tanto en las características personales de los individuos y, en las características familiares como en ciertos factores regionales. Las características personales se pueden expresar en diferencias en habilidad, motivación, etcétera. Las características familiares consistirían en el ingreso familiar, el nivel educativo y cultural de los padres, el tamaño del hogar y otro más. Los factores regionales consistirían tanto en el medio ambiente como en la oferta de servicios educativos disponible.

En forma simple, al calcular los retornos a la escolaridad se comparan las trayectorias de ingresos de individuos con y sin un nivel determinado de educación, ajustado por los costos directos de la educación. Por ejemplo, para calcular los retornos a la educación superior se comparan los ingresos, digamos, de un ingeniero con los ingresos de un técnico, tomando en cuenta que el ingreso del ingeniero durante el periodo de aprendizaje es negativo, lo que es reflejo de los costos directos. Sin embargo esta comparación puede no ser la mejor para calcular los retornos. El hecho de que la escolaridad sea una variable de elección indicaría que la persona que eligió llevar a cabo estudios superiores lo hizo porque —en su percepción— esto era óptimo para él; del mismo modo, la persona que eligió terminar sus estudios de nivel intermedio lo hizo porque juzgó mejor esto.⁴ Presumiblemente la persona que eligió ser ingeniero habría obtenido ingresos inferiores a los que registra el técnico si no hubiera llevado a cabo estudios superiores. En este caso los costos de oportunidad del ingeniero, que en este caso son iguales a los ingresos del técnico, resultarían sobreestimados, por lo que el retorno a la educación superior resultaría subestimado.

El hecho de que la escolaridad sea una variable de elección, hace que la información disponible sobre los individuos sea una información truncada, ya que solamente se tiene información sobre los individuos una vez que éstos han tomado una decisión. Sin embargo, para calcular los retornos se requeriría idealmente, a fin de incorporar el costo de oportunidad correcto, información sobre los individuos para las dos elecciones en consideración en la trayectoria educativa: continuar o no continuar con los estudios;⁵ sin embargo, esto es imposible con la información disponible. En este caso, la distribución observada en el nivel de escolaridad entre los individuos no es el resultado de una distribución aleatoria de educación entre los individuos, sino el resultado de la elección propia de los individuos. Esto quiere decir que los mismos individuos se clasifican entre diferentes niveles educacionales de acuerdo con sus preferencias o restricciones, por lo cual, los ingresos observados por nivel educativo corresponden a ingresos óptimos. Por ello, los costos de oportunidad para una persona que desea continuar con un cierto nivel de educación resultan sobreestimados por los ingresos de una persona con un nivel educativo inferior, con lo que los retornos a la inversión en

⁴ Por ahora mantenemos esta terminología de las elecciones educativas; sin embargo, reconocemos que el punto "óptimo" de elección en la trayectoria educativa puede ser juzgada por el sujeto como no óptimo (o deseable desde su percepción de habilidades), sino como "necesario", en condiciones de restricción de oferta, necesidades de trabajo, etcétera.

⁵ En el límite, la elección sería educarse o no educarse.

educación resultan subestimados. Si los retornos a la escolaridad son calculados de una manera simple, los resultados no se pueden generalizar como aplicables a cualquier individuo tomado al azar.

La forma estadística más usada para abordar, o ajustar por, el problema de sesgo por elección se utiliza cuando se tiene una ecuación de escolaridad dicotómica, es decir, una ecuación con solamente dos alternativas para el individuo: llevar un programa particular de educación o no hacerlo. Por ejemplo, considérese el siguiente modelo.⁶ Se tiene un cierto individuo que tiene la elección de llevar a cabo un programa educativo, digamos estudios superiores. El ingreso que obtendrá dependerá de su elección, y la decisión de continuar con el programa educativo dependerá de ciertas variables socioeconómicas. Estas relaciones se pueden expresar con las siguientes ecuaciones:

$$\log(Y_{si}) = y_{si} = X_{si}\beta_s + u_{si} \quad (1)$$

$$\log(Y_{mi}) = y_{mi} = X_{mi}\beta_m + u_{mi} \quad (2)$$

$$I^*_i = Z_i\delta - v_i \quad (3)$$

$$I_i = 1 \quad \text{si } I^*_i > 0 \quad (4)$$

$$I_i = 0 \quad \text{de otro modo}$$

$$E(u_{si}) = E(u_{mi}) = E(v_i) = 0, \quad (5)$$

$$\text{Var}(u_{si}) = \sigma_{ss}, \quad \text{Var}(u_{mi}) = \sigma_{mm}, \quad \text{Var}(v_i) = \sigma_{vv}$$

$$\text{Cov}(u_{si}, v_i) = \sigma_{sv}, \quad \text{Cov}(u_{mi}, v_i) = \sigma_{mv}$$

