

EDUCACIÓN Y SALARIOS: UN ANÁLISIS DE SENSIBILIDAD^{o^}

EDUCATION AND WAGES: A SENSITIVITY ANALYSIS

*Julián Martínez-Iriarte**

enviado: 22 abril 2021 - aceptado: 30 junio 2021

Resumen

Este trabajo propone un análisis de sensibilidad para analizar la relación entre educación y salarios. Para diferentes políticas que incrementan la proporción de graduados universitarios, derivamos cotas para los cuantiles de la distribución contra factual de salarios. Luego usamos esas cotas para cuantificar la robustez de distintas conclusiones de interés. Nuestro análisis empírico sugiere que una política de educación dirigida a individuos en los cuantiles más bajos de la distribución de salarios puede contribuir a reducir la desigualdad aun en presencia de un considerable sesgo de selección.

Palabras clave: análisis de sensibilidad, cuantiles, efectos de tratamiento, desigualdad.

Códigos JEL: C1, C31, C54, J3, I26.

Abstract

This paper proposes a sensitivity analysis to analyze the relationship between education and wages. For different policies that increase the proportion of university graduates, we derive bounds for the quantiles of the counterfactual distribution of

^o El presente trabajo está basado en el tercer capítulo de mi tesis doctoral “Sensitivity Analysis in Unconditional Quantile Effects” que ganó el primer premio en IX Congreso Nacional de Estudiantes de Posgrado en Economía, Argentina. Esta versión contiene una nueva aplicación empírica.

[^] Martínez-Iriarte, J. (2022). Educación y salarios: un análisis de sensibilidad, *Estudios económicos*, 39(78), pp. 5-31. DOI: <https://doi.org/10.52292/j.estudecon.2022.2756>

* Departamento de Economía, Universidad de California, Santa Cruz, Estados Unidos. ORCID: <https://orcid.org/0000-0003-2639-3909>. Correo electrónico: jmart425@ucsc.edu

wages. We then use these bounds to quantify the robustness of various conclusions of interest. Our empirical analysis suggests that an education policy targeting individuals in the lower quantiles of the wage distribution can contribute to reducing inequality even in the presence of considerable selection bias.

Keywords: sensitivity analysis, quantiles, treatment effects, inequality.

JEL Codes: C1, C31, C54, J3, I26.

I. INTRODUCCIÓN

La evaluación de políticas contrafactuales es de gran importancia en economía. Preguntas como “¿cuál es el efecto de un aumento en los años de educación de los trabajadores?” o “¿cómo impacta un aumento en la proporción de trabajadores sindicalizados en los salarios?” son de gran relevancia práctica. Usualmente, para poder identificar los efectos de una política contrafactual, es necesario hacer supuestos fuertes sobre la interacción entre las variables observadas e inobservadas. Por ejemplo, si estamos pensando en el efecto de una política que incrementaría los años de experiencia de ciertos trabajadores, tenemos que hacer algunos supuestos sobre cómo van a interactuar las habilidades de los trabajadores (una variable no observada) con los años de experiencia y posiblemente con otras variables observadas (educación, género, edad, por nombrar algunas).

Por otro lado, las políticas contrafactuales pueden tomar muchas formas. En este trabajo en particular nos enfocamos en políticas que incrementan la proporción de individuos bajo cierto tratamiento. Dicho tratamiento podría ser pertenecer a algún sindicato o haberse graduado de la universidad. La pregunta de interés es entonces: cuál es el efecto de dicha política en los cuantiles de una determinada variable de resultado. En lugar de hacer supuestos que nos permitan identificar dicho efecto, vamos a proponer un análisis de sensibilidad que funciona de la siguiente manera. Primero caracterizamos el conjunto de efectos compatibles con relajaciones graduales y continuas de los supuestos que arrojan identificación puntual. Luego, dada una cierta conclusión de interés, por ejemplo, que el efecto en la mediana sea positivo, el análisis de sensibilidad provee un punto de quiebre con la siguiente interpretación: una relajación de los supuestos más allá del punto de quiebre no nos permite inferir que la conclusión se cumple.

Nuestra aplicación empírica analiza la relación entre educación universitaria y salarios. En este caso, es de esperar un considerable sesgo de selección entre graduados y no graduados universitarios. Usando datos de la EPH, consideramos diferentes tipos de políticas contrafactuales que incrementan la proporción de graduados universitarios. Luego hacemos un análisis de sensibilidad para analizar la robustez de conclusiones del tipo: “un aumento en la proporción de graduados universitarios incrementa en un 10% el decil 20 de la distribución de salarios”. Un tipo de política contrafactual que parece ser relativamente robusta a la heterogeneidad es la siguiente: un incremento en la proporción de graduados universitarios dado por aquellos individuos que están en la cola izquierda de la distribución de salarios. Si nuestra conclusión de interés es un efecto positivo en los cuantiles más bajos de la distribución de salarios, dicha política es

relativamente robusta. Esto sugiere que la educación podría contribuir a reducir la desigualdad.

Un ejemplo sencillo nos va a ayudar a entender cómo relajamos los supuestos de identificación. Supongamos que estamos en una situación en que el tratamiento ha sido asignado aleatoriamente a una proporción p y los individuos cumplen perfectamente. Esto es, los individuos toman el tratamiento que les fue asignado. La distribución incondicional de la variable que nos interesa es un promedio de las distribuciones de los individuos bajo tratamiento y los individuos bajo control. Los pesos de dicho promedio son p y $1 - p$. Ahora, nos podemos preguntar cómo cambiaría esta distribución si, en vez de asignar una proporción p al tratamiento, asignamos una proporción $p + \delta$. La respuesta es simple: dado que la asignación original fue aleatoria, entonces basta tomar las dos distribuciones condicionales y promediarlas con los pesos $p + \delta$ y $1 - p - \delta$. Sin embargo, quizá nos preguntamos: cómo cambiaría la distribución incondicional si tomáramos ciertos individuos con ciertas características del grupo de control y los asignáramos a tratamiento. En este caso, los nuevos tratados ya no van a ser comparables a los que ya estaban en tratamiento. Entonces es poco probable que el procedimiento de promediar las distribuciones condicionales con los pesos $p + \delta$ y $1 - p - \delta$ nos dé la respuesta que buscamos. El análisis de sensibilidad parte del supuesto de que los nuevos tratados son iguales a los que ya están tratados, en cuyo caso el nuevo promedio es válido, y de a poco relaja el supuesto de igualdad. Esta idea incluso funciona si el tratamiento original no es aleatorio en primer lugar, es decir, si hay algún tipo de sesgo de selección. El análisis de sensibilidad solo se enfoca en cuán diferentes (en cierto sentido que vamos a definir más abajo) pueden ser los nuevos tratados de los tratados originalmente.

El análisis de políticas que modifican la proporción de individuos bajo tratamiento ha sido tratado anteriormente en Firpo, Fortin y Lemieux (2009), Heckman y Vytlacil (1999, 2001, 2005), Carneiro, Heckman y Vytlacil (2010, 2011), Martínez-Iriarte (2020), Martínez-Iriarte y Sun (2020) y Rothe (2010, 2012). Nuestro trabajo se distingue de los anteriores en el sentido de que no ponemos prácticamente restricciones en cómo se lleva a cabo el incremento en cuestión. Por otro lado, la literatura de análisis de sensibilidad es bastante grande y un buen resumen se puede encontrar en Masten y Poirier (2020). Asimismo, nuestro trabajo hace uso de la frontera de quiebre por cuantiles introducida por Martínez-Iriarte (2020).

