

Llamas, I. y Garro, N.(1993). Capacitación de los trabajadores mexicanos. *Ensayos*. CIE, Facultad de Economía, Universidad Autónoma de Nuevo León. Vol. XII, No. 2, nov. 1993, Pp. 97-120.

Llamas, I.(1989). *Educación y mercado de trabajo en México*. Universidad Autónoma Metropolitana.

Llamas, I.(1993). Gastos en educación e incorporación al mercado de trabajo de los jóvenes de los hogares pobres de México. *Análisis Económico*. Vol. XI, No. 22, Pp. 115-130.

Llamas, I.(1994). Education and labor markets in developing countries. En *The International Encyclopedia of Education*. Oxford, Pergamon Press.

Llamas, I. y Garro, N.(1991). La inversión en educación. *Ensayos*. CIE, Facultad de Economía, Universidad Autónoma de Nuevo León. Vol. X, No. 2, nov. 1991. Pp. 33-50.

Meléndez, J.(1993). La asignación del trabajo heterogéneo en el mercado laboral y la distribución del ingreso: evidencia para el Área Metropolitana de Monterrey. *Ensayos*. CIE, Facultad de Economía, Universidad Autónoma de Nuevo León. Vol. XII, No. 2, nov. 1993. Pp. 39-98.

Mincer, J.(1974). *Schooling, experience and earnings*. NBER, Nueva York.

Psacharopoulos, G. y Woodhall, M.(1985). *Education for development: an analysis of investment choices*. World Bank, Oxford University Press.

Silos M.(1980). *Los rendimientos de la escolaridad en el Área Metropolitana de Monterrey*. CIE, Facultad de Economía, Universidad Autónoma de Nuevo León.

Secretaría de Trabajo y Previsión Social (STPS)(1986). *Oferta y necesidades de capacitación 1985-1988: técnicos, operarios calificados y semicalificados*. México.

Valdez, F.(1995). *Condiciones del mercado laboral para hombres y mujeres en el Area Metropolitana de Monterrey*. Tesis de Licenciatura. Facultad de Economía, UANL.

Willis, R.(1986). Wage determinants: a survey and reinterpretation of human capital earnings functions. En Ashenfelter, O. & Layard, R. *Handbook of Labor Economics*. Elsevier Science Publishers BV, Vol.1, cap. 10.

RENDIMIENTOS A LA INVERSIÓN EDUCATIVA EL PROBLEMA DE SESGO POR ELECCIÓN: ESTIMACIONES PARA MÉXICO

Andrés Zamudio Carrillo¹

Revisión de la evidencia empírica para México sobre el sesgo por elección o el problema de la endogeneidad de la educación en las estimaciones simples de la tasa de retorno. Se calculan tasas de retorno a la educación con y sin ajuste para llevar a cabo la comparación. Se utiliza la información proveniente de las Encuestas Ingreso-Gasto de los Hogares 1989, de donde se desprende la existencia de un sesgo negativo por selección propia en estimaciones simples de la tasa de retorno a la inversión educativa -10.6% de rendimiento estimado sin ajuste, contra una rentabilidad de 17.4% corrigiendo.

1. Introducción

Se ha señalado que la educación o los años de escolaridad explican una parte importante de la variación observada en los ingresos de los individuos, donde individuos con un mayor nivel de educación obtienen, en promedio, ingresos mayores. Este hecho empírico ha llevado a muchos investigadores a pensar que la educación es una forma de inversión para los individuos². Los costos de la inversión vienen dados por los costos directos de la educación (colegiaturas, inscripciones, material didáctico, etc.), y los costos indirectos o de oportunidad que consisten en los ingresos no ganados por el individuo durante el período de aprendizaje. Los beneficios de la inversión consisten en el diferencial de ingresos que se logra como resultado del mayor nivel educativo. Bajo este enfoque de la educación, esto es como una forma de inversión en capital (capital humano), ha resultado importante el calcular la tasa de retorno a este tipo de inversión. Si los retornos a la inversión en proyectos alternativos son menores a los retornos que se producen en la inversión educativa, entonces quiere decir que la inversión en educación es un proyecto rentable.