Si el individuo decide continuar con estudios superiores su ingreso estará dado por (1). Si decide no continuar, sus ingresos serán (2). En estas dos ecuaciones X representa un vector de características personales que influyen en los ingresos, β es un vector de parámetros por estimar, el cual depende de la elección tomada, u es un residual. La decisión a llevar a cabo el programa educativo se expresa en (3), donde Z es un vector de características personales, que influyen en la toma de decisión, δ es un vector de parámetros por estimar y v un residual. En esta ecuación I^* es una variable latente, la cual solamente es observada en forma binaria, es decir, cuando la condición (4) se cumple. Esta ecuación de selección determina cuál de las ecuaciones de ingreso es observada, por ejemplo, si $I = 1$ entonces la ecuación (1) es observada, si $I = 0$ entonces (2) es observada. La ecuación (5) expresa los momentos de las distribuciones para los residuales de las tres ecuaciones.

Como la ecuación (1) o (2) solamente es observada si se cumple una cierta condición, el valor medio del logaritmo del ingreso tendrá la forma (asumiendo que los residuales tienen una distribución normal trivariada y omitiendo el subíndice que denota al individuo),

⁶ Véanse J. Heckman, (1979) o L. Lee, y R. Trost (1978).

$$E(y_s) = X_s\beta_s + E[u_s | I = 1] = X_s\beta_s - \sigma_{sv} \varphi(Z_i\delta) / F(Z_i\delta)$$

$$E(y_m) = X_m\beta_m + E[u_m | I = 0] = X_m\beta_m + \sigma_{mv} \varphi(Z_i\delta) / [1 - \Phi(Z_i\delta)]$$

donde, $\varphi(t)$ es la función de densidad normal estandarizada evaluada en el punto t , $\Phi(t)$ es la función de distribución normal evaluada en t .

Si la covarianza entre los residuales de las ecuaciones de ingreso y de escolaridad (ecuación 3) son diferentes de cero, esto es, los parámetros φ_{sv} y Φ_{mv} son distintas de cero, el uso de mínimos cuadrados ordinarios para estimar las ecuaciones (1) y (2) daría por resultado estimadores sesgados. Este sesgo sería el resultado de la omisión de variables en la estimación, ya que los noobservables de la ecuación de escolaridad deberían ser utilizados en la estimación de las ecuaciones de ingreso.

La solución al problema de sesgo consiste en llevar a cabo una estimación en dos etapas. Primeramente, estimar la ecuación de escolaridad con el uso de modelos *probit*. Con los parámetros estimados se pueden calcular las expresiones $\varphi(t)$ y $\Phi(t)$ y después incorporarlas en la estimación de las ecuaciones de ingreso, las cuales ahora sí pueden ser estimadas por mínimos cuadrados ordinarios. A los parámetros estimados en esta segunda etapa se les puede interpretar como libres del sesgo por elección.

El problema principal que tiene el método anteriormente descrito radica en que solamente es aplicable a ecuaciones de escolaridad dicotómicas, esto es, cuando el nivel de escolaridad es una variable discreta y cuando solamente se tienen dos alternativas educacionales.

Cuando se considera el ciclo escolar completo, resulta que la elección del nivel de escolaridad no es una elección dicotoma sino policótoma, es decir se tienen más de dos alternativas. Esta elección policótoma se puede modelar como una elección dicotómica secuencial, es decir en un momento dado el individuo no tiene que decidir sobre el total de alternativas que ofrece el ciclo escolar, sino básicamente tiene dos alternativas: continuar con el siguiente ciclo escolar o no hacerlo,⁷ con lo cual se podría extender el modelo dicotómico para cubrir el ciclo escolar completo. Sin embargo, en este documento se trabaja con una variable de escolaridad continua, lo que hace más sencillo el método de ajuste.

