



Universidad Autónoma del Estado de Hidalgo

Instituto de Ciencias Básicas e Ingeniería
Área Académica de Matemáticas y Física

Efectos de los programas de becas de
capacitación para el trabajo sobre las
condiciones laborales en México

T E S I S

que para obtener el título de:

Licenciada en Matemáticas Aplicadas

presenta

Daniela Hernández Nicolás

Bajo la dirección de la

Dra. Alma Sofía Santillán Hernández

Mineral de la Reforma, Hidalgo.

Julio 2023



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
Instituto de Ciencias Básicas e Ingeniería
School of Engineering and Basic Sciences

Mineral de la Reforma, Hgo., a 1 de septiembre de 2023

Número de control: ICBI-D/1412/2023
Asunto: Autorización de impresión.

**MTRA. OJUKY DEL ROCÍO ISLAS MALDONADO
DIRECTORA DE ADMINISTRACIÓN ESCOLAR DE LA UAEH**

Con fundamento en lo dispuesto en el Título Tercero, Capítulo I, Artículo 18 Fracción IV; Título Quinto, Capítulo II, Capítulo V, Artículo 51 Fracción IX del Estatuto General de nuestra Institución, por este medio le comunico que el Jurado asignado a la Pasante de la Licenciatura en Matemáticas Aplicadas **Daniela Hernández Nicolás**, quien presenta el trabajo de titulación "**Efectos de los programas de capacitación para el trabajo sobre las condiciones laborales en México**", después de revisar el trabajo en reunión de Sinodales ha decidido autorizar la impresión del mismo, hechas las correcciones que fueron acordadas.

A continuación, firman de conformidad los integrantes del Jurado:

Presidente Dr. Roberto Ávila Pozos

Secretario: Mtra. Margarita Tetlalmatzi Montiel

Vocal: Dra. Alma Sofia Santillán Hernández

Suplente: Dr. Raúl Temoltzi Ávila

Sin otro particular por el momento, reciba un cordial saludo.

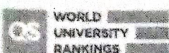
Atentamente
"Amor, Orden y Progreso"

Dr. Otilio Arturo Acevedo Sandoval
Director del ICBI



OAAS/YCC

Ciudad del Conocimiento
Carretera Pachuca-Tulancingo km 4.5 Colonia
Carboneras, Mineral de la Reforma, Hidalgo,
México. C.P. 42184
Teléfono: 771 71 720 00 ext. 2231 Fax 2109
direccion_icbi@uaeh.edu.mx



www.uaeh.edu.mx

Resumen

El problema del empleo percibido a través del desempleo, el subempleo y la dificultad para conseguir un trabajo digno, tiene una importancia considerable debido a los efectos negativos que conlleva tanto a nivel personal como social. En México se han implementado desde hace muchos años programas de apoyo a la capacitación laboral con la finalidad de mejorar las condiciones de empleo de las personas al adquirir habilidades y conocimientos que demanda el mercado laboral. El objetivo de este trabajo es determinar cuál es la mejora en las condiciones laborales de los individuos que han sido beneficiarios de programas de becas de capacitación en términos de algunas variables de interés como el ingreso mensual, las horas de trabajo, las prestaciones laborales, las prestaciones de salud y la colocación de empleo. Específicamente, se estima el impacto de las becas de capacitación sobre el ingreso, las horas trabajadas, las prestaciones de salud y las prestaciones laborales en las personas ocupadas que fueron beneficiarios de las becas y el impacto en la colocación en el empleo en las personas buscadoras de empleo que recibieron beca de capacitación. Para efectuar la estimación de los impactos de las becas de capacitación se utiliza la metodología de diferencias en diferencias y los datos empleados para ello son provenientes de la Encuesta Nacional de Ocupación y Empleo (ENOE) elaborada por INEGI. Los resultados obtenidos muestran un efecto positivo en las variables ingreso mensual, horas de trabajo y prestaciones laborales; el efecto en la variable prestaciones de salud es significativo únicamente en el largo plazo, es decir al año de la recepción de becas. En cuanto a la colocación de empleo, se observa que la probabilidad de acceder a un empleo disminuye para aquellos que fueron beneficiarios de las becas.

Agradecimientos

Quiero expresar mi gratitud a aquellas personas que de alguna manera contribuyeron a este logro, directa o indirectamente. Su apoyo ha sido invaluable.

Primeramente, agradezco a la Dra. Alma Sofía Santillán Hernández, directora de esta tesis, por todo el tiempo dedicado a la orientación y la supervisión de este trabajo, por su paciencia, y por todo el conocimiento recibido como tesista y como su alumna durante la licenciatura. Su experiencia, conocimiento y dedicación han sido una fuente de inspiración para mí. Gracias por brindarme la oportunidad de crecer y aprender bajo su tutela.

También quiero agradecer a mi jurado evaluador: Dr. Roberto Ávila Pozos, Mtra. Margarita Tetlalmatzi Montiel y Dr. Raúl Temoltzi Ávila, por su tiempo dedicado a la revisión de este trabajo y por sus valiosos comentarios.

Agradezco de manera especial a mis padres por su apoyo incondicional. Mamá y papá, sus esfuerzos y aliento han sido el motor que me ha impulsado a lo largo de este recorrido. Gracias por ser mis guías, mis héroes y mis amigos.

A mis hermanos, Fernando y Jesús, les agradezco por ser mi red de apoyo constante, por su inquebrantable fe en mí, por sus palabras de ánimo y aliento constante.