II. MODELO

El modelo que vamos a utilizar es el de resultados potenciales o *potential outcomes*. En este modelo, hay dos realizaciones latentes $Y(0)$ e $Y(1)$. Para cada individuo solo observamos una de ellas. Sea D la indicadora de si el individuo está en el grupo de tratamiento o en control. Esto es, $D = 1$ si el individuo pertenece al grupo de tratamiento y $D = 0$ si pertenece al grupo de control. Para aquellos individuos bajo tratamiento observamos $Y(1)$ y para aquellos individuos en el grupo de control observamos $Y(0)$. La clave, y la dificultad, radica en que nunca observamos ambas realizaciones para la misma persona. El resultado observado para cada persona se denota Y , que satisface $Y = D Y(1) + (1 - D) Y(0)$.

Nuestro enfoque es no paramétrico. Formalmente, el modelo que proponemos es el siguiente:

$$Y(0) = h_0(X, U),$$

$$Y(1) = h_1(X, U),$$

donde h_0 y h_1 son dos funciones desconocidas sobre las que no imponemos ningún tipo de restricción más allá de las necesarias para asegurar la continuidad de las distribuciones de $Y(0)$ e $Y(1)$. Las variables observadas son denotadas por X y las no observadas por U . Dado que el resultado observado satisface $Y = D Y(1) + (1 - D) Y(0)$, podemos simplificar el modelo escribiendo

$$Y = h(D, X, U),$$

donde h contiene las funciones que determinan $Y(0)$ e $Y(1)$. Como nos vamos a concentrar en políticas contrafactuales que manipulan la proporción de individuos bajo tratamiento, un parámetro importante es $p \stackrel{\text{def}}{=} Pr(D = 1)$. Una política contrafactual es una asignación alternativa D_δ que satisface $Pr(D_\delta = 1) = p + \delta$. Cuando reemplazamos D por D_δ en la ecuación para Y , obtenemos un resultado contrafactual:

$$Y_\delta = h(D_\delta, X, U).$$

Implícitamente estamos haciendo el supuesto de que la función estructural h no se ve afectada por la manipulación en la asignación.

Supuesto 1. Continuidad. Las variables aleatorias Y e Y_δ tienen funciones de distribución absolutamente continuas con densidades positivas en el interior de su soporte.

Este supuesto es útil para minimizar la notación. El supuesto de densidades positivas es necesario dado que estimaremos cuantiles y la varianza de un cuantil estimado es inversamente proporcional al valor de la densidad evaluada en dicho cuantil.

Para analizar el efecto de una política contrafactual, nos vamos a enfocar en los cuantiles. Esto es, vamos a comparar los cuantiles de Y con los cuantiles de Y_δ . Primero introducimos un poco de notación. Vamos a denotar por $F(y)$ la distribución de Y , por $f(y)$ su densidad, y por $F(y, \delta)$ la distribución de Y_δ . Esto es, $F(y) = Pr(Y < y)$, $f(y) = dF(y)/dy$, y $F(y, \delta) = Pr(Y_\delta < y)$. Para un dado $\tau \in (0,1)$, el cuantil τ de Y se denota $F^{-1}(\tau)$. El cuantil contrafactual se denota $F^{-1}(\tau, \delta)$.

Siguiendo a Martínez-Iriarte (2020), vamos a enfocarnos en dos efectos. El primero de ellos es un efecto global:

$$G(\tau, \delta) \stackrel{\text{def}}{=} F^{-1}(\tau, \delta) - F^{-1}(\tau).$$

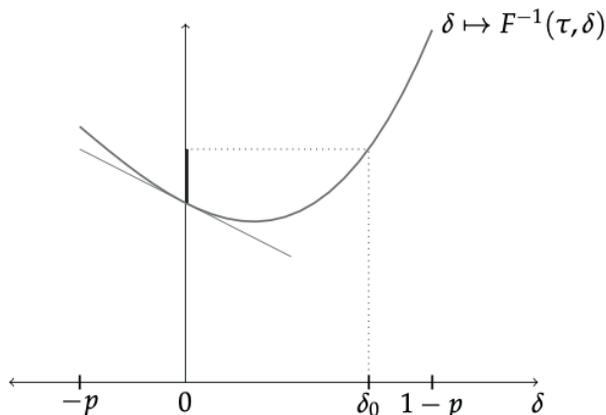
El segundo efecto que vamos a mirar es un efecto marginal, el límite del efecto global reescalado:

$$M(\tau) \stackrel{\text{def}}{=} \lim_{\delta \rightarrow 0} \frac{F^{-1}(\tau, \delta) - F^{-1}(\tau)}{\delta}$$

La existencia de este límite depende de la secuencia de políticas contrafactuales. Una manera de asegurarse que el límite exista es tomando secuencias de políticas que son aleatorizaciones para distintos valores de δ . Por ejemplo, para un dado valor de $\delta > 0$, se selecciona al azar una proporción δ del grupo de control. Esta secuencia garantiza la existencia del efecto marginal. Los detalles se pueden encontrar en Martínez-Iriarte (2020).

La diferencia entre el efecto global y el efecto marginal se puede apreciar en figura 1.

Figura 1. Diferencia entre el efecto global y marginal.



Fuente: elaboración propia.

En la figura 1 se muestra la función que, para un dado cuantil τ , mapea los cuantiles contrafactuales como una función de δ . Esta función es desconocida. También se muestra su tangente en $\delta = 0$. La pendiente de esta recta tangente es el efecto marginal. El efecto global es el segmento resaltado sobre el eje de las ordenadas. Para dado $\delta = \delta_0$, este segmento compara el valor del cuantil contrafactual con el valor del cuantil observado, que corresponde a $\delta = 0$.

Dado que no podemos obtener la distribución contrafactual de Y_δ , vamos a intentar aproximarla. Para ello vamos a hacer el siguiente supuesto sobre D_δ .

Supuesto 2. Monotonicidad. Las políticas contrafactuales D_δ satisfacen $Pr(D_\delta = 1 | D = 0) = 0$.

El supuesto de monotonicidad solo permite políticas que llevan individuos del grupo de control al grupo de tratamiento. No permiten, sin embargo, que individuos pasen del grupo de tratamiento al grupo de control. El supuesto tiene un espíritu práctico y se puede levantar fácilmente a costa de más notación.

Un ejemplo de política contrafactual que cumple el supuesto 2 es el que se mencionó anteriormente en el contexto de la existencia del efecto marginal $M(\tau)$. Otro ejemplo es el siguiente: $D_\delta = 0$ si $D = 0$, y $D_\delta = 1$ si $D = 1$ o si $D = 0$ y

$F^{-1}\left(\frac{\delta}{1-p} \mid D = 0\right) > Y$. En este caso, los individuos que entran al tratamiento son aquellos que tienen un valor de Y lo suficientemente bajo.

Usando el supuesto 2, podemos obtener una expresión general para la distribución contrafactual.

Lema 1. Distribución contrafactual. *Bajo el supuesto 2, la distribución contrafactual está dada por*

$$F(y, \delta) = F_A(y) + \delta[F(y \mid D = 0, D_\delta = 1) - F(y \mid D = 1)],$$

donde $F_A(y)$ está dada por

$$F_A(y) = (1 - p - \delta)F(y \mid D_\delta = 0) + (p + \delta)F(y \mid D = 1).$$

Prueba Lema 1. *Primero notamos que, por el supuesto 2, la población se puede particionar en tres: los nunca tratados $\{D = 0, D_\delta = 0\} = \{D_\delta = 0\}$, los tratados originalmente $\{D = 1, D_\delta = 1\} = \{D = 1\}$ y los nuevos tratados $\{D = 0, D_\delta = 1\}$. Por lo tanto, podemos escribir:*

$$F(y, \delta) = (1 - p - \delta)F(y \mid D_\delta = 0) + p F(y \mid D = 1) + \delta F(y \mid D = 0, D_\delta = 1).$$

El resultado se obtiene sumando y restando $\delta F(y \mid D = 1)$.

