



UNIVERSIDAD
NACIONAL
DE LA PLATA

MAESTRÍA EN ECONOMÍA

Trabajo Práctico

*Coloma Conte-Grand, Carolina
Condori, Brayan Alexis
Deniard, Agustín*

Profesor: Javier Alejo

Materia: Regresiones por Cuantiles

7 de octubre de 2024

Ejercicio 1

El Cuadro 1 muestra las estadísticas descriptivas de las variables explicativas de la ecuación de Mincer. La muestra total incluye a 7903 individuos entre 18 y 65 años que residen en el Gran Buenos Aires. Para los ingresos horarios de la ocupación principal, observamos una tendencia creciente en la mediana a medida que aumenta la edad. En todos los grupos de edad, existe una brecha entre los individuos del primer y el último decil, siendo más pronunciado en el grupo de 65 años, aunque este grupo tiene una muestra pequeña (80 obs.). Excluyendo este grupo, el rango de personas entre 41 y 64 años muestra la mayor disparidad, siendo el último decil (91,0) casi seis veces superior al primero (15,5). Respecto a los años de educación recibidos, se observa un patrón similar entre el primer y último decil, con la mayor brecha en los grupos etarios de 25-40 y 41-64 años, excluyendo a los de 65 años. Sin embargo, los años de educación recibidos por individuo mediano en cada grupo etario es de 12 años, excepto para el grupo de 65 años, donde es 11 años. Finalmente, la muestra incluye un 42,2% de mujeres y un 26,6% de residentes en CABA.

Cuadro 1: Estadísticas descriptivas

| | N | Media | SD | Min | Max | Deciles | | | | | | | | | |
|--|------|-------|-------|------|--------|---------|------|------|------|------|------|------|------|-------|--|
| | | | | | | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | |
| <i>Ingreso horario en la ocupación principal</i> | | | | | | | | | | | | | | | |
| Grupo de edad: [15,24] | 998 | 34.5 | 28.8 | 0.6 | 515.5 | 13.5 | 18.2 | 21.9 | 25.8 | 30.2 | 33.4 | 37.6 | 43.3 | 60.1 | |
| Grupo de edad: [25,40] | 3339 | 47.6 | 46.5 | 0.6 | 1804.1 | 17.2 | 24.3 | 29.1 | 34.4 | 39.4 | 45.1 | 52.0 | 64.4 | 85.0 | |
| Grupo de edad: [41,64] | 3486 | 49.3 | 44.5 | 0.2 | 827.4 | 15.5 | 22.1 | 27.8 | 33.1 | 39.4 | 45.5 | 53.5 | 64.7 | 91.0 | |
| Grupo de edad: [65+] | 80 | 58.5 | 89.0 | 4.6 | 618.6 | 11.4 | 15.4 | 24.2 | 28.8 | 36.2 | 42.2 | 48.5 | 68.7 | 113.5 | |
| <i>Años de educación</i> | | | | | | | | | | | | | | | |
| Grupo de edad: [15,24] | 998 | 11.2 | 2.6 | 2.0 | 17.0 | 7.0 | 9.0 | 10.0 | 12.0 | 12.0 | 12.0 | 12.0 | 13.0 | 14.0 | |
| Grupo de edad: [25,40] | 3339 | 12.2 | 3.4 | 0.0 | 20.0 | 7.0 | 9.0 | 11.0 | 12.0 | 12.0 | 12.0 | 15.0 | 15.0 | 17.0 | |
| Grupo de edad: [41,64] | 3486 | 11.2 | 4.0 | 0.0 | 20.0 | 7.0 | 7.0 | 7.0 | 10.0 | 12.0 | 12.0 | 14.0 | 15.0 | 17.0 | |
| Grupo de edad: [65+] | 80 | 10.7 | 4.8 | 0.0 | 19.0 | 4.5 | 7.0 | 7.0 | 7.0 | 11.0 | 12.0 | 15.0 | 16.5 | 17.0 | |
| <i>Edad</i> | | | | | | | | | | | | | | | |
| Grupo de edad: [15,24] | 998 | 21.6 | 1.8 | 18.0 | 24.0 | 19.0 | 20.0 | 21.0 | 21.0 | 22.0 | 22.0 | 23.0 | 23.0 | 24.0 | |
| Grupo de edad: [25,40] | 3339 | 32.6 | 4.6 | 25.0 | 40.0 | 26.0 | 28.0 | 30.0 | 31.0 | 33.0 | 34.0 | 36.0 | 37.0 | 39.0 | |
| Grupo de edad: [41,64] | 3486 | 51.1 | 6.8 | 41.0 | 64.0 | 42.0 | 44.0 | 46.0 | 48.0 | 51.0 | 53.0 | 55.0 | 58.0 | 61.0 | |
| Grupo de edad: [65+] | 80 | 65.0 | 0.0 | 65.0 | 65.0 | 65.0 | 65.0 | 65.0 | 65.0 | 65.0 | 65.0 | 65.0 | 65.0 | 65.0 | |
| Mujer (=1) | 7903 | 0.422 | 0.494 | - | - | - | - | - | - | - | - | - | - | - | |
| Reside en CABA (=1) | 7903 | 0.266 | 0.442 | - | - | - | - | - | - | - | - | - | - | - | |