Índice general

Resumen	v
Agradecimientos	vii
Introducción	1
1. Análisis de regresión lineal	5
1.1. Modelo de regresión lineal simple	5
1.1.1. Supuestos	6
1.1.2. Estimación de los parámetros	7
1.1.3. Inferencia sobre los parámetros	12
1.2. Modelo de regresión lineal múltiple	13
1.2.1. Estimación	14
1.2.2. Inferencia	17
1.2.3. Análisis de regresión múltiple con variables independientes dicotómicas	19
2. Modelo de diferencias en diferencias	21
2.1. Planteamiento	21
2.2. Supuesto de tendencias paralelas	25
2.3. Modelo a estimar	26
3. Datos y métodos	29
3.1. Descripción de la ENOE	29
3.1.1. Composición de los beneficiarios	30
3.2. Muestras de estudio para la evaluación de impacto	33
3.2.1. Definición de las variables utilizadas en la modelación	34
3.2.2. Estadística descriptiva de la muestra del Panel 1	36
3.2.3. Estadística descriptiva de la muestra del Panel 2	37
4. Resultados	43
4.1. Resultados del panel 1	43
4.2. Resultados del panel 2	49

4.3. Discusión	52
4.4. Comprobación del supuesto de tendencias paralelas	54
5. Conclusiones	65
Bibliografía	69

Introducción

El problema del empleo percibido a través del desempleo, del subempleo, y en general, de la dificultad de acceder a un trabajo digno se ha convertido en un tema de gran interés debido a que los costos de dichos problemas son altos a nivel individual y social. Los costos se miden no solo en términos de ingresos, sino que incluyen el descenso de la producción, la disminución de los niveles de actividad y el aumento de las divisiones sociales (Rubio et al. 2022).

De acuerdo con la Ley Federal del Trabajo, el empleo digno debe tener, entre otras características, el acceso a la seguridad social y a un salario que le permita al trabajador y a su familia vivir en condiciones decentes, cubriendo sus necesidades básicas como alimentación, salud, educación, etc. Los bajos ingresos aumentan la posibilidad de que los hogares sean pobres, además representan limitaciones para alcanzar estándares mínimos en distintos aspectos del bienestar (R. M. Sánchez et al. 2022). La dificultad para acceder a un buen empleo se debe entre otras razones a que los individuos no cuentan con la experiencia o las habilidades básicas que demanda actualmente el mercado laboral. Esto aumenta la probabilidad de ingresar a trabajos informales lo que conlleva a laborar en condiciones precarias, sin salarios justos, ni protección social y sin derechos laborales (Yacila 2021).

Una política que el gobierno de México ha implementado desde hace muchos años para mejorar las condiciones de empleo es brindar capacitación laboral para que las personas puedan adquirir las habilidades y los conocimientos que el mercado laboral demanda. En México han existido distintos programas federales dirigidos al ámbito del empleo. En 2005 como parte del Programa de Apoyo al Empleo (PAE), se desarrolló el subprograma BÉCATE cuyo objetivo consistía en incorporar a la población desempleada y subempleada de 16 años o más a cursos de capacitación laboral de corto plazo con la finalidad de que los participantes adquirieran o fortalecieran sus conocimientos y habilidades para incrementar sus posibilidades de incorporarse al mercado laboral. El proceso de capacitación duraba entre 1 y 3 meses, durante este proceso el beneficiario recibía una beca económica de uno a tres salarios mínimos¹ (Aparicio 2006; Cervantes 2011). Este programa estuvo vigente hasta el año 2018, y contaba con las distintas modalidades de

¹El pago de la beca se otorgaba mensualmente, según los días de asistencia a la capacitación, se proporcionaba además apoyo para transporte, así como un seguro de accidentes.

capacitación: mixta,² en la práctica laboral,³ al autoempleo⁴ y vales de capacitación.⁵

En 2019 entró en vigor el programa Jóvenes Construyendo el Futuro, el cual representa uno de los programas sociales prioritarios del actual gobierno federal. Dicho programa brinda capacitación gratuita en empresas y centros de trabajo a jóvenes de entre 18 y 29 años que no están estudiando o no cuentan con un trabajo. Los jóvenes participantes del programa reciben una beca mensual como apoyo económico durante su participación, y cuentan con un tutor o instructor que los guía y supervisa en su proceso de aprendizaje. El programa Jóvenes Construyendo el Futuro busca fomentar la inclusión social y laboral de los jóvenes, brindándoles oportunidades para adquirir experiencia, habilidades y conocimientos que les faciliten su inserción en el mercado laboral, a través de su participación en empresas, negocios, comercios, organizaciones sociales e instituciones públicas y proyectos productivos.

A pesar de la importancia que tienen los programas de capacitación se han realizado pocos estudios sobre el impacto de la capacitación laboral en México. Uno de estos estudios es el realizado por Van Gameren (2010), quien estima el impacto del programa BÉCATE y sus distintas modalidades, sobre la colocación de empleo, los ingresos y las prestaciones laborales (principalmente acceso a servicios de salud en los empleos obtenidos). El autor emplea registros administrativos del programa para el año 2008 y el primer semestre del 2009; para generar el grupo de comparación usa datos de la ENOE 2008 y 2009. El método de estimación que utiliza es el modelo de diferencias en diferencias combinado con emparejamiento por puntajes de propensión. De manera reciente, Rubio et al. (2022) evalúan el impacto del programa Jóvenes Construyendo el Futuro sobre el desempleo de jóvenes de 20 a 29 años, para ello usa datos estatales obtenidos de la ENOE y del Banco de Indicadores Económicos (BIE), ambos presentados por INEGI, así como un modelo basado en la Ley Okun y el modelo de diferencias en diferencias.