El método de ajuste utilizado en este trabajo es el método propuesto por Garen (1984).⁸ El método de estimación se describe con detalle en el apéndice y consiste básicamente en la estimación de las siguientes ecuaciones,

⁷ Esto no es del todo cierto para algunos nodos. Por ejemplo cuando el individuo termina la secundaria tiene realmente tres alternativas: terminar con el ciclo educativo, continuar con el siguiente ciclo de educación universal (preparatoria) o cambiar a una carrera técnica corta.

⁸ Un problema con el método de Garen consiste en que el nivel de escolaridad debe ser una variable acotada, la cual no puede asumir cualquier valor real. En este sentido, un modelo *tobit* para la escolaridad sería el más adecuado.

$$y = \beta_0 + \beta_1 X + \beta_2 S + \beta_3 X^2 + \beta_4 S^2 + (e + \varphi \cdot S)$$

$$S = \beta_0 + \beta_1 R + w$$

La variable X denota el vector de variables que determinan los ingresos, como pueden ser la experiencia, horas trabajadas, etcétera, S los años de educación formal y R es un vector de variables que determinan la decisión a escolarizarse, como las características familiares, zona de origen, etcétera ($e + \varphi \cdot S$) y w representan los residuales. El término $\varphi \cdot S$ permite que los no observables afecten la formación de ingresos de manera distinta, dependiendo del nivel de educación. A partir del proceso de selección del individuo, los residuales de ambas ecuaciones se encuentran correlacionados,⁹ por lo cual el valor esperado del residual ($e + \varphi \cdot S$) —dada una elección del nivel de escolaridad S — no es cero. En este caso los estimadores mínimocuadráticos de los parámetros de la primera ecuación son sesgados, razón por la cual este sistema de dos ecuaciones no podría estimarse de una manera recursiva simple.

Como se discute en un apéndice al final de este capítulo, la solución propuesta por Garen (1984), la cual es similar al caso cuando se tiene una ecuación de escolaridad dicotómica,¹⁰ es incluir el sesgo en la primer ecuación, y estimar ésta por MCO. Así, se tendría que estimar primero la ecuación de escolaridad aplicando MCO,¹¹ utilizando los residuales de esta última ecuación, denotados por w , se construirían dos nuevas variables, w y $w \cdot S$, las cuales se incluirían en la estimación de la ecuación de ingresos. De este modo, la ecuación de ingresos por estimar sería,

$$y = \beta_0 + \beta_1 X + \beta_2 S + \beta_3 X^2 + \beta_4 S^2 + d_1 w + d_2 w \cdot S + v$$

siendo que los estimadores de los parámetros de esta ecuación se encuentran ahora libres del sesgo por elección.

Estimaciones para México

Para realizar la estimación se utilizó información de la Encuesta Nacional de Ingresos y Gastos de los Hogares, 1989 (ENIGH, 1989). En esta encuesta se encuentran datos sobre características de los individuos como nivel de educación, edad, tipo de trabajo, rama de actividad e ingresos. Con esta información es posible el llevar a cabo estima-

⁹ Véase J. Garen (1984).

¹⁰ J. Heckman (1979).

¹¹ Aplicar MCO a la ecuación de escolaridad no es del todo correcto, ya que la variable de escolaridad, aunque continua, es una variable acotada, la cual no puede tomar ya sea valores negativos o valores positivos muy grandes. En este caso, se podrían hacer mejoras en la estimación si se utilizan algunos de los métodos estadísticos disponibles para el caso de variables dependientes limitadas.

ciones sobre los retornos a la educación para toda la muestra. Sin embargo, existe el problema de que la ENIGH, 1989 no contiene datos sobre las características familiares para todos los individuos, siendo que este tipo de información es de suma importancia para la estimación de la ecuación de escolaridad. Por esta razón, se tuvo que trabajar con una submuestra de la ENIGH, 1989.