El Cuadro 2 muestra los resultados de la estimación de la ecuación de Mincer mediante MCO. Dado que el efecto marginal no es constante para las variables de edad y educación, tomamos como referencia a un individuo de 30 años de edad y 12 años de educación formal. Un incremento de un año en la edad se asocia con un aumento promedio del 1,36% en el ingreso laboral horario, mientras que un incremento de un año de educación se asocia con un aumento promedio del 8,50%. Por otro lado, ser mujer se asocia con una reducción promedio del 13,27% en el ingreso laboral horario

en comparación con ser hombre. Por último, residir en CABA se asocia con un aumento promedio del 14,53 % en el ingreso laboral horario, respecto a quienes viven en el Conurbano Bonaerense. Todos los resultados descritos son significativos al 1 %.

Cuadro 2: Estimación de la ecuación de Mincer por MCO y QR

| | MCO | QR | | | | |
|-------------------------------|----------------------|---------------------------|----------------------------|----------------------------|----------------------------|--------------------------|
| | | Q05 | Q25 | Q50 | Q75 | Q95 |
| Años de educación | -0.039*** (0.013) | 0.0328 (0.0367) | -0.0197 (0.0153) | -0.0257** (0.0126) | -0.0579*** (0.0166) | -0.0623* (0.0353) |
| Años de educación al cuadrado | 0.005*** (0.001) | 0.00270* (0.00162) | 0.00444*** (0.000632) | 0.00442*** (0.000557) | 0.00596*** (0.000713) | 0.00606*** (0.00140) |
| Edad | 0.035*** (0.004) | 0.0274** (0.0110) | 0.0445*** (0.00580) | 0.0389*** (0.00380) | 0.0348*** (0.00453) | 0.0228*** (0.00866) |
| Edad al cuadrado | -0.000*** (0.000) | -0.000320** (0.000132) | -0.000506*** (7.45e-05) | -0.000398*** (4.68e-05) | -0.000333*** (5.47e-05) | -0.000182* (0.000105) |
| Mujer (=1) | -0.142*** (0.015) | -0.323*** (0.0499) | -0.116*** (0.0180) | -0.125*** (0.0156) | -0.129*** (0.0157) | -0.107*** (0.0321) |
| Reside en CABA (=1) | 0.136*** (0.017) | 0.311*** (0.0601) | 0.117*** (0.0219) | 0.123*** (0.0151) | 0.0954*** (0.0178) | 0.139*** (0.0325) |
| Constante | 2.534*** (0.108) | 1.276*** (0.325) | 1.972*** (0.135) | 2.454*** (0.0990) | 2.985*** (0.143) | 3.750*** (0.277) |
| Observaciones | 7,903 | 7,903 | 7,903 | 7,903 | 7,903 | 7,903 |

Errores estándar robusto en paréntesis. *** p<0.01, ** p<0.05, * p<0.1. Error estándar en QR por Bootstrap: 500 repeticiones.