La distribución contrafactual depende de $F_A(y)$, que llamamos la distribución aparente, y de la diferencia entre $F(y \mid D = 0, D_\delta = 1)$, la distribución de los nuevos tratados, y $F(y \mid D = 1)$, la distribución de los tratados originalmente. Esta diferencia refleja la heterogeneidad que tiene el grupo de nuevos tratados, aquellos caracterizados por $\{D = 0, D_\delta = 1\}$, con respecto al grupo tratado originalmente. La distribución aparente se llama así porque no tiene esta diferencia y simplemente considera que esos individuos nuevos tratados provienen de la misma distribución que aquellos que ya fueron tratados. Por eso el peso de $(p + \delta)$ dado a la distribución $F(y \mid D = 1)$.

La estrategia para realizar el análisis de sensibilidad es la siguiente. Primero postulamos un valor mínimo y máximo para la diferencia $F(y \mid D = 0, D_\delta = 1) - F(y \mid D = 1)$. Luego, usando esos valores obtenemos cotas para la distribución contrafactual. Estas cotas a su vez nos permiten acotar los cuantiles contrafactuales, que nos sirven para acotar los efectos globales y marginales. Si

supiésemos estos valores mínimos y máximos de la diferencia $F(y | D = 0, D_\delta = 1) - F(y | D = 1)$, entonces nuestro ejercicio terminaría aquí. Sin embargo, como no los sabemos, vamos a proceder de manera inversa. Supongamos que estamos interesados en que el efecto global para un dado cuantil tome cierto valor. Entonces nos preguntamos cuales son los valores de la diferencia $F(y | D = 0, D_\delta = 1) - F(y | D = 1)$ que nos permiten concluir esto. En la siguiente sección desarrollamos este enfoque.

III. ANÁLISIS DE SENSIBILIDAD

Como se mencionó anteriormente, la diferencia $F(y | D = 0, D_\delta = 1) - F(y | D = 1)$ no se puede calcular de los datos. En vez de ello definimos las siguientes cotas mínimas y máximas:

$$L \stackrel{\text{def}}{=} \inf \{F(y | D = 0, D_\delta = 1) - F(y | D = 1)\},$$

$$U \stackrel{\text{def}}{=} \sup \{F(y | D = 0, D_\delta = 1) - F(y | D = 1)\},$$

donde ambos *inf* y *sup* se toman sobre el soporte de las distribuciones. Es fácil ver que L vive entre -1 y 0, mientras que U vive entre 0 y 1. Cuando ambos L y U son iguales a 0, esto corresponde al caso donde la distribución aparente es la distribución contrafactual. Caso contrario, hay una discrepancia entre la distribución aparente y la distribución contrafactual. El resultado del lema 1 nos permite concluir rápidamente que, para todo y , tenemos la siguiente desigualdad:

$$\delta L \leq F(y, \delta) - F_A(y) \leq \delta U.$$

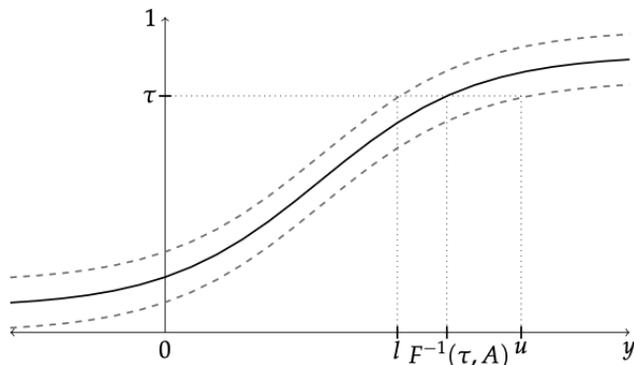
La utilidad de esta desigualdad para acotar los cuantiles de la distribución contrafactual $F(y, \delta)$ es que podemos usar los cuantiles de $F_A(y)$, que se pueden calcular de los datos. Denotamos estos cuantiles de la distribución aparente por $F_A^{-1}(\tau)$.

La figura 2 muestra cómo se procede. La línea sólida es la distribución aparente con las respectivas cotas $F_A(y) + \delta L$ y $F_A(y) + \delta U$. La distribución contrafactual está en algún lugar entre esas cotas, pero no sabemos dónde exactamente. Sin embargo, podemos ver que el cuantil de la distribución contrafactual, $F^{-1}(\tau, \delta)$, tiene que estar entre l y u . Estos dos valores satisfacen las siguientes dos ecuaciones:

$$F_A(l) + \delta U = \tau,$$

$$F_A(u) + \delta L = \tau.$$

Figura 2. Cotas para los cuantiles contrafactuales



Fuente: elaboración propia.

Si resolvemos ambas ecuaciones encontramos que

$$l = F_A^{-1}(\tau - \delta U),$$

$$u = F_A^{-1}(\tau - \delta L),$$

por lo que, siguiendo a la figura 2, podemos concluir que

$$F_A^{-1}(\tau - \delta U) \leq F^{-1}(\tau, \delta) \leq F_A^{-1}(\tau - \delta L).$$

Es importante notar que la cota superior para la distribución contrafactual, δU , ahora juega el rol de cota inferior para el cuantil contrafactual. Lo mismo sucede con la cota inferior δL . Esta inversión es característica de los problemas que involucran cotas en cuantiles. Con estas cotas podemos ahora acotar los efectos marginales y globales.

Teorema 1. Cotas para los efectos. *Bajo los supuestos 1 y 2, el efecto global está acotado por*

$$F_A^{-1}(\tau - \delta U) - F^{-1}(\tau) \leq G(\tau, \delta) \leq F_A^{-1}(\tau - \delta L) - F^{-1}(\tau)$$

para todo $\tau \in (\delta, 1 - \delta)$. El efecto marginal, si existe, está acotado por

$$-\frac{U}{f(F^{-1}(\tau))} \leq M(\tau) - M_A(\tau) \leq -\frac{L}{f(F^{-1}(\tau))}$$

para todo $\tau \in (0, 1)$, y donde $M_A(\tau)$ es el efecto marginal aparente:

$$M_A(\tau) \stackrel{\text{def}}{=} -\frac{F(F^{-1}(\tau)|D = 1) - F(F^{-1}(\tau)|D = 0)}{f(F^{-1}(\tau))}.$$

Prueba Teorema 1. La cota para el efecto global $G(\tau, \delta)$ se desprende inmediatamente de la definición y de la desigualdad establecida en el texto: $F_A^{-1}(\tau - \delta U) \leq F^{-1}(\tau, \delta) \leq F_A^{-1}(\tau - \delta L)$. Las cotas para el efecto marginal $M(\tau)$ se obtienen tomando el límite cuando $\delta \rightarrow 0$ en las cotas (reescaladas) para el efecto global. Es decir:

$$\lim_{\delta \rightarrow 0} \frac{F_A^{-1}(\tau - \delta U) - F^{-1}(\tau)}{\delta} = -\frac{U}{f(F^{-1}(\tau))} + M_A(\tau),$$

$$\lim_{\delta \rightarrow 0} \frac{F_A^{-1}(\tau - \delta L) - F^{-1}(\tau)}{\delta} = -\frac{L}{f(F^{-1}(\tau))} + M_A(\tau).$$

