

Proyecto econometría aplicada: Retornos a la educación

Sergio Sarmiento

September 2025

1 Introducción

A lo largo de los últimos años la economía como ciencia ha empezado a tomar en cuenta nuevas perspectivas, tales ideas se han alejado de los análisis clásicos de la economía como PIB, economía de la empresa y de los consumidores. Estos cambios han permitido que la ciencia económica empiece a ver nuevos temas de análisis de igual importancia que los anteriores. Un ejemplo de esto, aunque no el único, son los temas relacionados al capital humano. Uno de los principales economistas que han sentado las bases sobre el análisis del capital humano fue Gary Stanley Becker, el cuál en su artículo titulado *Investment in Human Capital: A theoretical analysis* plantea los beneficios de invertir en recursos intangibles, tales como conocimiento y habilidades. De esta forma Becker cambia el paradigma respecto a que las diferencias salariales ya no son necesariamente exclusivas de diferencias en capital físico, como fuerza o resistencia, sino que hay mecanismos intangibles que hacen que las personas puedan aspirar a un mayor salario si invierten en educación, salud o entrenamiento. Para lo que atañe a este texto, retornos a la educación, Becker entiende los retornos a la educación como el incremento en los ingresos futuros que obtiene un individuo gracias a las mejoras en su productividad, una vez que se descuentan los costos (monetarios y de oportunidad) de adquirir esa educación.

Dicho lo anterior, varios investigadores subsecuentes han continuado con la línea de investigación del capital humano, ejemplo de lo anterior son Kathryn Wilson y David Card, los cuales en sus textos *The determinants of educational attainment* y *Estimating the return to schooling*, respectivamente, han intentado medir los efectos de invertir en educación, principalmente el efecto en el salario futuro. Wilson define los retornos a la educación como los incentivos económicos esperados que influyen en la decisión de continuar estudiando. Los jóvenes eligen su nivel de educación considerando los beneficios en ingresos futuros, al igual que otros factores. Por otro lado, Card entiende los retornos a la educación como el efecto causal de la escolaridad sobre los ingresos laborales. Card reconoce que estos retornos pueden variar entre individuos debido a diferencias en

costos y habilidades. Su investigación empírica ha demostrado que, especialmente en los grupos marginales, los retornos tienden a ser altos, reflejando su mayor costo marginal de estudiar. De esta forma, desde una perspectiva más simple, los retornos a la educación pueden ser vistos como el costo de oportunidad en el beneficio futuro en el ingreso de continuar invirtiendo en educación.

2 Modelo

El presente texto busca continuar con los análisis anteriores al utilizar un modelo econométrico con el fin de medir los retornos a la educación, en particular en México. Para ello, se utilizan datos de la EMOVI 2015 (Encuesta de Movilidad Social 2015), encuesta realizada por el Centro de Estudios Económicos (CEE) del Colegio de México. Por un lado, a continuación se presenta el modelo econométrico empleado. El modelo está estimado por Mínimos Cuadrados Ordinarios (MCO):

$$\ln(\text{ingreso}_i) = \alpha + \beta \text{educ}_i + X_i' \delta + u_i$$

El modelo anterior utiliza el logaritmo natural del ingreso del individuo i como variable dependiente, los años de educación del individuo i como variable independiente y un vector de controles X_i' , en los que se encuentran variables como edad, sexo, la clave del estado y el color de piel del individuo i .

Por otro lado, se utilizan variables instrumentales, como la educación de los padres o la educación del padre, con el fin de comparar con la estimación dada por MCO, más adelante se compararán y explicarán mejor los modelos utilizados. El modelo estimado por VI está definido de la siguiente manera:

Primera etapa:

$$\text{educ}_i = \pi_0 + \pi_1 \text{educ padres}_i + X_i' \delta + v_i$$

Segunda etapa:

$$\ln(\text{ingreso}_i) = \alpha + \beta \hat{\text{educ}}_i + X_i' \delta + u_i$$

En el modelo anterior se mantienen las variables de interés y los controles usados en el modelo por MCO, con la diferencia de las variables instrumentales que es la educación de ambos padres o la educación del padre. A partir de los dos modelos descritos anteriormente, podremos estimar la relación promedio de la educación en los salarios.

3 Discusión de los modelos

Usar MCO sería preferible en muchas ocasiones ya que permite obtener un estimador del efecto promedio de un año adicional de escolaridad sobre los ingresos laborales. Es sencillo de implementar, eficiente bajo supuestos y en la práctica suele arrojar estimaciones con buena precisión. Sin embargo, un posible problema es que la educación puede estar correlacionada con el término de error, lo que generaría sesgo de endogeneidad. Esto ocurriría, por ejemplo, si las personas con mayor habilidad no observada tienden a estudiar más y también a obtener mayores ingresos, de manera que MCO sobreestimaría el retorno a la educación.