Ejercicio 2

El Cuadro 2 presenta las estimaciones de los coeficientes del modelo para el τ -ésimo cuantil condicional de la ecuación de Mincer:

$$Q_{\tau}(y|x) = x_i^T \beta(\tau)$$

en donde $\tau \in \{0,05, 0,25, 0,50, 0,75, 0,95\}$, y es el logaritmo natural del ingreso laboral por hora y x es el vector de variables explicativas para esta especificación de la ecuación de Mincer, conformada por los años de educación (y su cuadrado), la edad (y su cuadrado), el sexo y la ubicación geográfica. Cabe destacar que la interpretación de los efectos parciales no es la misma que en el modelo para la media condicional estimado por OLS en el punto anterior:

- Un año adicional de educación para una persona con 12 años de educación formal se traduce, *ceteris paribus*, en un aumento del 9,76 % en el ingreso laboral horario del cuantil 0,05; del 8,69 % para el cuantil 0,25; del 8,04 % para el cuantil 0,50; del 8,51 % para el cuantil 0,75; y del 8,31 % para el cuantil 0,95. En todos los casos, el coeficiente que acompaña al término cuadrático de la variable *educ* es significativo al 10 % (y para algunos cuantiles también lo es al 1 %), mientras que el coeficiente que acompaña al término lineal de esta variable no es significativo para algunos cuantiles (0,05 y 0,25) y presenta distintos niveles de significatividad para otros.

- Un año adicional de edad para una persona de 30 años se traduce, *ceteris paribus*, en un aumento del 0,82 % en el ingreso laboral horario del cuantil 0,05; del 1,41 % para el cuantil 0,25; del 1,5 % para el cuantil 0,50; del 1,48 % para el cuantil 0,75; y del 1,19 % para el cuantil 0,95. En todos los casos, el coeficiente que acompaña al término lineal de la variable *edad* es significativo al 5 % (y para algunos cuantiles también lo es al 1 %), mientras que el coeficiente que acompaña al término cuadrático de esta variable lo es al 10 % (y para algunos cuantiles también lo es al 5 % y al 1 %).
- El ingreso laboral horario del cuantil 0,05 es un 27,6 % menor para las mujeres en comparación con los hombres; un 10,95 % para el cuantil 0,25; un 11,75 % para el cuantil 0,50; un 12,1 % para el cuantil 0,75; y un 10,15 % para el cuantil 0,95. En todos los casos, el coeficiente que acompaña a la variable *female* es significativo al 1 %.
- El ingreso laboral horario del cuantil 0,05 es un 36,48 % mayor para quienes residen en Capital Federal en comparación con aquellos que viven en el conurbano bonaerense; un 12,41 % para el cuantil 0,25; un 13,09 % para el cuantil 0,50; un 10 % para el cuantil 0,75; y un 14,91 % para el cuantil 0,95. En todos los casos, el coeficiente que acompaña a la variable *caba* es significativo al 1 %.

Ejercicio 3

Realizamos un test para estudiar si el efecto parcial de los años de educación sobre el salario es el mismo para los cuantiles condicionales 0,05 y 0,95. Encontramos que el p-valor asociado es de 0,0452, lo cual resulta significativo en un 0,05 y nos permite rechazar para este nivel la hipótesis nula de que el efecto parcial de la educación sobre el salario es el mismo en ambos cuantiles.

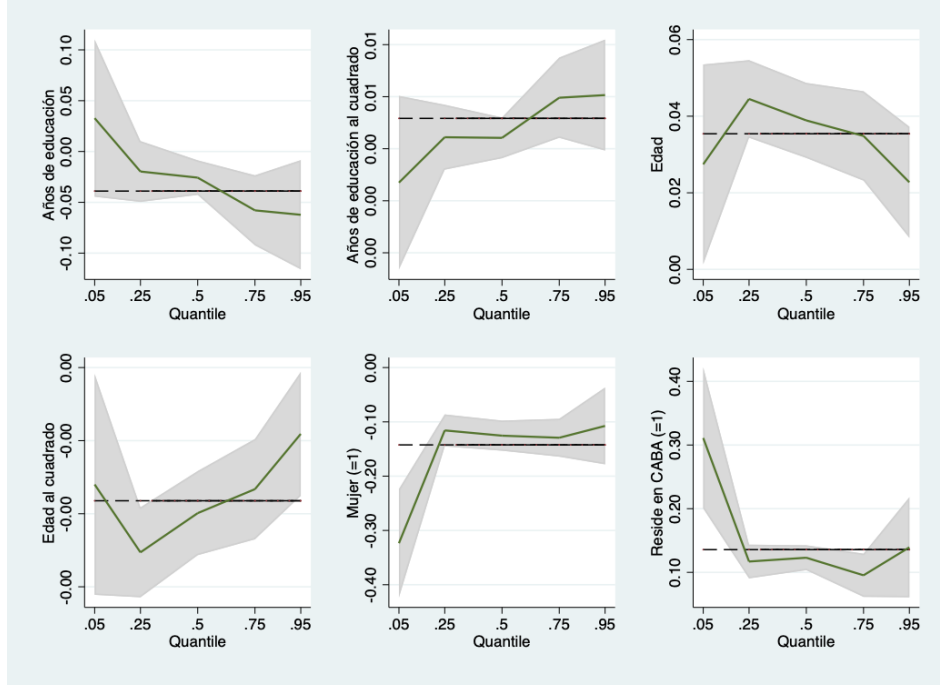
A su vez, realizamos un test bajo la hipótesis nula de que los coeficientes del modelo (a excepción de la ordenada al origen) no difieren entre cuantiles condicionales. Encontramos que el p-valor asociado es casi nulo, por lo que podemos rechazar la hipótesis nula para un nivel de 0,0001.