El presente trabajo abona a la escasa literatura sobre el tema de los efectos de la capacitación laboral en México. El objetivo es determinar cuál es la mejora en las condiciones laborales de los individuos que han sido beneficiarios de programas de becas de capacitación para el trabajo en términos de algunas variables de interés como el ingreso mensual, las horas de trabajo, las prestaciones laborales, las prestaciones de salud y la colocación de empleo. Específicamente, se estima el impacto de la becas de capacitación sobre el ingreso, las horas trabajadas, las prestaciones de salud y las prestaciones laborales sobre las personas ocupadas que fueron beneficiarios de las becas y el impacto en la colocación en el empleo para las personas buscadoras de empleo que

²Dirigida a la población desempleada, los cursos eran gestionados y realizados por el sector empresarial con el objetivo de que el desempleado obtuviera la calificación requerida por la empresa.

³Orientada principalmente para jóvenes entre 16 y 29 años con estudios técnicos o profesionales, con el objetivo de que obtuvieran experiencia laboral.

⁴Diseñada para personas con interés en el desarrollo de una actividad productiva por cuenta propia. Tenía la finalidad de mejorar los conocimientos técnicos y administrativos relacionados con el trabajo.

⁵Dirigida a personas que requerían mejorar o adquirir nuevas habilidades y destrezas laborales. El propósito consistía en actualizar y/o adquirir nuevas competencias que les permitan encontrar empleo de manera más efectiva

recibieron beca de capacitación. Contrario a Rubio et al. (2022) y Van Gameren (2010), en este trabajo se busca estimar el efecto de la recepción de becas de capacitación laboral sin distinción sobre el tipo de programa del cual reciben la beca, además se emplean datos desde el primer trimestre del año 2005 al primer trimestre de 2020.

Para efectuar la estimación del impacto de las becas de capacitación laboral sobre las variables de interés se utilizan datos provenientes de la Encuesta Nacional de Ocupación y Empleo (ENOE) elaborada por INEGI y se emplea la metodología de diferencias en diferencias. Los resultados obtenidos muestran un efecto positivo de las becas de capacitación laboral en las variables ingreso mensual, horas de trabajo y prestaciones laborales; el efecto en la variable prestaciones de salud es significativo únicamente en el largo plazo, es decir al año de la recepción de becas. En cuanto a la colocación de empleo, se encuentra que la probabilidad de acceder a un empleo se reduce para aquellos que fueron beneficiarios de las becas.

El trabajo se estructura de la siguiente manera: en el primer capítulo se abordan los fundamentos matemáticos de los modelos de regresión lineal simple y múltiple. En el segundo capítulo se explica la metodología de diferencias en diferencias y se describe el modelo a utilizar para estimar el impacto de las becas de capacitación laboral. En el tercer capítulo se describen los datos y muestras de estudio empleadas. En el cuarto capítulo se presentan los resultados de las estimaciones, y finalmente, en el quinto capítulo se presentan las conclusiones y se mencionan las limitaciones del trabajo.

Capítulo 1

Análisis de regresión lineal

El objetivo de este capítulo es presentar los antecedentes matemáticos que se emplean para estimar el impacto de la recepción de las becas de capacitación laboral. Primero se describen los supuestos del modelo de regresión lineal simple, seguido del planteamiento del problema que se usa para estimar los parámetros del modelo y, posteriormente, se habla sobre la inferencia de los coeficientes del modelo. Finalmente, se explica de manera análoga el modelo de regresión lineal múltiple.

1.1. Modelo de regresión lineal simple

La mayor parte de esta sección está basada en el libro de Wooldridge (2010). El modelo de regresión lineal simple es un modelo estadístico utilizado para determinar la relación que existe entre alguna variable de respuesta y y una variable independiente o explicativa x . Este modelo también se emplea para hacer estimaciones sobre el comportamiento de y . El modelo de regresión lineal es

$$y = E(y|x) + u \quad (1.1)$$

Usualmente se asume que $E(y|x)$ es lineal, es decir,

$$y = \beta_0 + \beta_1 x + u. \quad (1.2)$$

donde los parámetros β_0 y β_1 son constantes desconocidas llamadas coeficientes de regresión y u es una variable llamada término de error que representa factores distintos a x que no son observados y que afectan a y .

Si los factores en u permanecen constantes entonces x tiene un efecto lineal sobre y . De modo que el cambio en y es simplemente β_1 multiplicado por el cambio en x . Esto significa que β_1 es el parámetro de la pendiente en la relación entre y y x , cuando todos los demás factores en u permanecen constantes y β_0 es el parámetro del intercepto.

1.1.1. Supuestos

A partir de una muestra aleatoria es posible obtener estimadores confiables de β_0 y β_1 , para ello es necesario hacer una suposición sobre el término de error u y su relación con la variable explicativa x , así como otros supuestos. Enseguida se mencionan los supuestos del modelo de regresión lineal simple.

Supuesto 1. Linealidad en los parámetros

Los parámetros β_0 y β_1 deben ser lineales. Sin embargo, la relación entre la variable de respuesta y y la variable explicativa x no necesariamente es lineal.

Supuesto 2. Variación muestral en la variable explicativa

Los valores muestrales de la variable explicativa x , $\{x_i, i = 1, \dots, n\}$ no son todos iguales.