De esta forma, el uso de Variables Instrumentales ofrece una alternativa para corregir esta endogeneidad. Un buen instrumento debe estar correlacionado con la educación del individuo, pero no debe afectar directamente el ingreso salvo a través de la educación, lo que haría que la relación de la educación con el ingreso sea exógena. En este caso, la educación de los padres puede considerarse como un posible instrumento, está relacionada con la probabilidad de que los hijos acumulen más años de educación, pero no debería influir directamente en el salario del hijo una vez que controlamos por la propia escolaridad y otras características. Usar VI permite entonces identificar un efecto más cercano al causal, especialmente si MCO sobreestima el impacto de la educación.

4 Datos

Los datos usados en este análisis son, como fué expuesto anteriormente, los dados por la EMOVI 2015, encuesta realizada por el Colegio de México, con el fin de analizar la movilidad social. El subconjunto de la población usado es el de los adultos. En este módulo de adultos podemos encontrar varias variables relevantes para el análisis, las cuales son las utilizadas, por ejemplo: ingreso de la persona encuestada, sexo, color de piel, localidad en donde viven, nivel de educación, último año de educación cursado y nivel de escolaridad de los padres.

En el apéndice del presente texto se muestran las gráficas relevantes para el análisis. Por un lado, la gráfica 1 muestra la relación entre los años de educación y el logaritmo del ingreso mensual, donde se observa una ligera pendiente positiva en la recta de regresión. Aunque hay una gran dispersión en los datos, la tendencia indica que más años de educación se asocian con un mayor ingreso promedio. La gráfica 2 muestra la distribución del ingreso mensual, donde se observa una fuerte concentración de individuos con ingresos bajos. A medida que el ingreso aumenta, la densidad disminuye rápidamente y se mantiene cercana a cero en los niveles altos de ingreso. Es por esta distribución del ingreso que se recorta la distribución al eliminar los valores extremos; para este análisis se utilizan únicamente las observaciones entre el percentil 1 y el percentil 99. Con

esto se logra trabajar con una muestra menos afectada por outliers, lo que facilita la estimación.

5 Resultados

En la siguiente sección se presentan los resultados de estimar los retornos a la educación por MCO y por variables instrumentales, el instrumento usado en el modelo (2) es la educación del padre y la madre del individuo i ; mientras que en el modelo (3), el instrumento es solamente la educación del padre. En el apéndice se encuentran los resultados de la primera etapa de los modelos (VI 2) y (VI 3). Los resultados muestran diferencias claras entre MCO y VI. En la estimación por MCO (columna 1), los años de educación tienen un efecto positivo y significativo al 5%. Por lo que un año adicional de escolaridad se asocia con un aumento aproximado de 2% en el ingreso mensual. Sin embargo, al usar variables instrumentales, educación de ambos padres (columna 2) o solo educación del padre (columna 3), el coeficiente de la educación aumenta en magnitud, pero los errores estándar crecen y la significancia estadística desaparece. Esto sugiere que, aunque los IV corrigen posibles sesgos de endogeneidad, los instrumentos son débiles y generan estimaciones imprecisas, por lo que la evidencia a favor de retornos significativos bajo estos instrumentos es limitada.

Table 1: Retornos a la educación: MCO vs VI (variantes y misma muestra)

| | (1) MCO | (2) VI (Educ padres) | (3) VI (Educ padre) |
|--------------------------|--------------------|-------------------------|------------------------|
| Años de educación | 0.020** (0.010) | 0.277 (0.232) | 0.378 (0.643) |
| Experiencia | -0.001 (0.022) | 0.182 (0.145) | 0.247 (0.421) |
| Experiencia ² | 0.000 (0.000) | -0.001 (0.001) | -0.002 (0.003) |
| Observaciones | 1,599 | 979 | 1,154 |
| R ² | 0.085 | 0.945 | 0.897 |

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Diagnósticos de IV

Los resultados de diagnóstico sugieren lo siguiente: la prueba de endogeneidad de Hausman no rechaza la hipótesis nula ($p - value = 0.19$), por lo que MCO parece consistente. La prueba de sobreidentificación de Sargan tampoco rechaza la validez de los instrumentos ($p=0.83$). Sin embargo, la prueba de instrumentos débiles (*first stage regression 2*) muestra un $F = 1.73$ y un R^2 parcial de 0.0042, esto señala un instrumento debil. En este caso, aunque los IV son válidos producen estimaciones imprecisas y se prefiere MCO.