El cálculo de los retornos a la educación ha sido muy importante para los investigadores de la educación. Desde el punto de vista privado, una alta tasa de retorno observada significaría que los individuos demandarían más educación, por lo cual sería necesario, para la sociedad, el estar en posibilidad de ofrecer una mayor cantidad de servicios educativos en el futuro. Desde el punto de vista social, un rendimiento alto a la Inversión educativa indicaría que es óptimo, para la sociedad, el transferir recursos a la educación, ya que de este modo se obtendría un ingreso nacional mayor. Por esta razón los retornos a la educación han tenido su importancia en cuanto a la planeación y asignación de recursos se refiere.

Investigaciones sobre los rendimientos a la educación se han producido en gran cantidad a partir de los 60's. La mayoría de los estudios han sido sobre los Estados Unidos, aunque también existe una literatura importante sobre países en desarrollo, y en particular sobre Latinoamérica.

El método para llevar a cabo los cálculos ha variado a través del tiempo. Primero, el método de estimación consistió en el simple cálculo de la tasa interna de retorno a la inversión educativa.

¹ Profesor Investigador de Tiempo Completo del Centro de Investigación y Docencia Económica, CIDE.

² Véase por ejemplo Schultz (1967).

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 altos³. A partir de mediados de los 70's 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 trabajo se revisa la evidencia empírica que hay para México, sobre el sesgo por elección, o el problema de la endogeneidad de la educación, en las estimaciones simples de la tasa 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 Ingreso-Gasto de los Hogares-1989.

2. El problema del sesgo por elección

Considérese un problema simple, estimar la tasa de retorno a la inversión en educación superior. La forma usual para estimar el retorno, ya sea por el método de la tasa interna de retorno o el de la ecuación minceriana, es el comparar las trayectorias de ingresos de individuos con solamente educación media y personas con educación superior. La diferencia entre los ingresos, una vez terminada la educación superior, representa el beneficio del incremento en el nivel de escolaridad. Mientras que el negativo de la diferencia, durante el periodo de aprendizaje, corresponde al costo de oportunidad o indirecto, esto es, el ingreso no ganado.

Al llevar a cabo la comparación se está dejando de lado la posibilidad de que los individuos, o grupo de individuos, no sean comparables, esto es, se está dejando de lado el problema de la heterogeneidad de los individuos. Por ejemplo, el costo de oportunidad de un licenciado no es necesariamente igual al ingreso de un técnico con solamente educación media. Es posible que si el licenciado no hubiera llevado a cabo estudios superiores su ingreso sería menor al del técnico, suponiendo que la persona que escogió ser técnico medio es, por alguna razón, más apto para este tipo de trabajo; del mismo modo, si el técnico hubiera hecho estudios superiores su ingreso no sería el mismo que el recibido por el licenciado. En este caso las variables no observables, como sería el caso de la habilidad, motivación, o diferentes características familiares, tendrían un efecto diferente sobre la formación de ingresos, dependiendo del nivel de escolaridad; o puesto de otra forma, los no-observables explicarían no sólo la formación de los ingresos, sino también el logro escolar.

Si la heterogeneidad descrita es cierta, el costo de oportunidad para el licenciado estaría sobrestimado, por lo cual los retornos resultarían subestimados, esto es, los retornos resultarían sesgados hacia abajo.

La solución a este problema sería el comparar las trayectorias de ingreso, para un individuo típico, cuando se lleva a cabo el proyecto educativo y cuando no se realiza. Sin embargo, este tipo de

³ Véase Gaston y Tenjo (1992).

información no se tiene, ya sea que se tenga información para cuando el individuo llevó a cabo el proyecto educativo, o para cuando no lo hizo, pero no para ambos casos. La información que se tiene es truncada, siendo que esta truncación es el resultado de la elección del individuo. Esto es, sólo se tiene información del individuo una vez que este tomó una decisión.