Supuesto 3. Media condicional cero

Dado cualquier valor de la variable explicativa x , el valor esperado del término de error u es cero.

$$E(u|x) = 0$$

Esto indica que el valor promedio de los factores no observables es cero, de modo que en promedio, lo no observable no le afecta a la variable de resultado.

Supuesto 4. Homogeneidad de la varianza

Para cualquier valor de la variable explicativa x , el error u tiene la misma varianza.

$$Var(u|x) = \sigma^2$$

Esto indica que la variabilidad del término de error en torno a su media es constante en todos los valores de la variable explicativa.

Supuesto 5. Exogeneidad

La variable explicativa x no está correlacionada con u , es decir x es exógena.

$$Corr(x, u) := \frac{Cov(x, u)}{\sqrt{Var(x)Var(u)}} = 0$$

Esto significa que la variable explicativa no está relacionada con los factores no observables que afectan a la variable de resultado.

Supuesto 6. No autocorrelación

Dados dos valores cualesquiera de x , x_i y $x_j (i \neq j)$, la correlación entre los términos de error u_i y u_j es cero.

$$Cov(u_i, u_j) = 0$$

donde i y j son dos observaciones diferentes. El incumplimiento de este supuesto significa que las observaciones del término de error mantienen una relación.

Supuesto 7. Normalidad

El error u tiene distribución normal con media cero y varianza σ^2

$$u \sim Normal(0, \sigma^2)$$

1.1.2. Estimación de los parámetros

Dada una muestra aleatoria $\{(x_i, y_i) : i = 1, \dots, n\}$, se usa el método de mínimos cuadrados ordinarios para estimar los parámetros β_0 y β_1 . De acuerdo a la ecuación (1.2), para toda i puede escribirse

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (1.3)$$

Por lo que se puede considerar que la ecuación (1.1) es un modelo poblacional de regresión, mientras que la ecuación (1.3) es un modelo muestral de regresión.

Para todo $\hat{\beta}_0$ y $\hat{\beta}_1$ se define el valor ajustado para y cuando $x = x_i$ como

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

este es el valor que se predice para y cuando $x = x_i$ con valores dados del intercepto $\hat{\beta}_0$ y la pendiente $\hat{\beta}_1$.

El residual de la observación i es la diferencia entre el verdadero valor y_i y su valor ajustado \hat{y}_i

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$\hat{\beta}_0$ y $\hat{\beta}_1$ se eligen de manera que la suma de residuales cuadrados sea tan pequeña como sea posible. Es decir, se quiere hallar $\hat{\beta}_0, \hat{\beta}_1$ que resuelvan el siguiente problema

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n \hat{u}_i^2 = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Los estimadores por mínimos cuadrados ordinarios de β_0 y β_1 son:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (1.4)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.5)$$

en donde $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ y $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Proposición 1. *Bajo el supuesto 3 (Media condicional cero),*

$$E(\hat{\beta}_0) = \beta_0 \text{ y } E(\hat{\beta}_1) = \beta_1$$

Es decir, $\hat{\beta}_0$ es un estimador insesgado de β_0 y $\hat{\beta}_1$ es un estimador insesgado de β_1

Demostración. Veamos que el estimador de β_1 de la ecuación (1.5) se puede escribir como

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i - \sum_{i=1}^n (x_i - \bar{x})\bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.6)$$

Observe que

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0 \quad (1.7)$$

entonces (1.6) se puede escribir como

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.8)$$

sustituyendo el lado derecho de la ecuación (1.3) en la ecuación (1.8) se tiene

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.9)$$

empleando propiedades de la suma, (1.9) se puede escribir como

$$\hat{\beta}_1 = \frac{\beta_0 \sum_{i=1}^n (x_i - \bar{x}) + \beta_1 \sum_{i=1}^n x_i (x_i - \bar{x}) + \sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.10)$$

notemos que el denominador de la expresión en el lado derecho de la ecuación (1.10) se puede escribir como

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) \\ &= \sum_{i=1}^n x_i (x_i - \bar{x}) - \bar{x} \sum_{i=1}^n (x_i - \bar{x}) \\ &= \sum_{i=1}^n x_i (x_i - \bar{x}) - \bar{x}(0) \quad \text{por (1.7)} \\ &= \sum_{i=1}^n x_i (x_i - \bar{x}) \end{aligned}$$

de modo que

$$\begin{aligned} \hat{\beta}_1 &= \frac{\beta_1 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned} \quad (1.11)$$

Calculando el valor esperado de $\hat{\beta}_1$ condicionado sobre los valores muestrales de la variable independiente x , como $\sum_{i=1}^n (x_i - \bar{x})^2$ y $(x_i - \bar{x})$ son funciones sólo de las x_i , no son aleatorias bajo el condicionamiento. Por tanto, de acuerdo con la ecuación (1.11) y manteniendo implícito el condicionamiento sobre $\{x_1, x_2, \dots, x_n\}$, por propiedades del valor esperado, se tiene

$$\begin{aligned} E(\hat{\beta}_1) &= E(\beta_1) + E\left(\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) = \beta_1 + \frac{E(\sum_{i=1}^n (x_i - \bar{x})u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 + \frac{\sum_{i=1}^n E((x_i - \bar{x})u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})E(u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})(0)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta_1 \end{aligned}$$

donde se ha usado que el valor esperado de cada u_i condicionado sobre $\{x_1, x_2, \dots, x_n\}$ es cero (supuesto 3). Dado que esto es válido para cualesquiera valores de $\{x_1, x_2, \dots, x_n\}$ también es válido sin el condicionamiento de $\{x_1, x_2, \dots, x_n\}$.