Así, la submuestra que se eligió para llevar a cabo la estimación por sesgo, requirió ciertas acotaciones. Las restricciones son las siguientes:

1) A fin de incluir en la estimación datos sobre las características de los padres y, teniendo en cuenta que la información que proporciona INEGI se refiere a las familias, se incluyeron en la muestra solamente individuos que todavía vivían con sus padres.

2) Como es muy importante en la estimación el ingreso familiar, o en este caso el ingreso del jefe del hogar, se introdujo una limitante adicional al considerar solamente los casos en que el ingreso familiar fuera positivo.

3) También fue necesario acotar la edad de los individuos, por lo que solamente se consideraron individuos entre 15 y 40 años.

4) Finalmente sólo se consideraron los individuos que trabajaran y reportaran un sueldo positivo.

Dadas estas restricciones impuestas sobre la muestra, los resultados de esta estimación no se pueden generalizar a toda la población. Sin embargo, los resultados presentados son importantes en un sentido exploratorio, en cuanto que indican la posible existencia del sesgo por elección en las estimaciones simples de las tasas de retorno a la educación.

La muestra resultante tiene un tamaño de 3 100 individuos.¹² De estos 1 154 son mujeres y 1 946 hombres. La muestra utilizada cubre a toda la república mexicana, donde 72% de los casos corresponden a la población urbana y 28% a la rural.¹³

Las variables utilizadas en la estimación son las siguientes.

- *LW* representa el logaritmo natural del sueldo del individuo del mes anterior.
- *LHT* es el logaritmo natural de las horas trabajadas durante una semana (*HRSEMT*).
- *S* representa los años de educación formal del individuo.
- *EX* es una aproximación a los años de experiencia laboral, ésta se calculó, como es estándar en este tipo de estudios, como $EX = EDAD - S - 6$.
- *EX2* es el cuadrado de la experiencia.
- *SJ* son los años de educación del jefe del hogar.
- *INGRESOJ* representa al ingreso del jefe del hogar.

¹² Para la estimación llevada a cabo en este capítulo no se incluyeron los factores de expansión.

¹³ Como ya se señaló en los capítulos anteriores, el INEGI divide las zonas entre alta y baja densidad poblacional. Aquí estamos identificando estas zonas como urbana y rural respectivamente.

- *ZONA* es una variable *dummy* que toma el valor unitario cuando el individuo reside en una zona urbana y cero en caso contrario.
- *TAMH* es el tamaño del hogar.

En el cuadro 1 se resumen algunas estadísticas sobre las variables utilizadas.

Cuadro 1
Descripción estadística de las variables utilizadas

<i>Variable</i>	<i>Media</i>	<i>Desv. Estándar</i>
Sueldo	357570.8	275942.1
<i>HRSEMT</i>	43.9	11.7
<i>S</i> (total)	8.7	3.8
<i>S</i> (mujeres)	9.7	3.6
<i>S</i> (hombres)	8.2	3.8
<i>EDAD</i> (total)	22.3	5.3
<i>EDAD</i> (mujeres)	23.2	5.6
<i>EDAD</i> (hombres)	22.2	5.0
<i>TAMH</i>	7.1	2.6
<i>SJ</i>	4.2	3.5

Cabe destacar del cuadro la diferencia tan grande que hay entre los promedios de escolaridad del jefe familiar (*SJ*) y el de los hijos (*S*) y que se puede explicar en alguna medida por las diferencias generacionales en la distribución educativa.¹⁴

Como se indicó anteriormente, la estimación de los retornos se hace en dos etapas. En primer lugar, se estimó la llamada ecuación de escolaridad, en la cual la variable dependiente son los años de educación, mientras que las variables explicativas son el nivel de educación y el ingreso del jefe familiar, el tamaño del hogar y la zona de residencia.¹⁵ Después se utilizan los residuales de esta ecuación para estimar los retornos con la ecuación de ingresos.

En el cuadro 2 se presentan los resultados de la estimación de la ecuación de escolaridad para tres casos: total de la población, para las mujeres y para los hombres. En el cuadro puede observarse que los parámetros estimados son estadísticamente significativos y que corresponden al signo esperado. Por ejemplo, se esperaría que un in-

¹⁴ Véase Bracho (1989) para la comparación entre escolaridad de los jóvenes estudiantes, sus hermanos mayores y sus padres.