Estos límites son derivadas con respecto a δ , evaluadas en $\delta = 0$. Es importante notar que δ juega un rol doble: en el argumento de F_A^{-1} y en la construcción de la función aparentemente. Por lo que la derivada con respecto a δ es la suma de las derivadas parciales. Para el caso de la primera ecuación, tenemos:

$$\lim_{\delta \rightarrow 0} \frac{F_A^{-1}(\tau - \delta U) - F^{-1}(\tau)}{\delta} = \frac{\partial F^{-1}(\tau - \delta U)}{\partial \delta} + \frac{\partial F_A^{-1}(\tau)}{\partial \delta},$$

donde hemos usado el hecho que en $\delta = 0$, la distribución aparente F_A es la distribución de Y . Para el primer término tenemos

$$\frac{\partial F^{-1}(\tau - \delta U)}{\partial \delta} = -\frac{U}{f(F^{-1}(\tau))}.$$

Para el segundo término recordamos que la distribución aparente está dada por

$$F_A(y) = (1 - p - \delta)F(y | D = 0) + (p + \delta)F(y | D = 1).$$

Donde hemos impuesto el supuesto que $F(y | D_\delta = 0) = F(y | D = 0)$, algo que se cumple si seleccionamos individuos al azar del grupo de control. Ahora, cuando δ se mueve, se mueve la función F_A y luego se mueve la inversa F_A^{-1} . Por lo que necesitamos de una derivada funcional. Hay muchos tipos de derivadas funcionales, pero para el caso de inversas de distribuciones vamos a usar la derivada de Hadamard. Siguiendo el teorema 21.4 de van der Vaart (1998), tenemos que:

$$\frac{\partial F_A^{-1}(\tau)}{\partial \delta} = - \frac{F(F^{-1}(\tau) | D = 1) - F(F^{-1}(\tau) | D = 0)}{f(F^{-1}(\tau))}.$$

Juntando ambos resultados obtenemos la cota inferior para el efecto marginal. Para el caso de la cota superior, la prueba es análoga y la omitimos.

El teorema 1 merece algunos comentarios. La cota para el efecto global solo se puede aplicar para aquellos cuantiles que satisfacen $\tau \in (\delta, 1 - \delta)$. Mientras más pequeño sea δ , o sea, mientras menos individuos se muden de un grupo a otro, más cuantiles podemos acotar. Intuitivamente, mientras menos manipulamos, más sabemos. La razón matemática por la cual requerimos $\tau \in (\delta, 1 - \delta)$ es que necesitamos que simultáneamente $0 \leq \tau - \delta U \leq 1$ y que $0 \leq \tau - \delta L \leq 1$. Tomando los peores casos posibles en términos de U y L , llegamos a que $\tau \in (\delta, 1 - \delta)$. Esta restricción no está presente para el caso del efecto marginal por la sencilla razón que allí δ está yendo a 0, y por consiguiente puede alcanzar más cuantiles.

Las cotas para el efecto marginal involucran el efecto marginal aparente. Este es el estimando propuesto por Firpo, Fortin y Lemieux (2009). Este resultado muestra que, si hay suficiente heterogeneidad entre las subpoblaciones, el estimando de Firpo, Fortin y Lemieux (2009) no va a converger al verdadero efecto marginal. Las cotas para el efecto marginal son más difíciles de computar dado que involucran la densidad f evaluada en el cuantil $F^{-1}(\tau)$. Ambos objetos deben ser estimados. A menos que se impongan restricciones paramétricas, uno tiene que recurrir a un estimador no paramétrico como, por ejemplo, un kernel para estimar la densidad. Por otro lado, la presencia de la densidad evaluada en $F^{-1}(\tau)$ en el denominador, nos dice que a medida que τ se acerca a 0 o a 1, $F^{-1}(\tau)$ se acerca a

$-\infty$ o $+\infty$, por lo que la densidad se acerca a 0 en ambos casos. Esto ocasiona que las cotas sean demasiado amplias y por lo tanto posiblemente no informativas.

Con el resultado del teorema 1 estamos listos para explicar en detalle cómo funciona el análisis de sensibilidad. Lo ilustraremos usando el efecto global. Supongamos que nos interesa que la conclusión $G(\tau, \delta) > g(\tau)$ se cumpla para un dado τ y $g(\tau)$. Inspeccionando las cotas, esto va a cumplirse solo si

$$g(\tau) \leq F_A^{-1}(\tau - \delta U) - F^{-1}(\tau),$$

porque en ese caso el efecto global vive por encima de $g(\tau)$. Esto nos permite encontrar un valor límite de U , que denotamos $U(\tau)$, donde la desigualdad anterior se cumple con igualdad:

$$g(\tau) = F_A^{-1}(\tau - \delta U(\tau)) - F^{-1}(\tau),$$

lo que implica que podemos resolver para $U(\tau)$:

$$U(\tau) = \min \left\{ \max \left\{ 0, \frac{1}{\delta} [\tau - F_A(F^{-1}(\tau) + g(\tau))] \right\}, 1 \right\}.$$

Un valor de U mayor que $U(\tau)$ no nos permite inferir que $G(\tau, \delta) \geq g(\tau)$. Es una posibilidad que para todos los valores de U , no podamos inferir $G(\tau, \delta) > g(\tau)$. En este caso la conclusión no se cumple ni aun relajando al máximo las cotas. Sin embargo, si nos fijamos en la conclusión opuesta, $G(\tau, \delta) < g(\tau)$, en este caso vamos a restringir el valor de L , de tal manera que

$$F_A^{-1}(\tau - \delta L) - F^{-1}(\tau) \leq g(\tau).$$

De manera análoga, vamos a definir $L(\tau)$ como el valor del L para el cual esa desigualdad se cumple con igualdad:

$$L(\tau) = \max \left\{ \min \left\{ 0, \frac{1}{\delta} [\tau - F_A(F^{-1}(\tau) + g(\tau))] \right\}, -1 \right\}.$$

Una comparación de $U(\tau)$ con $L(\tau)$ nos muestra que en realidad son la misma fórmula, con las composición max/min alternada. Esta composición nos asegura que $U(\tau) \in [0, 1]$ y que $L(\tau) \in [-1, 0]$. Lo único que cambia es que cuando $\frac{1}{\delta} [\tau - F_A(F^{-1}(\tau) + g(\tau))]$ es positivo, entonces es $U(\tau)$, y cuando es negativo es $L(\tau)$. Por lo tanto, el procedimiento es: primero fijamos $g(\tau)$, y luego

de calcular $\frac{1}{\delta}[\tau - F_A(F^{-1}(\tau) + g(\tau))]$, sabemos si es que podemos inferir $G(\tau, \delta) > g(\tau)$ o $G(\tau, \delta) < g(\tau)$. Luego, interpretamos $\frac{1}{\delta}[\tau - F_A(F^{-1}(\tau) + g(\tau))]$ como el valor máximo $U(\tau)$ que podemos relajar U , o como el valor máximo (en valor absoluto) $L(\tau)$ que podemos relajar L .