Ahora, para probar que $E(\hat{\beta}_0) = \beta_0$, notemos primero que

$$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u}$$

entonces, sustituyendo en la ecuación (1.4) se tiene que

$$\hat{\beta}_0 = \beta_0 + \beta_1 \bar{x} + \bar{u} - \hat{\beta}_1 \bar{x}$$

así, condicionando sobre los valores de las x_i

$$E(\hat{\beta}_0) = E(\beta_0) + E(\beta_1) \bar{x} + E(\bar{u}) - E(\hat{\beta}_1) \bar{x} = \beta_0 + \beta_1 \bar{x} - E(\hat{\beta}_1) \bar{x}$$

ya que $E(\bar{u}) = 0$ por el supuesto 3. Además como $E(\hat{\beta}_1) = \beta_1$, entonces

$$E(\hat{\beta}_0) = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0$$

□

Por lo tanto, $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores insesgados de β_0 y β_1 , respectivamente. Es decir, que el valor esperado de todas las estimaciones que se puedan realizar de cada uno de los parámetros para cada muestra diferente, es igual al parámetro poblacional.

Además de saber que $\hat{\beta}_0$ y $\hat{\beta}_1$ son insesgados, también es importante saber qué tanto puede esperarse que $\hat{\beta}_1(\hat{\beta}_0)$ se aleje, en promedio, de $\beta_1(\beta_0)$. Esto permite elegir el mejor estimador de todos. La medida de la dispersión de la distribución de $\hat{\beta}_0$ y de $\hat{\beta}_1$ con la que es más fácil trabajar es con la varianza.

Con los supuestos 1 a 3 puede ser calculada la varianza de los estimadores de mínimos cuadrados ordinarios. Sin embargo, hacerlo resulta complicado, por lo que se considera adicionalmente el supuesto de homogeneidad de la varianza, el cual establece que la varianza de los factores no observables, u , condicionados a x , es constante.

Proposición 2. *Bajo el supuesto 4 (Homogeneidad de la varianza),*

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.12)$$

y

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.13)$$

Demostración. Partiendo de que por (1.11) $\hat{\beta}_1$ se puede escribir como

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.14)$$

si se condiciona sobre las x_i , entonces $(x_i - \bar{x})$ y $\sum_{i=1}^n (x_i - \bar{x})^2$ son no aleatorios. Además, como las u_i son variables aleatorias independientes para toda i (por el muestreo aleatorio), y usando propiedades de la varianza se tiene

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \frac{1}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \text{Var} \left(\sum_{i=1}^n (x_i - \bar{x})u_i \right) \\ &= \frac{1}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}(u_i) \right) \\ &= \frac{1}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2 \right) \\ &= \sigma^2 \frac{1}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

donde se ha usado que $\text{Var}(u_i) = \sigma^2$ para toda i , por el supuesto 4.

Para $\hat{\beta}_0$, notemos que

$$\hat{\beta}_0 = \beta_0 + \beta_1 \bar{x} + \bar{u} - \hat{\beta}_1 \bar{x} = \beta_0 + \bar{u} - (\hat{\beta}_1 - \beta_1) \bar{x}$$

condicionando sobre los valores de las x_i y por propiedades de la varianza se tiene

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{u}) + \bar{x}^2 \text{Var}(\hat{\beta}_1 - \beta_1) - 2\bar{x} \text{Cov}(\hat{\beta}_1 - \beta_1, \bar{u})$$

notemos que

$$\text{Var}(\bar{u}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(u_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

$$\text{Var}(\hat{\beta}_1 - \beta_1) = \text{Var}(\hat{\beta}_1)$$

y

$$\text{Cov}(\hat{\beta}_1 - \beta_1, \bar{u}) = E[(\hat{\beta}_1 - \beta_1)\bar{u}] - E(\hat{\beta}_1 - \beta_1)E(\bar{u}) = E[(\hat{\beta}_1 - \beta_1)\bar{u}] - 0$$

por (1.14) se tiene que

$$E[(\hat{\beta}_1 - \beta_1)\bar{u}] = E\left[\frac{\sum_{i=1}^n (x_i - \bar{x})u_i\bar{u}}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = \frac{\sum_{i=1}^n (x_i - \bar{x})E(u_i\bar{u})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

dado que las $u_i, i = 1, \dots, n$ son independientes y que u_i^2 tiene distribución ji-cuadrada con 1 grado de libertad ($u_i^2 \sim \chi_1^2$), entonces $E(u_i, u_j) = 0$ y $E(u_i^2) = 1$, así

$$E[(\hat{\beta}_1 - \beta_1)\bar{u}] = \frac{\sum_{i=1}^n (x_i - \bar{x})(1)}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0 \quad \text{por 1.7}$$

Por lo tanto

$$\begin{aligned} \text{Var}(\hat{\beta}_0) &= \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \\ &= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

□

Las fórmulas en (1.12) y (1.13) son desconocidas, ya que la varianza del término de error, σ^2 , no se conoce, sin embargo, usando los datos puede estimarse y de este modo se puede estimar la varianza de $\hat{\beta}_0$ y $\hat{\beta}_1$. El estimador insesgado de σ^2 que se usa en la regresión lineal es

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2$$

Si en (1.12) y (1.13) se sustituye $\hat{\sigma}^2$ se obtienen estimadores insesgados de $\text{Var}(\hat{\beta}_0)$ y $\text{Var}(\hat{\beta}_1)$.