En una situación como la descrita habría que modelar tanto la formación de ingresos, para los cálculos de los retornos, como la decisión a escolarizarse. De este modo se tendría al nivel de educación como una variable endógena al modelo.

La forma más simple de calcular los retornos es utilizar la llamada ecuación minceriana. Este método consiste en estimar por mínimos cuadrados ordinarios (MCO), la siguiente ecuación de ingresos:

$$\log(Y) = y = a + bS + cE + u$$

donde Y denota el sueldo del individuo (o el ingreso); S denota al nivel de escolaridad, esto es los años de educación; E denota los años de experiencia laboral y u es un error aleatorio. En esta ecuación el parámetro b representa el retorno a la educación. Como se dijo anteriormente, la variable S es una variable endógena, por lo cual se debe especificar una ecuación para esta variable. Una vez especificada la ecuación de escolaridad, el sesgo por elección (o el problema de la endogeneidad de S), se reduce a la covarianza que pudiera existir entre los residuales, o no observables, de ambas ecuaciones. Cuando la covarianza es cero no existe problema de sesgo por elección. Si la covarianza es positiva, entonces una persona que presenta de manera inesperada un nivel alto de escolaridad, es una persona que podría presentar un ingreso mayor al esperado dada la ecuación de ingresos.

La importancia de este tipo de sesgo ha sido discutida por Z. Griliches (1977). La decisión por educarse es, en muchos casos, una decisión familiar. Por esta razón es de esperarse que los factores no observados de la decisión de educarse no estén correlacionados con los no observados de la ecuación de ingresos; en el primer caso los no-observados de la decisión de educarse estarían representados por las características familiares no observables, mientras que los no-observables de la ecuación de ingresos estarían representados por ciertas características personales. A pesar de esto, es de esperarse que los no-observables en ambas ecuaciones estén de alguna manera correlacionados, ya sea que se esté hablando de características personales, como habilidad o motivación; o familiares, como ingreso de las familias.

La forma tradicional de tratar este problema de sesgo es utilizar una ecuación de escolaridad dicotómica, esto es, una ecuación donde sólo se tienen dos posibilidades: llevar a cabo un programa particular de educación o no hacerlo. En este trabajo se utiliza el método propuesto por J. Garen (1984), el cual permite que el nivel de escolaridad sea una variable continua⁴. El método de estimación se describe en el apéndice, este consiste básicamente en la estimación de las siguientes ecuaciones:

$$y = \beta_0 + \beta_1 X + \beta_2 S + \beta_3 X^2 + \beta_4 S^2 + \beta_5 S \cdot X + (e + \phi \cdot S)$$

$$S = b_0 + b_1 R + w$$

⁴ Aunque un problema con el método utilizado es que el nivel de escolaridad no queda acotado.

Donde la variable X denote al vector de variables que determinan los ingresos, como pueden ser la experiencia, horas trabajadas, etc; R es un vector de variables que determinan la decisión a escolarizarse, como las características familiares, zona de origen, etc; $(e + \phi \cdot S)$ y w representan los residuales. El término $\phi \cdot S$ permite que los no-observables afecten de manera distinta, dependiendo del nivel de educación, la formación de ingresos. Dado el proceso de selección del individuo los residuales de ambas ecuaciones se encuentran correlacionados⁵, por lo cual el valor esperado del residual $(e + \phi \cdot S)$, dado que se escogió el nivel de escolaridad S , no es cero. En este caso, los estimadores mínimo cuadráticos de los parámetros de la primera ecuación son sesgados, por esta razón este sistema de dos ecuaciones no podría estimarse de manera recursiva. Como se discute en el apéndice, la solución propuesta por Garen (1984), y la cual es similar al caso cuando se tiene una ecuación de escolaridad dicotómica⁶, es incluir el sesgo del residual en la primer ecuación, y estimar esta por MCO. De esta manera se tendría que estimar primeramente la ecuación de escolaridad. 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 a estimar sería,

$$y = \beta_0 + \beta_1 X + \beta_2 S + \beta_3 X^2 + \beta_4 S^2 + \beta_5 S \cdot X + d_1 w + d_2 W \cdot S + V$$

3. Estimaciones para México

Para realizar la estimación se utilizó información de La Encuesta Nacional de Ingresos-Gastos de los Hogares-1989 (ENIGH89).