La descripción anterior nos motiva a definir el mapa:

$$\tau \rightarrow \theta(\tau) = \frac{1}{\delta}[\tau - F_A(F^{-1}(\tau) + g(\tau))]$$

que es una versión de la *frontera de quiebre por cuantiles* para el efecto global introducida por Martínez-Iriarte (2020). Mientras que anteriormente nos fijamos en un cuantil en particular, la frontera se puede computar para todos los cuantiles $\tau \in (\delta, 1 - \delta)$, que es donde podemos acotar el efecto global.

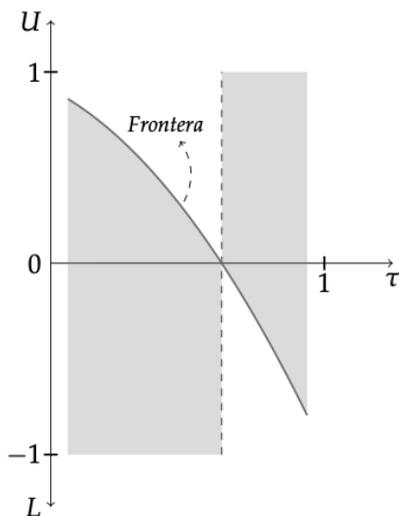
Para entender cómo usar la frontera, supongamos por un momento que $g(\tau) = 0$ para todo $\tau \in (\delta, 1 - \delta)$. Entonces la frontera se puede escribir:

$$\tau \rightarrow \theta(\tau) = \frac{1}{\delta}[\tau - F_A(F^{-1}(\tau))].$$

La figura 3 contiene una versión hipotética de la frontera. Podemos ver que el plano se divide en cuatro regiones. Donde la frontera es positiva, si nos movemos más arriba de la frontera, entonces la conclusión $G(\tau, \delta) > 0$ quizá no se cumple. Mientras U se mantenga por debajo de la frontera, entonces sabemos que $G(\tau, \delta) > 0$ se cumple. Esta región está pintada de gris. Denominamos a esta región, siguiendo a Masten y Poirier (2020) como la región robusta. Es importante notar que la región robusta para $G(\tau, \delta) > 0$ no restringe el valor de L .

Un análisis similar aplica al caso donde la frontera es negativa. En este caso L no puede ir mas allá de la frontera, es decir, no se puede acercarse demasiado a -1 si pretendemos que la conclusión $G(\tau, \delta) < 0$ se cumpla. Esta es la región en gris, arriba de la frontera, donde la frontera es negativa. Por lo tanto, un vistazo a la frontera nos permite tener una idea de para cuáles cuantiles la conclusión de interés es más robusta.

Figura 3. Ejemplo de la frontera de quiebre por cuantiles



Fuente: elaboración propia.

Por último, es importante remarcar que la frontera no tiene que ser decreciente. La continuidad tampoco está asegurada a menos que la ‘función’ de conclusiones $\tau \rightarrow g(\tau)$ sea continua.

Para el efecto marginal, si nos concentramos en el caso $g(\tau) = 0$, la frontera tiene la misma forma que en el efecto global, sin embargo la interpretación es con respecto a conclusiones del efecto marginal $M(\tau)$. Es posible construir una frontera para el efecto marginal para caso que $g(\tau)$ no es necesariamente 0. Sin embargo, dicha construcción es más complicada porque involucra la densidad de Y evaluada en el cuantil incondicional: $f(F^{-1}(\tau))$. Además, en términos de inferencia, la convergencia sería a una tasa no paramétrica. Por otro lado, dado que el efecto marginal tiene la interpretación de una derivada, concentrarse en el caso $g(\tau) = 0$ es equivalente a concentrarse en el signo de la derivada.

IV. ESTIMACIÓN E INFERENCIA

Para un dado τ y $g(\tau)$ tenemos que estimar tres objetos:

$$\theta(\tau) = \frac{1}{\delta} [\tau - F_A(F^{-1}(\tau) + g(\tau))]$$

y los valores de $U(\tau)$ y $L(\tau)$ que están dados por $U(\tau) = \min\{\max\{0, \theta(\tau)\}, 1\}$ y $L(\tau) = \max\{\min\{0, \theta(\tau)\}, -1\}$. Es conveniente empezar por $\theta(\tau)$, dado que para un estimador $\hat{\theta}(\tau)$, podemos estimar $\hat{U}(\tau) = \min\{\max\{0, \hat{\theta}(\tau)\}, 1\}$ y $\hat{L}(\tau) = \max\{\min\{0, \hat{\theta}(\tau)\}, -1\}$. Sin pérdida de generalidad, vamos a normalizar $g(\tau) = 0$ para aliviar notación.

Vamos a suponer que tenemos una muestra *i.i.d.* que consiste de $\{Y_i, D_i, D_{\delta i}\}_{i=1}^n$. Para estimar $\theta(\tau)$, empezamos con el cuantil τ de Y :

$$\hat{F}^{-1}(\tau) = \arg \min_q \sum_{i=1}^n \rho_{\tau}(Y_i - q)$$

donde $\rho_{\tau}(u)$ es la función *check* de Koenker y Bassett (1978): $\rho_{\tau}(u) = [\tau - 1(u \leq 0)]u$. Un resultado estándar que se puede encontrar, por ejemplo, en el corolario 21.5 de van der Vaart (1998) es que los cuantiles son asintóticamente normales con la siguiente función de influencia:

$$\begin{aligned} \sqrt{n} \left(\hat{F}^{-1}(\tau) - F^{-1}(\tau) \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\tau - 1(Y_i \leq F^{-1}(\tau))}{f(F^{-1}(\tau))} + o_p(1) \\ &\stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{Q,i} + o_p(1). \end{aligned}$$

La distribución aparente F_A se puede estimar por medio de una combinación convexa de distribuciones empíricas:

$$\hat{F}_A(y) = (1 - \hat{p} - \delta) \hat{F}(y|D_{\delta} = 0) + (\hat{p} + \delta) \hat{F}(y|D = 1)$$

donde los pesos se construyen con \hat{p} , un estimador de la proporción de individuos bajo tratamiento:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n D_i$$

y las distribuciones empíricas están dadas por

$$\hat{F}(y|D_\delta = 0) = \frac{1}{(1 - \hat{p} - \delta)} \frac{1}{n} \sum_{i=1}^n 1(Y_i \leq y) (1 - D_{\delta i})$$

y

$$\hat{F}(y|D = 1) = \frac{1}{(\hat{p} + \delta)} \frac{1}{n} \sum_{i=1}^n 1(Y_i \leq y) D_i$$

Por lo tanto, el estimador de la frontera es

$$\hat{\theta}(\tau) = \frac{1}{\delta} \left[\tau - \hat{F}_A \left(\hat{F}^{-1}(\tau) \right) \right]$$

donde remarcamos nuevamente que hemos normalizado $g(\tau) = 0$. Para mostrar que la frontera es consistente y asintóticamente normal, necesitamos el siguiente supuesto:

Supuesto 3. Teorema del límite central uniforme. Para un dado τ y δ , el siguiente teorema del límite central se cumple de manera uniforme para todo y :

$$\sqrt{n} \begin{pmatrix} \hat{F}(y|D_\delta = 0) - F(y|D_\delta = 0) \\ \hat{F}(y|D = 1) - F(y|D = 1) \end{pmatrix} \xrightarrow{d} N(0, \Sigma_F(y))$$

donde $\Sigma_F(y)$ es una matriz positiva definida de 2×2 para todo y .