El estimador natural de σ es

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

$\hat{\sigma}$ es una estimación de la desviación estándar de los factores no observables que afectan a y , y es usada para estimar las desviaciones estándar de $\hat{\beta}_0$ y $\hat{\beta}_1$.

Las desviaciones estándar de $\hat{\beta}_0$ y $\hat{\beta}_1$ se denotan por $de(\hat{\beta}_0)$ y $de(\hat{\beta}_1)$ y están dadas por

$$de(\hat{\beta}_0) = \frac{\sigma \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2}} \quad \text{y} \quad de(\hat{\beta}_1) = \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

y sus estimaciones son

$$ee(\hat{\beta}_0) = \frac{\hat{\sigma} \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2}} \quad \text{y} \quad ee(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$ee(\hat{\beta}_j)$ denota el error estándar de $\hat{\beta}_j$, este es el estimador de la desviación estándar de $\hat{\beta}_j$, para $j = 0, 1$.

1.1.3. Inferencia sobre los parámetros

Para hacer pruebas de hipótesis acerca de los parámetros del modelo de regresión se necesita determinar la distribución de los estimadores de mínimos cuadrados ordinarios. La distribución muestral de los estimadores de mínimos cuadrados ordinarios depende de la distribución de los errores. Para hacer manejables las distribuciones de muestreo de β_0 y β_1 se supone que el error u está distribuido normalmente. Esto se conoce como supuesto de normalidad (supuesto 7).

La normalidad del término de error se traduce en una distribución normal de los estimadores de mínimos cuadrados ordinarios:

$$\hat{\beta}_0 \sim Normal(\beta_0, Var(\hat{\beta}_0)) \quad \text{y} \quad \hat{\beta}_1 \sim Normal(\beta_1, Var(\hat{\beta}_1))$$

por lo tanto, si $Var(\hat{\beta}_0)$ y $Var(\hat{\beta}_1)$ fueran conocidas, por propiedades de la distribución normal, se tendría que

$$(\hat{\beta}_j - \beta_j)/de(\hat{\beta}_j) \sim Normal(0, 1) \quad j = 0, 1$$

donde $de(\hat{\beta}_j)$ es la desviación estándar de $\hat{\beta}_j$, la cual no se conoce, por lo que se emplea el estimador de la desviación estándar de $\hat{\beta}_j$, es decir $ee(\hat{\beta}_j)$. Notemos que

$$(\hat{\beta}_j - \beta_j)/ee(\hat{\beta}_j) = \frac{(\hat{\beta}_j - \beta_j)/de(\hat{\beta}_j)}{\sqrt{\hat{\sigma}^2/\sigma^2}} = \frac{(\hat{\beta}_j - \beta_j)/de(\hat{\beta}_j)}{\sqrt{\frac{(n-2)\hat{\sigma}^2/\sigma^2}{(n-2)}}}$$

como $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2$ y $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n$ provienen de una distribución $N(0, \sigma^2)$, entonces $(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi_{n-2}^2$ (distribución ji-cuadrada con $n-2$ grados de libertad). Así, por la definición de una variable aleatoria con distribución t (ver en el libro de Hogg et al. 2019) se tiene que

$$(\hat{\beta}_j - \beta_j)/ee(\hat{\beta}_j) \sim t_{n-2}$$

es decir, que el cociente entre $\hat{\beta}_j - \beta_j$ y $ee(\hat{\beta}_j)$ tiene distribución t con $n - 2$ grados de libertad.

Este resultado es importante porque permite probar hipótesis en las que intervienen las β_j . El principal interés es probar la hipótesis nula

$$H_0 : \beta_j = 0$$

contra la hipótesis alternativa

$$H_1 : \beta_j \neq 0$$

donde $j = 0, 1$.

El estadístico que se utiliza para probar H_0 contra H_1 se llama el estadístico t y se define como

$$t_{\hat{\beta}_j} = \hat{\beta}_j / ee(\hat{\beta}_j)$$

La regla de rechazo depende del nivel de significancia elegido para la prueba. Los niveles de significancia más empleados son 1, 5 y 10 %. Si el nivel de significancia es α , la región de rechazo es

$$|t_{\beta_j}| > c$$

donde c es un valor crítico que se elige de manera que el área en cada cola de la distribución t con $n - 2$ grados de libertad sea igual a $\alpha/2$.

Cuando se rechaza H_0 en favor de H_1 al nivel de α , se dice que β_j es estadísticamente significativo, o estadísticamente distinto de cero, a un nivel de significancia de α %. Cuando $j = 1$ y se rechaza H_0 en favor de H_1 al nivel de α , se dice que la variable independiente está significativamente relacionada con la variable dependiente.

1.2. Modelo de regresión lineal múltiple

Un modelo de regresión lineal donde intervienen más de una variable explicativa se llama modelo de regresión lineal múltiple. En general, se puede relacionar a la variable de respuesta y con k variables explicativas. El modelo de regresión lineal múltiple con k variables explicativas se expresa como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u \quad (1.15)$$

donde

β_0 es el intercepto.

β_j es el parámetro asociado con x_j , $j = 1, \dots, k$

Cada parámetro β_j , $j = 1, 2, \dots, k$ representa el cambio esperado en la respuesta y por cambio unitario en x_j cuando todas las demás variables explicativas x_i ($i \neq j$) se mantienen constantes.