¹⁵ Este tipo de variables explicativas para la ecuación de escolaridad ya se han utilizado en otra parte (véase Gaston y Trenjo, 1992).

dividuo de una zona urbana obtenga más altos niveles de educación que un individuo de una zona rural, por lo que el coeficiente de la variable *ZONA* tendría que ser positivo. El tamaño del hogar, se ha argumentado, influye de manera negativa en los logros educacionales, por lo que se esperaría que su coeficiente fuera negativo. Se esperaría que el nivel educativo del padre tuviera una influencia positiva sobre la educación del hijo, por lo que el coeficiente de la variable *SJ* tendría que ser positivo. Del mismo modo, es de esperarse que los ingresos familiares influyan de manera positiva en los logros educacionales (vía la eliminación o disminución de restricciones financieras).

Cuadro 2
Ecuación de escolaridad*
(Variable dependiente: *S*)

<i>Variable</i>	<i>Total</i>	<i>Mujeres</i>	<i>Hombres</i>
<i>CONSTANTE</i>	6.420 (31.7)	7.499 (20.9)	6.037 (24.5)
<i>ZONA</i>	2.428 (18.1)	2.018 (8.0)	2.356 (14.7)
<i>TAMH</i>	0.137 (6.3)	0.148 (4.2)	0.128 (4.7)
<i>SJ</i>	0.353 (19.7)	0.323 (11.7)	0.366 (15.7)
<i>INGRESOJ</i>	0.148 (3.8)	0.138 (2.2)	0.155 (3.2)
<i>R</i> ²	0.293	0.234	0.305
<i>Muestra</i>	3100	1154	1946

* Estadístico *t* entre paréntesis

Los resultados de la estimación para los tres casos registrados son muy similares. Sin embargo, es de destacar el mayor efecto negativo que tiene el tamaño del hogar sobre los logros educacionales de la mujer en comparación con los del hombre (0.148 contra 0.128). También se puede ver que el nivel educativo del padre, su ingreso y la zona de residencia tiene un mayor efecto positivo sobre la elección de educarse en los hombres que en las mujeres.¹⁶

El siguiente paso en la estimación es obtener los retornos vía la ecuación de in-

¹⁶ La diferencia en el valor absoluto de estos parámetros no es muy grande, por esta razón se tendrían que llevar a cabo pruebas de hipótesis formales para verificar si esta diferencia es estadísticamente importante.

gresos. Para este efecto se procedió del siguiente modo: en primer lugar, se estimó esta ecuación sin considerar el ajuste por el sesgo, esto es, se estimó la ecuación minceriana simple. Después se estimó la misma ecuación añadiendo las variables que fueron construidas con los residuales de la ecuación de escolaridad. En el cuadro 3 se muestran los resultados de la estimación a partir de la ecuación minceriana simple y el cuadro 4 presenta los resultados cuando se incorpora el ajuste por el sesgo.¹⁷

La estimación indica retornos simples (cuadro 3) algo bajos en comparación con los obtenidos para toda la muestra (presentados en el segundo documento de esta serie: *Estimaciones para el caso mexicano, 1989*).

Cuadro 3
Ecuación de Ingresos sin ajuste*
Variable dependiente: log (Sueldo)

<i>Variable</i>	<i>Población Total</i>	<i>Mujeres</i>	<i>Hombres</i>
<i>CONSTANTE</i>	9.724 (76.6)	9.924 (51.7)	9.562 (57.1)
<i>S</i>	0.106 (33.2)	0.108 (22.4)	0.107 (24.6)
<i>EX</i>	0.057 (10.0)	0.043 (5.2)	0.066 (8.7)
<i>EX2</i>	0.001 (5.8)	0.001 (2.3)	0.002 (5.4)
<i>LHT</i>	0.423 (13.4)	0.372 (7.8)	0.457 (11.0)
<i>R²</i>	0.289	0.320	0.270
<i>Muestra</i>	3100	1154	1946

* Estadístico *t* entre paréntesis

En cuanto a los resultados a partir de la estimación con ajuste, se registra la existencia del sesgo por elección en la estimación simple de los retornos. Compárense para ello los coeficientes que corresponden a la variable *S* en las dos ecuaciones, con y sin ajuste. Esta diferencia se presenta en el cálculo para el total de la población y

¹⁷ En la estimación no se consideró la posible heterogeneidad del modelo, por esta razón, las estimaciones de las varianzas de coeficientes podrían estar subestimadas.

para hombres y mujeres. En el caso de la población total, el salto es de un retorno simple estimado en 10.6%, a un retorno de 17.4% cuando se incluye el ajuste.¹⁸

Es importante señalar que, en relación con el resto de las variables explicativas consideradas, el ajuste hecho en la estimación no produce resultados muy diferentes, esto es, los parámetros estimados por la ecuación minceriana simple son robustos al ajuste por sesgo realizado.