Nos va a ser útil definir las siguientes funciones de influencia:

$$\begin{aligned}
& \sqrt{n} \left(\hat{F}(y|D_\delta = 0) - F(y|D_\delta = 0) \right) \\
&= \frac{1}{(1 - \hat{p} - \delta)} \frac{1}{\sqrt{n}} \sum_{i=1}^n [1(Y_i \leq y)(1 - D_{\delta i}) - F(y|D_\delta = 0)] \\
&\stackrel{\text{def}}{=} \frac{1}{(1 - p - \delta)} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{0,i}(y).
\end{aligned}$$

De manera similar, tenemos que

$$\begin{aligned}
& \sqrt{n} \left(\hat{F}(y|D = 1) - F(y|D = 1) \right) \\
&= \frac{1}{(\hat{p} + \delta)} \frac{1}{\sqrt{n}} \sum_{i=1}^n [1(Y_i \leq y)D_i - F(y|D = 1)] \\
&\stackrel{\text{def}}{=} \frac{1}{(p + \delta)} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{1,i}(y) + o_p(1),
\end{aligned}$$

donde en ambos casos, el término $o_p(1)$ absorbe la incertidumbre asociada con la estimación de p .

El supuesto anterior, junto con una ley de los grandes que se cumple para \hat{p} , es decir, \hat{p} converge en probabilidad a p , y el teorema de Slutsky, nos permite concluir que $\hat{F}_A(y) \xrightarrow{p} F_A(y)$ para todo y . Combinando las funciones de influencias, tenemos la siguiente representación lineal:

$$\sqrt{n}(\hat{F}_A(y) - F_A(y)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{A,i}(y) + o_p(1),$$

donde $\psi_{A,i} \stackrel{\text{def}}{=} \psi_{0,i} + \psi_{1,i}$. Es decir, el estimador de la distribución aparente es asintóticamente normal.

Por otro lado, en la expresión de la frontera tenemos que \hat{F}_A está evaluada en un punto aleatorio: $\hat{F}^{-1}(\tau) + g(\tau)$. También se cumple que $\hat{F}^{-1}(\tau) + g(\tau) \xrightarrow{p} F^{-1}(\tau) + g(\tau)$ por la consistencia de estimador de los cuantiles. Sin embargo, esto no es suficiente para concluir que $\hat{F}_A(\hat{F}^{-1}(\tau) + g(\tau)) \xrightarrow{p} F_A(F^{-1}(\tau) + g(\tau))$. Es por eso que necesitamos el supuesto de uniformidad. Esta situación es análoga al caso de una secuencia convergente de funciones evaluada en una secuencia

convergente de puntos. Necesitamos convergencia uniforme para asegurar que la doble secuencia converja a la función convergente evaluada en el punto de convergencia. Ver ejercicio 9 del capítulo 7 de Rudin (1976). Por lo tanto, podemos concluir que el estimador de la frontera es consistente: $\hat{\theta}(\tau) \xrightarrow{p} \theta(\tau)$.

Para obtener la distribución asintótica de la frontera, luego de algo de álgebra, escribimos la diferencia como $\hat{\theta}(\tau) - \theta(\tau) = -\frac{1}{\delta} \left[\hat{F}_A \left(\hat{F}^{-1}(\tau) \right) - F_A \left(F^{-1}(\tau) \right) \right]$. Primero vamos a dejar de lado el factor $-\frac{1}{\delta}$ y obtener la distribución asintótica de la expresión entre corchetes. Para ello, descomponemos la diferencia $\hat{F}_A \left(\hat{F}^{-1}(\tau) \right) - F_A \left(F^{-1}(\tau) \right)$ en dos términos.

El primero de ellos es:

$$\hat{F}_A \left(\hat{F}^{-1}(\tau) \right) - F_A \left(\hat{F}^{-1}(\tau) \right)$$

y el segundo es

$$F_A \left(\hat{F}^{-1}(\tau) \right) - F_A \left(F^{-1}(\tau) \right).$$

El primer término, usando la uniformidad de su convergencia, se puede escribir como

$$\hat{F}_A \left(\hat{F}^{-1}(\tau) \right) - F_A \left(\hat{F}^{-1}(\tau) \right) = \hat{F}_A \left(F^{-1}(\tau) \right) - F_A \left(F^{-1}(\tau) \right) + o_p \left(n^{-1/2} \right)$$

y el segundo término, mediante una aplicación del método Delta, es

$$F_A \left(\hat{F}^{-1}(\tau) \right) - F_A \left(F^{-1}(\tau) \right) = f_A \left(F^{-1}(\tau) \right) \left(\hat{F}^{-1}(\tau) - F^{-1}(\tau) \right) + o_p \left(n^{-\frac{1}{2}} \right).$$

Esta descomposición nos permite obtener la distribución asintótica de la frontera y separar la contribución de cada uno de sus ingredientes. Juntando todas las funciones de influencia, obtenemos la siguiente representación para la frontera:

$$\sqrt{n} \left(\hat{\theta}(\tau) - \theta(\tau) \right) = -\frac{1}{\delta \sqrt{n}} \sum_{i=1}^n \left[\psi_{A,i} \left(F^{-1}(\tau) \right) + f_A \left(F^{-1}(\tau) \right) \psi_{Q,i} \right] + o_p(1),$$

es decir, la frontera es asintóticamente normal, con varianza dada por

$$V_{\theta(\tau)} = E \left([\psi_{A,i}(F^{-1}(\tau)) + f_A(F^{-1}(\tau))\psi_{Q,i}]^2 \right),$$

que puede ser estimada con su análogo muestral. La consistencia de dicho estimador de la varianza se omite por brevedad, pero está asegurada por una ley de los grandes números uniforme. Alternativamente, el bootstrap es válido por el teorema 23.5 de van der Vaart (1998).

Ahora nos resta obtener las distribuciones asintóticas de $\hat{U}(\tau) = \min\{\max\{0, \hat{\theta}(\tau)\}, 1\}$ y $\hat{L}(\tau) = \max\{\min\{0, \hat{\theta}(\tau)\}, -1\}$. Ambas cantidades son funciones $\hat{\theta}(\tau)$. Sin embargo, son funciones que no son diferenciables en ciertos puntos por la presencia del max y min. Esto no nos permite aplicar el método Delta tradicionalmente. Siguiendo a Fang y Santos (2019) y Masten y Poirier (2020), aplicamos el método Delta direccional. Para ello necesitamos encontrar la derivada direccional de las funciones $r(x) = \min\{\max\{0, x\}, 1\}$ y $s(x) = \max\{\min\{0, x\}, -1\}$. Dado que x es un parámetro de una dimensión, la derivada direccional corresponde a acercarse por derecha o por izquierda. Cuando estos dos límites son iguales, la función es diferenciable en ese punto. Dado que las funciones no son muy complicadas, podemos encontrar sus derivadas considerando distintos valores de su dominio y cómo nos podemos acercar a ellos. Para el caso de $r(x)$, si $x \in (0,1)$, entonces $r(x) = x$, y es diferenciable, con $r'(x) = 1$. Cuando $x > 1$, entonces $r(x) = 1$, y su derivada $r'(x) = 0$. Lo mismo sucede cuando $x < 0$. En este caso, $r(x) = 0$, y su derivada $r'(x) = 0$. Los casos problemáticos son $x = 0$ y $x = 1$. Si nos acercamos por izquierda a $x = 0$, la derivada es 0, pero es igual a 1 si nos acercamos por derecha. Al revés sucede con $x = 1$. Vamos a escribir la derivada como un mapa que aproxima $r(x+h) - r(x)$, y la vamos a denotar por $r'_x(h)$. Entonces

$$h \rightarrow r'_x(h) = \begin{cases} 0 & \text{si } x > 1 \text{ o } x < 0 \\ h & \text{si } x \in (0,1) \\ \min(h, 0) & \text{si } x = 1 \\ \max(h, 0) & \text{si } x = 0 \end{cases}$$

La idea es que, como el mapa original es una función lineal por partes, entonces $r(x+h) - r(x) = r'_x(h)$. Entonces la diferencia depende de donde estamos en el dominio, el valor de x , y de hacia donde nos movemos: a la derecha, con un $h > 0$, o hacia la izquierda, con un $h < 0$. Para el caso de $s(x)$, de igual manera tenemos:

$$h \rightarrow s'_x(h) = \begin{cases} 0 & \text{si } x < -1 \text{ o } x > 0 \\ h & \text{si } x \in (-1,0) \\ \max(h, 0) & \text{si } x = 1 \\ \min(h, 0) & \text{si } x = 0 \end{cases}$$

y la derivada es el mapa definido por $r(x + h) - r(x) = s'_x(h)$.