En esta encuesta se encuentran datos sobre características de los individuos como nivel de educación, edad, tipo de trabajo, rama de actividad y diferentes tipos de ingresos. Con esta información es posible llevar a cabo estimaciones sobre los retornos a la educación para toda la muestra; sin embargo, existe el problema de que la ENIGH89 no aporta datos sobre las características familiares para todos los individuos, cuando 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 ENIGH89.

Con el objeto de incluir en la estimación datos sobre las características de los padres y teniendo en cuenta que la información que proporciona INEGI es referente a las familias, se incluyeron, en la muestra solamente datos sobre los hijos que todavía vivían con sus padres⁷. De hecho, como es muy importante en la estimación el ingreso familiar (o en este caso el ingreso del jefe familiar), se consideró solamente a los casos en que el ingreso familiar fuera positivo. También fue necesario el acotar la edad de los individuos, por lo que solamente se consideraron individuos entre los 15 y 40 años. Finalmente, solamente se consideraron a los individuos que, estando trabajando, reportaron un sueldo positivo.

Dadas las restricciones impuestas sobre la muestra por trabajar, los resultados de esta estimación no se pueden generalizar a toda la población. La utilidad de los resultados es simplemente verificar, para una

⁵ Véase Garen (1984).

⁶ Véase Heckman (1979).

⁷ Esta restricción sobre la muestra seleccionada tiene obvias consecuencias sobre la generalidad de los resultados obtenidos.

cierta muestra, el posible sesgo por elección que pueda estar presente en las estimaciones simples de las tasas de retorno a la educación.

La muestra resultante tiene un tamaño de 3100 individuos. De éstos, 1154 son mujeres y 1946 hombres. La muestra utilizada cubre a toda la república mexicana, donde el 72% de los casos corresponden a la población urbana⁸ y 28% a la rural. Las variables utilizadas en la estimación son las siguientes. LW representa al logaritmo natural del sueldo del individuo del mes anterior. $HRSEMT$ representa las horas trabajadas durante una semana y LHT es el logaritmo natural de esta variable. 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 familiar. $INGRESOJ$ representa al ingreso del jefe familiar. $ZONA$ es una variable dummy la cual toma el valor unitario cuando el individuo reside en una zona urbana, y cero cuando lo es de la zona rural. $TAMH$ es el tamaño del hogar. En el cuadro 1 se presentan algunas estadísticas sobre las variables utilizadas.

Cuadro 1. Estadísticas de las variables utilizadas

Variable	Media	Desviación Estándar
Sueldo	357570.8	275942.1
HRSEMT	43.9	11.7
S (Total)	8.7	3.8
S (Mujeres)	9.7	3.6
S (Hombres)	8.2	3.8
EDAD (Total)	22.3	5.3
EDAD (Mujeres)	23.2	5.6
EDAD (Hombres)	22.2	5.0
TAMH	7.1	2.6
SJ	4.2	3.5

La estimación de los retornos se hace en dos partes. Primero 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 residencia⁹. Después se utilizan los residuales de esta última ecuación para estimar los retornos con la ecuación de ingresos. En el cuadro 2 se presentan los resultados de la estimación, por MCO, de la ecuación de escolaridad para tres casos: el total de la población, para las mujeres y para los hombres. En esta tabla se puede ver que los parámetros estimados son significativos y que corresponden al signo correcto o esperado. Por ejemplo, se esperaría que un individuo 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 $ZONA$

⁸ El INEGI divide a las zonas entre alta y baja densidad.

⁹ Este tipo de variables explicativas para la ecuación de escolaridad ya se han utilizado en otra parte (véase Gaston y Trenjo, 1992).