Cuadro 4
Ecuación de Ingresos con ajuste*
Variable dependiente: log (Sueldo)

<i>Variable</i>	<i>Población Total</i>	<i>Mujeres</i>	<i>Hombres</i>
<i>Constante</i>	9.086 (71.4)*	9.067 (44.2)	9.001 (54.5)
<i>S</i>	0.174 (33.5)	0.181 (19.9)	0.178 (26.0)
<i>EX</i>	0.058 (10.7)	0.044 (5.6)	0.067 (9.3)
<i>EX2</i>	0.002 (7.1)	0.001 (3.0)	0.002 (6.5)
<i>LHT</i>	0.429 (14.2)	0.403 (8.8)	0.445 (11.2)
<i>ERR</i>	0.133 (14.9)	0.139 (9.4)	0.134 (11.7)
<i>ERR•S</i>	0.004 (5.8)	0.005 (4.6)	0.003 (3.6)
<i>R²</i>	0.349	0.374	0.334
<i>Muestra</i>	3100	1154	1946

* Estadístico *t* entre paréntesis

Los signos para las variables de ajuste, *ERR* y *ERR•S*, resultaron ser negativo y positivo, respectivamente. Este resultado no es nuevo, ya que otros estudios¹⁹ han registrado resultados similares. Esto quiere decir que residuales negativos en la ecua-

¹⁸ Resultados similares se presentan en los trabajos de Garen (1984) y Gaston y Trenjo (1992).

¹⁹ *Ibid.*

ción de escolaridad tenderán a estar asociados con residuales positivos de la ecuación de ingresos en niveles bajos de S . De esta manera, si un individuo tuviera un nivel inesperadamente bajo de educación, esto es, bajo en relación con lo que se esperaría, dada la ecuación de escolaridad y el valor asumido por las variables explicativas, se esperaría que esta persona registrara un nivel de ingresos inesperadamente alto para su nivel educativo. Del mismo modo, las personas que tuvieran un nivel inesperadamente alto de educación (igualmente a niveles bajos de S) serían personas que tenderían a reportar ingresos inesperadamente bajos, dado su nivel de educación. Esta relación negativa entre los noobservables se irá reduciendo a medida que aumente el nivel de escolaridad. Garen interpreta esta combinación de signos para los coeficientes de las variables de ajuste como una clara indicación de la presencia de ventajas comparativas en la elección de la escolaridad. Esto es, los individuos elegirán el nivel educativo que corresponda al tipo de trabajo para el que son comparativamente mejores. En este caso la variable habilidad no debería ser modelada como un solo factor, es decir ésta no podría estar asociada simplemente a un índice de habilidad como son los IQ, habría necesidad de incorporar a otros aspectos de la habilidad o características personales.

Observaciones finales.

El ejercicio llevado a cabo en este documento sugiere la existencia del sesgo por elección propia en las estimaciones simples de la tasa de retorno a la inversión educativa, esto es, en las estimaciones basadas en la ecuación minceriana simple. Aunque este resultado pudiera ser válido únicamente para la muestra restringida que tuvimos que seleccionar, es un resultado que muestra congruencia con los obtenidos en otros estudios.

La posible existencia del sesgo en las estimaciones simples de los retornos hace que estos cálculos simples deban ser vistos con precaución, sobre todo, cuando en estos resultados se basan políticas educativas.