Podemos ver que una característica de estas derivadas direccionales, es que no son un mapa lineal de h . En particular la linealidad se rompe en los puntos donde la derivada tradicional no existe. Sin embargo, sí son continuas.

El teorema 2.1 en Fang y Santos (2019) nos permite utilizar esta derivada direccional para encontrar la distribución asintótica de $\hat{U}(\tau)$ y $\hat{L}(\tau)$. En particular tenemos que:

$$\sqrt{n}(\hat{U}(\tau) - U(\tau)) = r'_{\theta(\tau)}(\sqrt{n}(\hat{\theta}(\tau) - \theta(\tau))) + o_p(1),$$

y que

$$\sqrt{n}(\hat{L}(\tau) - L(\tau)) = s'_{\theta(\tau)}(\sqrt{n}(\hat{\theta}(\tau) - \theta(\tau))) + o_p(1).$$

Dado que las derivadas son continuas, entonces podemos tomar el límite en distribución “adentro” de las derivadas. Pero, como estas derivadas no son lineales, las distribuciones asintóticas no van a ser normales: estamos aplicando una transformación no lineal a una distribución normal. Por ejemplo, $aN(0,1) + b$ se distribuye como una normal, pero $N(0,1)^2$ tiene una distribución χ^2 con un grado de libertad. Esto complica la inferencia porque, como muestran Fang y Santos (2019), el bootstrap clásico no va a ser consistente. Para poder hacer inferencia correctamente, vamos a usar el método Delta numérico de Hong y Li (2018) como en Masten y Poirier (2020).

V. APLICACIÓN EMPÍRICA

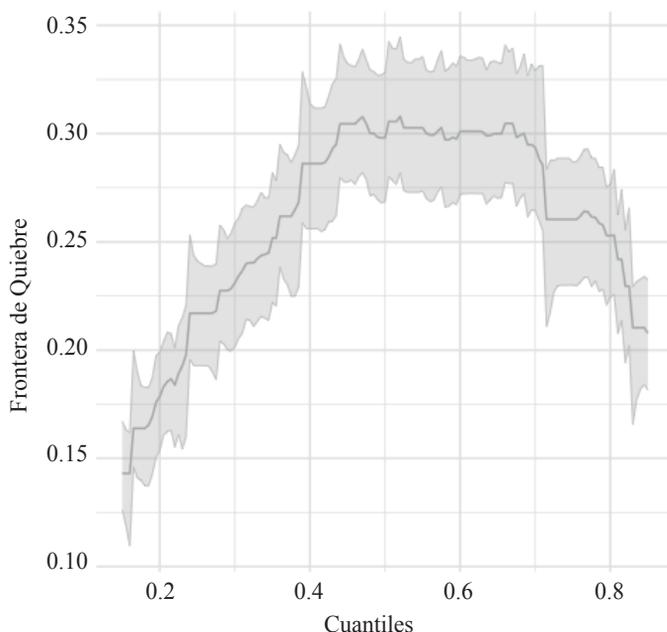
En nuestra aplicación empírica vamos a analizar la relación entre la educación universitaria y los salarios. Nuestra política contrafactual es un incremento en la proporción de graduados universitarios. La variable de resultado que miramos es el salario por hora. Usamos datos de la Encuesta Permanente de Hogares Base Individual del INDEC para el cuarto trimestre de 2019 (Instituto

Nacional de Estadísticas y Censos [INDEC], 2019). La muestra se circunscribe a hombres, entre 18 y 64 años, ocupados, cuya categoría ocupacional es obrero o empleado (asalariados). El salario por hora es construido como el ingreso en la ocupación principal (variable P21) dividido en el total de horas mensuales trabajadas (calculadas como $4*PP3E_TOT$). Esta es nuestra variable Y . Dejamos de lado aquellos individuos que reportan 0 horas trabajadas o 0 ingreso en la ocupación principal.

Para construir la variable binaria D , usamos la variable NIVEL_ED y definimos: $D = 1$ si el individuo reporta “Superior universitario completo” y $D = 0$ si el individuo reporta “Superior universitario incompleto” o “Secundario completo”. Solo consideramos aquellos individuos que no están activamente siendo educados (aquellos que responden “No asiste, pero asistió” en la variable CH10). En total tenemos 4712 observaciones.

La primera política contrafactual que analizamos es una política que incrementa marginalmente la proporción de graduados universitarios. Estos son seleccionados al azar del grupo $D = 0$. Construimos la frontera para el efecto marginal y miramos el caso de $g(\tau) = 0$. Los resultados se muestran en la figura 4. La zona sombreada contiene intervalos de confianza (punto por punto) al 95%. Podemos ver que la frontera es mayormente positiva, lo que a priori nos permitiría concluir que el efecto marginal es positivo. Sin embargo, el valor máximo de la frontera es 0.31, lo que se interpreta como: si $U > 0.31$, entonces para ningún cuantil podemos asegurar que el efecto marginal sea positivo. Si nos enfocamos en la mediana, tenemos que $\hat{U}(0.5) = 0.3$ y un intervalo de confianza del 95% calculado con el bootstrap numérico de $[0.27, 0.33]$. Esto se interpreta como: mientras U sea menor que 0.3, entonces podemos asegurar que el efecto marginal en la mediana será positivo.

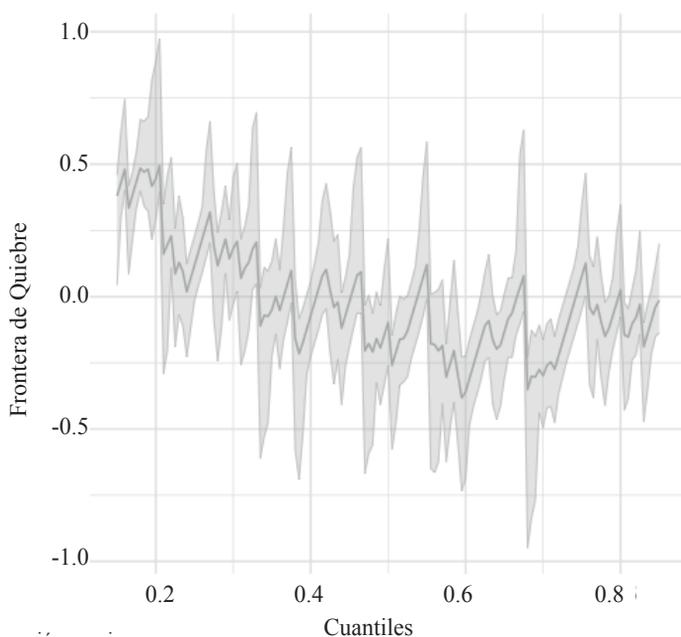
Figura 4. Frontera para el efecto marginal



Fuente: elaboración propia.