1.2.1. Estimación

Dada una muestra de n observaciones, se emplea el método de mínimos cuadrados para estimar los coeficientes $\beta_0, \beta_1, \dots, \beta_k$. La ecuación (1.15) puede escribirse en términos de las observaciones como:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i, \quad i = 1, \dots, n \\ &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + u_i \end{aligned}$$

La representación matricial es la siguiente

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (1.16)$$

donde

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

En general \mathbf{y} es un vector de dimensión n , \mathbf{X} es una matriz de dimensión $n \times (k+1)$, $\boldsymbol{\beta}$ es el vector de los coeficientes de regresión de dimensión $(k+1)$ y \mathbf{u} es el vector de los errores de dimensión n .

El método de mínimos cuadrados consiste en elegir $\boldsymbol{\beta}$ de tal modo que la suma de los cuadrados de los errores, L , se minimice

$$L = \sum_{i=1}^n u_i^2 = \mathbf{u}^T \mathbf{u} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}$$

de modo que los estimadores de mínimos cuadrados deben satisfacer que

$$\left. \frac{\partial L}{\partial \boldsymbol{\beta}} \right|_{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} \frac{\partial L(\hat{\boldsymbol{\beta}})}{\partial \beta_0} \\ \frac{\partial L(\hat{\boldsymbol{\beta}})}{\partial \beta_1} \\ \vdots \\ \frac{\partial L(\hat{\boldsymbol{\beta}})}{\partial \beta_k} \end{bmatrix} = 0$$

esto es

$$\left. \frac{\partial L}{\partial \boldsymbol{\beta}} \right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = 0$$

de donde se obtiene

$$\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y}$$

si se asume que la matriz $\mathbf{X}^T \mathbf{X}$ de dimensión $(k+1) \times (k+1)$ es no singular, se obtiene que el estimador de mínimos cuadrados de $\boldsymbol{\beta}$ es

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

El supuesto de que $\mathbf{X}^T \mathbf{X}$ es invertible es equivalente al supuesto de que la matriz \mathbf{X} tiene rango $k+1$, lo que significa que las columnas de \mathbf{X} deben ser linealmente independientes.

De este modo el modelo de regresión ajustado es

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}}$$

El residuo de la observación i es $\hat{u}_i = y_i - \hat{y}_i$, así el vector de los residuales es

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$$

Los supuestos para el modelo de regresión lineal múltiple son los siguientes

Supuesto 1. Linealidad en los parámetros

Los parámetros β_0 y β_1, \dots, β_k deben ser lineales. Sin embargo, la relación entre la variable de respuesta y y las variables independientes x_1, \dots, x_k no necesariamente es lineal.

Supuesto 2. Variación muestral en la variable explicativa

Los valores muestrales de las $x_j, i = 1, \dots, k$ no son todos iguales.

Supuesto 3. No colinealidad perfecta

No existe relación lineal exacta entre ninguna de las variables independientes.

La versión matricial de este supuesto es:

$$\text{El rango de } \mathbf{X} \text{ es } (k+1)$$

Supuesto 4. Media condicional cero

Condicionando sobre la matriz \mathbf{X} , cada error u_i tiene una media cero

$$E(u_i | \mathbf{X}) = 0, \quad i = 1, 2, \dots, n$$

En la forma vectorial esto se puede escribir como

$$E(\mathbf{u} | \mathbf{X}) = \mathbf{0}$$

Supuesto 5. Homogeneidad de la varianza y no autocorrelación

a) La varianza de u_i no depende de ningún elemento de \mathbf{X} y es constante en todas las observaciones

$$\text{Var}(u_i | \mathbf{X}) = \sigma^2, \quad i = 1, 2, \dots, n$$

b) Los errores no se correlacionan a través de las observaciones

$$\text{Cov}(u_i, u_j | \mathbf{X}) = 0$$

donde i y j son dos observaciones diferentes.

Estos dos supuestos se pueden escribir de forma matricial como

$$\text{Var}(\mathbf{u} | \mathbf{X}) = \sigma^2 \mathbf{I}_n$$

donde \mathbf{I}_n es la matriz identidad $n \times n$.

Supuesto 6. Normalidad

Condicionadas a \mathbf{X} , las u_i son independientes y se distribuyen como $Normal(0, \sigma^2)$. De modo que \mathbf{u} dada \mathbf{X} se distribuye como normal multivariada

$$\mathbf{u} \sim Normal(0, \sigma^2 \mathbf{I}_n)$$

Proposición 3. *Bajo el supuesto 4 (Media condicional cero),*

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

es decir, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ son estimadores insesgados.

Demostración. Por (1.16) podemos escribir a $\hat{\boldsymbol{\beta}}$ como

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X} \boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \\ &= \boldsymbol{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \end{aligned}$$

por lo tanto, al calcular el valor esperado de $\hat{\boldsymbol{\beta}}$ condicionando implícitamente sobre \mathbf{X} se tiene

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E(\boldsymbol{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}) \\ &= E(\boldsymbol{\beta}) + E((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}) \\ &= \boldsymbol{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{u}) \\ &= \boldsymbol{\beta} \end{aligned}$$

donde se ha usado que el valor esperado de \mathbf{u} condicionado sobre \mathbf{X} es cero por el supuesto de media condicional cero. Como esto es válido para cualquier valor de \mathbf{X} también es válido sin el condicionamiento. Por lo tanto $\hat{\boldsymbol{\beta}}$ es insesgado. □

Proposición 4. *Bajo el supuesto 5 (Homogeneidad de la varianza y no autocorrelación),*