Apéndice. Método de ajuste por sesgo para el caso de escolaridad continua

En este apéndice se presenta una descripción del método de estimación empleado para el sesgo por elección.²⁰

Si la variable S denota los años de escolaridad, la cual toma valores entre 0 y n ; la variable " y " denota el logaritmo natural de los ingresos y la variable (posiblemente un vector de variables) X denota otras variables que determinan los ingresos. La relación estocástica entre los ingresos y las variables explicativas se puede escribir de la siguiente manera:

²⁰ Para más detalles véase Garen (1984).

$$\begin{aligned}
 y &= a_0 + b_0X + e_0 & \text{si } S = 0 \\
 &\cdot \\
 &\cdot \\
 y &= a_n + b_nX + e_n & \text{si } S = n
 \end{aligned}
 \tag{1}$$

donde las diferentes ecuaciones indican que cada nivel de escolaridad afecta de manera distinta a los ingresos, lo cual se ve reflejado en los diferentes parámetros y en los diferentes componentes no observables (los residuales). Cuando n es grande se puede suponer, como aproximación, que S es continua. En este caso las ecuaciones (1) se pueden aproximar por medio de:

$$y = \beta_0 + \beta_1X + \beta_2S + \beta_3X^2 + \beta_4S^2 + \beta_5S \cdot X + (e + \delta \cdot S)$$

o

$$\tag{2}$$

$$y = \beta_0 + \beta_1X + \beta_2S + \beta_3X^2 + \beta_4S^2 + \beta_5S \cdot X + u$$

En esta ecuación de ingresos los términos cuadráticos y de interacción permiten que los ingresos dependan del nivel de escolaridad de manera distinta. La variable e denota el residual y la expresión $\delta \cdot S$ permite que los factores no observables de los individuos afecten de manera distinta, dependiendo del nivel de escolaridad. Las variables aleatorias, e y Φ , se suponen normalmente distribuidas con medias cero y varianzas y covarianzas no restringidas.

El aspecto principal del sesgo por elección propia del individuo consiste en que una de las variables explicativas, el nivel de escolaridad, es una variable endógena, es decir es una variable de elección del individuo. La decisión del nivel de educación del individuo va a depender de los ingresos esperados, denotados por (2), y de otros factores que determinan tanto los costos de educación como su demanda. De esta manera, la ecuación de escolaridad (la cual se obtiene como resultado de optimizar los beneficios netos, esto es ingresos menos costos) se puede escribir como:

$$S = b_0 + b_1R + w$$

$$\tag{3}$$

Donde R representa un vector de variables explicativas, las cuales pueden incluir a algunas de las variables en X . El modelo por estimar consiste en las ecuaciones (2) y (3), lo cual se parece a un modelo triangular. Sin embargo, dada la dependencia del residual en (3) de los parámetros y residuales de (2),²¹ el valor esperado del resi-

²¹ Lo cual es el resultado del proceso de optimización del individuo; en particular, para el presente caso se tiene que $w = -\Phi/2\beta_4$.

dual en (2) no es cero, por lo cual la estimación directa de la ecuación (2) da por resultado estimadores sesgados.

La solución de este problema consiste en obtener una expresión sobre la media del residual en (2) y usar ésta en la estimación. La expresión que se busca es la siguiente:

$$E(e + \varphi \cdot S | w = S - b_0 + b_1 R, S, X) = \\ \{Cov(e, w) / Var(w)\}w + \{Cov(\Phi, w) / Var(w)\}S \cdot w \quad (4)$$

De esta manera, la estimación consistiría en utilizar mínimos cuadrados ordinarios (MCO) en la siguiente ecuación,

$$y - \beta_0 + \beta_1 X + \beta_2 S + \beta_3 X^2 + \beta_4 S^2 + \beta_5 S \cdot X + d_1 w + d_2 w \cdot S + v \quad (5)$$

Como no se conoce w se tiene que utilizar un estimador de éste. Para tal efecto se estima (3) por MCO, con lo cual se obtienen estimadores de w . Una vez obtenido el estimador del residual, se sustituye su valor en (5) y se estima esta última ecuación por MCO. Mediante este procedimiento se tiene que los residuales de la nueva ecuación (5) son heterocedásticos, ya que son función de la variable S . En este caso se logra una mayor eficiencia en la estimación, si se utilizan mínimos cuadrados ponderados.