La segunda política contrafactual que analizamos es un incremento del 10% en la proporción de graduados universitarios. Para construir la frontera para el efecto global nos fijamos en $g(\tau) = 0.1$. Dado que las variables están en logaritmos, esto se corresponde a un aumento del 10% de los cuantiles. La política elige individuos que están por debajo del cuantil 0.14 en los salarios de los universitarios. Si dichos individuos se unen al grupo $D = 1$, entonces la proporción de graduados universitarios aumenta en un 10% aproximadamente. La frontera, junto con los intervalos de confianza al 95% (punto a punto), se muestran en la figura 5. Se desprende de dicha figura que la frontera tiende a ser positiva para los cuantiles más bajos y negativa para los cuantiles más altos. Esto se debe a la forma de la política contrafactual: elegimos los que tienen salarios más bajos. Si nos fijamos en el cuantil .25, tenemos que $\hat{U}(0.25) = 0.12$ con un intervalo de confianza al 95% dado por $[0.007, 0.14]$. Esto se interpreta como que un valor de U mayor que 0.12 no nos permite concluir que la política tendrá un efecto de incrementar los salarios más del 10% en el cuantil .25. Este valor de $\hat{U}(0.25)$ es relativamente bajo, por lo que la conclusión es poco robusta a diferencias no observables entre las dos subpoblaciones.

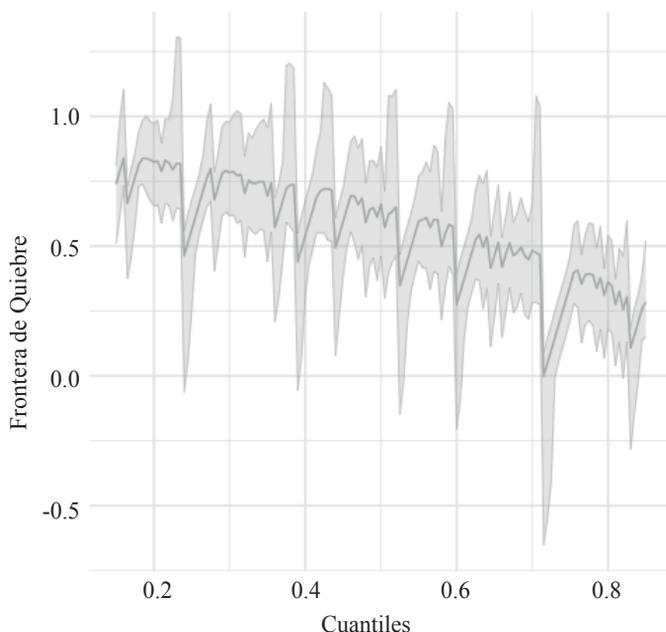
Figura 5. Frontera para el efecto global



Fuente: elaboración propia.

Podemos analizar la misma política contrafáctica, pero esta vez usando una conclusión más débil: $g(\tau) = 0$. Los resultados se muestran en la figura 6. Notamos, nuevamente, que la frontera tiende a ser positiva en los cuantiles más bajos y se acerca a cero en los cuantiles más altos. En este caso, tenemos que $\hat{U}(0.25) = 0.56$ con un intervalo de confianza al 95% dado por $[0.46, 0.67]$, por lo que la conclusión es relativamente más robusta que en el caso de $g(\tau) = 0.1$. Es decir, con un U de hasta 0.56 todavía podemos asegurar que los salarios para el cuantil 25% van a aumentar como consecuencia de esta política. Sin embargo, vemos por ejemplo que este valor de U está por encima de la frontera para los cuantiles más altos. Esto significa que para esos cuantiles no podemos asegurar qué efecto va a ser positivo.

Figura 6. Frontera para el efecto global



Fuente: elaboración propia.

VI. CONCLUSIÓN

En este trabajo nos enfocamos en el efecto en los cuantiles de políticas contrafactuales que incrementan la proporción de individuos bajo tratamiento. En vez de imponer supuestos que nos permitan identificar dicho efecto, realizamos un análisis de sensibilidad para cuantificar la robustez de ciertas conclusiones en diferentes cuantiles. Nuestra aplicación empírica analiza la relación entre educación universitaria y salarios. En particular, encontramos que en una política contrafactual que aumenta la proporción de graduados universitarios con individuos de la cola izquierda de la distribución de salarios es relativamente robusta la heterogeneidad entre graduados y no graduados. Dicha política podría ser útil para reducir la desigualdad en salarios dado que comprimiría la distribución.

BIBLIOGRAFÍA

- Carneiro, P., Heckman, J. J., & Vytlacil, E. (2010). Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica*, 78(1), 377-394. doi: <https://doi.org/10.3982/ECTA7089>
- Carneiro, P., Heckman, J. J., & Vytlacil, E. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6), 2754-2781.
- Fang, Z., & Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1), 377-412. doi: <https://doi.org/10.1093/restud/rdy049>
- Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3), 953-973.
- Heckman, J. J., & Vytlacil, E. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8), 4730-4734. doi: <https://doi.org/10.1073/pnas.96.8.4730>
- Heckman, J. J., & Vytlacil, E. (2001). Policy relevant treatment effects. *American Economic Review*, 91(2), 107-111. doi: 10.1257 / aer.91.2.107
- Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3), 669-738. doi: <https://doi.org/10.1111/j.1468-0262.2005.00594.x>
- Hong, H., & Li, J. (2018). The numerical delta method. *Journal of Econometrics*, 206(2), 379-394. doi: <https://doi.org/10.1016/j.jeconom.2018.06.007>
- Instituto Nacional de Estadísticas y Censos - INDEC (2019). *Encuesta Permanente de Hogares Base Individual*. Recuperado de <https://www.indec.gov.ar/indec/web/Institucional-Indec-BasesDeDatos>
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33-50. doi: 10.2307/1913643
- Kozłowski, D., Tiscornia, P., Weksler, G., Rosati, G., & Shokida, N. (2020). *eph: Argentina's Permanent Household Survey Data and Manipulation Utilities*. R package version doi: <https://doi.org/10.5281/zenodo.3462677>
- Martinez-Iriarte, J. (2020). *Sensitivity Analysis in Unconditional Quantile Effects*. (RedNIE Working Papers No. 52).
- Martinez-Iriarte, J., & Sun, Y. (2020). *Identification and Estimation of Unconditional Policy Effects of an Endogenous Binary Treatment*. (UC San Diego, Working Paper Series qt2bc57830) Recuperado de <https://escholarship.org/uc/item/2bc57830>
- Masten, M. A., & Poirier, A. (2020). Inference on Breakdown Frontiers. *Quantitative Economics, Econometric Society*, 11(1), 41-111. doi: <https://doi.org/10.3982/QE1288>

- Rothe, C. (2010). Identification of Unconditional Partial Effects in Nonseparable Models. *Economics Letters*, 109(3), 171-174.
- Rothe, C. (2012). Partial Distributional Policy Effects. *Econometrica*, 80(5), 2269-2301. doi: <https://doi.org/10.3982/ECTA9671>
- Rudin, W. (1976). *Principles of Mathematical Analysis*. New York: McGraw-Hill.
- Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511802256

© 2022 por los autores; licencia no exclusiva otorgada a la revista Estudios económicos. Este artículo es de acceso abierto y distribuido bajo los términos y condiciones de una licencia Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0) de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc/4.0>