Para la especificación con un solo instrumento (educación del padre), la prueba de endogeneidad de Hausman no rechaza la hipótesis nula de exogeneidad ($p = 0.41$), por lo que MCO parece consistente. Aunque, la prueba de instrumentos débiles (*first stage regression 3*) indica falta de fuerza: $F = 0.65$ y R^2 parcial = 0.0007, por debajo de un $F \approx 10$, por lo que parece ser un instrumento débil. Dado que no hay evidencia de endogeneidad, el IV es débil y produce estimaciones imprecisas, por lo que es recomendable usar MCO.

Apéndice

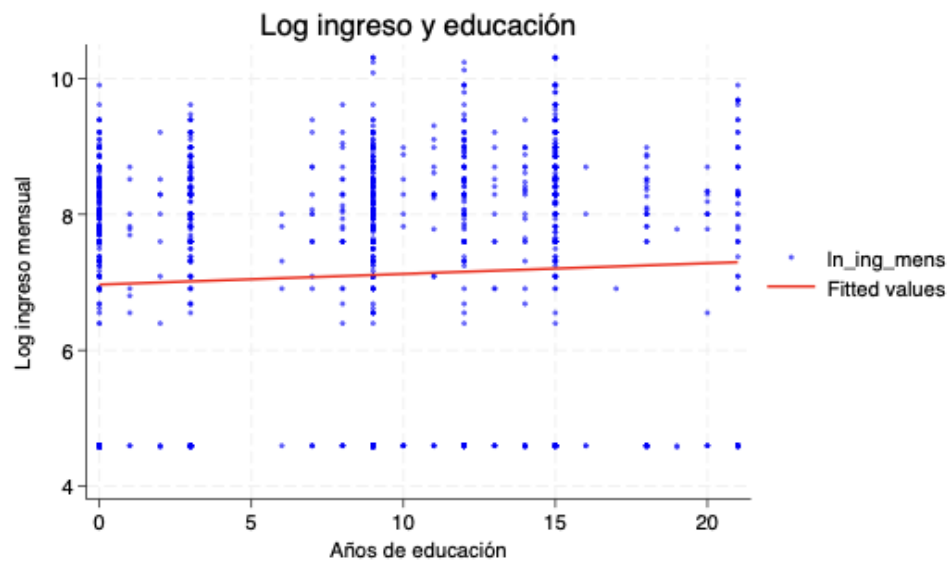


Figure 1: Gráfica de dispersión, años de educación e ingreso

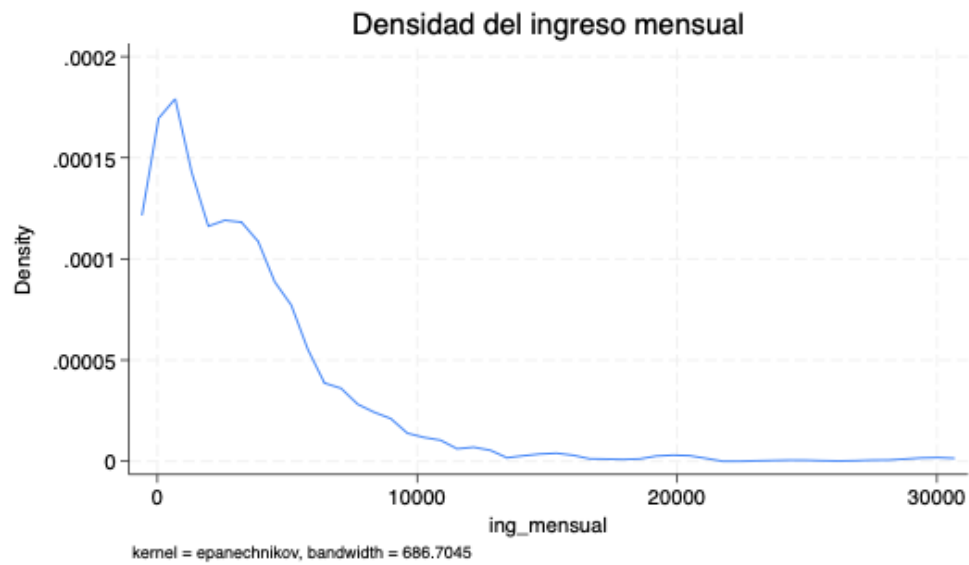


Figure 2: Distribución del ingreso mensual

Table 2: First-stage regression summary statistics

| Variable | R-sq. | Adjusted R-sq. | Partial R-sq. | Robust F(2,941) | Prob > F |
|-----------|--------|----------------|---------------|-----------------|----------|
| anios_edu | 0.4485 | 0.4269 | 0.0042 | 1.73072 | 0.1777 |

Table 3: First-stage regression summary statistics

| Variable | R-sq. | Adjusted R-sq. | Partial R-sq. | Robust F(1,1117) | Prob > F |
|-----------|--------|----------------|---------------|------------------|----------|
| anios_edu | 0.4655 | 0.4483 | 0.0007 | 0.646925 | 0.4214 |