Este test es útil para estudiar la relevancia del modelo ya que muestra si existe o no información adicional proveniente de las regresiones por cuantiles. Si la hipótesis nula no fuera rechazada, esto implicaría que los efectos parciales son iguales sin importar el cuantil, por lo que no habría diferencia entre la información proporcionada por una estimación cuantílica y la obtenida en una regresión por MCO.

Ejercicio 4

Graficamos los coeficientes de la regresión en función a los cuantiles τ estimados. Los resultados pueden verse en la Figura 1. Encontramos que, en general, todos difieren respecto de la estimación de MCO, aunque esta recta suele encontrarse contenida en la mayoría de los intervalos de confianza estimados por bootstrap¹.

Figura 1: Coeficientes de la regresión respecto al cuantil τ



El método de regresiones por cuantiles se considera uno semiparamétrico. Tiene un componente paramétrico ya que modela a los cuantiles condicionales de la variable dependiente suponiendo una estructura lineal $Q_y(\tau|X) = X\beta(\tau)$. A su vez, cuenta con un componente no paramétrico en que no asume una distribución específica para los errores estándares y estima los cuantiles condicionales a partir de la forma de los datos. Por este motivo, se dice que es un método *distribution free*.

¹El gráfico muestra la forma funcional de $\beta(\tau)$ para cada uno de los coeficientes estimados, sin embargo, debemos ser cuidadosos al interpretar los gráficos de las variables edad y educación, dado que son no lineales

Ejercicio 5

Cuadro 3: Residuos de la estimación de la ecuación de Mincer por QR

| | u(25) | u(50) | u(75) |
|----------|-----------------|-----------------|-----------------|
| Negativo | 1.972 (24,95 %) | 3.946 (49,93 %) | 5.924 (74,96 %) |
| Nulo | 8 (0,10 %) | 8 (0,10 %) | 7 (0,09 %) |
| Positivo | 5.923 (74,95 %) | 3.949 (49,97 %) | 1.972 (24,95 %) |
| Total | 7.903 (100 %) | 7.903 (100 %) | 7.903 (100 %) |

El Cuadro 3 presenta la frecuencia absoluta (relativa) de residuos negativos, nulos y positivos que surgen del modelo para el τ -ésimo cuantil condicional de la ecuación de Mincer, con $\tau \in \{0,25, 0,50, 0,75\}$. Podemos ver que el cómputo de la diferencia entre los valores observados para el ingreso laboral horario y los valores predichos por el modelo para el τ -ésimo cuantil condicional da como resultado en los 3 casos que aproximadamente un τ % de las observaciones cuentan con un residuo negativo, mientras que aproximadamente un $(1 - \tau)$ % de las observaciones cuentan con un residuo positivo.

Sin embargo, encontramos una discrepancia correspondiente a las observaciones para las cuales el residuo es nulo. Como regla general, la frecuencia absoluta debería ser exactamente igual a la cantidad de parámetros a estimar en nuestra especificación de la ecuación de Mincer pero, en este caso, los residuos nulos correspondientes a la estimación del modelo para los cuantiles condicionales 0,25 y 0,5 son 8 en un caso donde contamos con 7 parámetros a estimar. Esto se debe a que, en ambos casos, tenemos dos observaciones cuyo residuo es igual a 0 con el mismo valor para el ingreso laboral horario. Como esto no sucede en la estimación del modelo para el cuantil condicional 0,75, la cantidad de residuos nulos sí coincide con la cantidad de parámetros a estimar.