Referencias bibliográficas

- Becker, G. (1975), *Human capital: A theoretical and empirical analysis, with special reference to education*, National Bureau for Economic Research.
- Blau, D., J. Behrman y B. Wolf (1988), "Schooling and earnings distribution with endogenous labor force participation, marital status and family size", *Economica*, vol. 55, núm. 219, agosto.
- Blaug, M. (1972), *Economía de la educación*, Madrid, Tecnos.
- Carnoy, M. (1967), "Rates of return to schooling in Latin America", *Journal of Human Resources*, vol. 2, núm. 3, pp. 359-374.
- Freeman, R. (1986), "Demand for education", en O. Ashenfelter y R. Layard (eds.), *Handbook of Labor Economics*, vol. 1, Elsevier Science Publishers.
- Garen, J. (1984), "The returns to schooling: a selectivity bias approach with a continuous choice variable", *Econometrica*, vol. 52 núm. 52, pp. 1199-1218.
- Gaston, N. y J. Trejo (1992), "Education attainment and earnings determination in Colombia", *Economic development and cultural change*, vol. 41 núm. 1, pp. 125-139.
- Glewwe, P. (1991), "Schooling, skills, and the returns to government investment in education: An exploration using data from Ghana". *Living Standards Measurement Study*.
- Heckman, J. (1976), "The common structure of statistical models of truncation, sample selec-

- tion and limited dependent variables and a simple estimation for such models”, *Annals of Economic and Social Measurement*, vol. 5, núm. 4, pp. 475-492.
- (1979), “Sample selection bias as a specification error”, *Econometrica*, vol. 47, núm. 1, pp. 153-161.
- Hollenbeck, K. (1993), “Postsecondary education as a triage: returns to academic and technical programs”, *Economics of Education Review*, vol. 12, núm. 3, pp. 213-232.
- Kenny, L., L. Lee, G. Maddala y R. Trost (1979), “Returns to college education: an investigation of selfselection bias based on the proyect talent data”, *International Economic Review*, vol. 20, núm. 3, pp. 775-789.
- King, E. (1990), “Does education pay in the labor market?: The labor force participation, occupation, and earnings of peruvian women”, *Living Standards Measurement Study*, documento de trabajo núm. 67, The World Bank.
- Lee, Lung-Fei (1983), “Generalized econometric models with selectivity”, *Econometrica*, vol. 51 núm. 2, pp. 507-512.
- Lee, Lung-Fei y R. Trost (1978), “Estimation of some limited dependent variable models with applications to housing demand”, *Journal of Econometrics*, vol. 8, pp. 357-382.
- Olsen, R. (1980), “A least squares correction for selectivity bias”. *Econometrica*, vol. 48, núm. 47, pp. 1815-1820.
- Rosen, Sherwin (1977), “Human capital: relations between education and earnings”, en M. Intriligator (ed.), *Frontiers of Quantitative Economics*, vol. 3B, pp.731-753, NorthHolland.
- Rosen, S. (1987), “Self-selection and education”, en Psacharopoulos (ed.), *Economics of Education. Research and Studies*. Oxford, Pergamon Press, pp. 298-300.
- Rosen, Sh. y R. Willis (1979), “Education and selfselection”, *Journal of Political Economy*, núm. 87, (5 parte 2), pp. s7-s36.
- Trost R. y L. Lee (1984), “Technical training and earnings: a polichotomous choice model with selectivity”, *The Review of Economics and Statistics*, vol. 66, núm. 1, pp. 151-156.
- Willis, R.J. (1986), “Wage determinants: a survey a reinterpretation of Human Capital earnings functions”, en, O. Ashenfelter, y R. Layard (eds.), *Handbook of Labor Economics*, vol I, Elsevier Science Publishers.
- Zamudio, A. (1993), “Rendimientos a la inversión educativa: el problema de sesgo por elección: estimaciones para México”, Ponencia presentada en el Segundo Congreso Nacional de Investigación Educativa, Guadalajara, Jal.

Índice

<i>Prólogo</i>	1
<i>Introducción</i>	2
<i>El problema del sesgo por elección</i>	2
<i>Estimaciones para México</i>	6
<i>Observaciones finales</i>	12
<i>Apéndice. Método de ajuste por sesgo para el caso de escolaridad continua</i>	12
<i>Referencias Bibliográficas</i>	14