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

Demostración. Por la definición de varianza de un vector se tiene

$$\text{Var}(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^T]$$

por el resultado anterior y condicionando implícitamente sobre \mathbf{X} se obtiene

$$\begin{aligned} \text{Var}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\ &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}^T] \\ &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \mathbf{u}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] \\ &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \mathbf{u}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{u} \mathbf{u}^T) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned}$$

dado que $\text{Var}(\mathbf{u}|\mathbf{X}) = E(\mathbf{u} \mathbf{u}^T | \mathbf{X}) = \sigma^2 \mathbf{I}$ por el supuesto 5, se tiene que

$$\begin{aligned} \text{Var}(\hat{\beta}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{I} \sigma^2 \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned}$$

□

$(\mathbf{X}^T \mathbf{X})^{-1}$ es una matriz de dimensión $(k+1) \times (k+1)$

$$(\mathbf{X}^T \mathbf{X})^{-1} = \{c_{ij}\}, i, j = 0, \dots, k$$

cuyos elementos en la diagonal multiplicados por σ^2 son las varianzas de los $\hat{\beta}_j$

$$\text{Var}(\hat{\beta}_j) = \sigma^2 c_{jj}$$

El estimador insesgado de σ^2 en el caso de la regresión lineal múltiple es

$$\hat{\sigma}^2 = \frac{1}{n - (k+1)} \left(\sum_{i=1}^n \hat{u}_i \right)$$

1.2.2. Inferencia

Prueba de significancia individual

La prueba de hipótesis para la significancia de cualquier coeficiente individual de la regresión es

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

donde j corresponde a cualquiera de las k variables independientes.

El estadístico de prueba es

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{ee(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}}$$

Y la región de rechazo es

$$|t_{\hat{\beta}_j}| > c$$

donde c es un valor crítico que se elige de manera que el área en cada cola de la distribución t con $n - (k + 1)$ grados de libertad sea igual a $\alpha/2$, cuando el nivel de significancia establecido es $\alpha\%$. Cuando se rechaza H_0 en favor de H_1 al nivel de $\alpha\%$, se suele decir que x_j es estadísticamente significativa al nivel de $\alpha\%$.

Prueba de significancia conjunta

La prueba de significancia conjunta se emplea para determinar si el conjunto de variables explicativas tiene efecto sobre la variable dependiente. Las hipótesis son:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ para al menos un } \beta_j$$

El estadístico de prueba es

$$F = \frac{\frac{SC}{k}}{\frac{SRC}{(n-k-1)}} \quad (1.17)$$

donde SC es la suma de cuadrados de la regresión que se define como

$$SC = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Y SRC es la suma de residuales cuadrados que se define como

$$SRC = \sum_{i=1}^n \hat{u}_i^2$$

El estadístico en (1.17) se distribuye como una variable aleatoria F con $(k, n - k - 1)$ grados de libertad. Esto se escribe

$$F \sim F_{k, n-k-1}$$

Se puede demostrar que la ecuación (1.17) resulta del cociente de dos variables aleatorias independientes ji-cuadradas, divididas entre sus respectivos grados de libertad. La variable aleatoria ji-cuadrada del numerador tiene k grados de libertad y la variable aleatoria ji-cuadrada del denominador tiene $n - k - 1$ grados de libertad. Esta es la definición de una variable aleatoria con distribución F . Más detalles sobre este resultado se pueden encontrar en el libro de Wooldridge (2010)

Si se elige un nivel de significancia de $\alpha\%$, sea c el percentil $100 - \alpha$ en la distribución $F_{k, n-k-1}$, entonces H_0 se rechaza a favor de H_1 al nivel de significancia elegido si

$$F > c$$

El rechazo de H_0 implica que al menos una de las variables independientes x_1, x_2, \dots, x_k contribuye de manera significativa al modelo.

Con esta prueba no es posible saber cuál o cuáles de las variables tienen efecto sobre y . Cuando se rechaza H_0 se dice que x_1, x_2, \dots, x_k son estadísticamente significativas conjuntamente. En ocasiones las variables pueden no resultar significativas individualmente, pero al emplear una prueba F se puede determinar si como grupo tienen un efecto.

1.2.3. Análisis de regresión múltiple con variables independientes dicotómicas

En ocasiones es de interés incluir variables cualitativas en los modelos de regresión. Una variable cualitativa es una variable que expresa distintas cualidades, características o categorías. Una variable dicotómica es un tipo de variable cualitativa que solo toma dos valores posibles, estos valores comúnmente son cero, como ausencia y uno, como presencia de determinada cualidad. En realidad, estos valores son arbitrarios, la ventaja de usar una variable que solo tome ceros y unos es que esto conduce a un modelo de regresión en el que los parámetros tienen fáciles interpretaciones.

Si z es una variable dicotómica esta simplemente se agrega a la ecuación como una variable independiente.

$$y = \beta_0 + \delta z + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

el parámetro δ tiene la interpretación siguiente: δ es la diferencia promedio en la variable de resultado entre un grupo que tiene la cualidad y un grupo que no la tiene, manteniendo a las demás variables constantes.

Las variables binarias se usan también en caso de tener categorías múltiples. Sea z una variable categórica con g grupos o categorías, podemos suponer entonces que z toma los valores: $1, 2, \dots, g$. Para estimar las diferencias entre los g grupos en el modelo de regresión, se define un conjunto de $g - 1$ variables binarias de la siguiente manera:

$$z_1 = \begin{cases} 1 & \text{si } z = 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$z_2 = \begin{cases} 1 & \text{si } z = 2 \\ 0 & \text{en otro caso} \end{cases}$$

$$z_{g-1} = \begin{cases} 1 & \text{si } z = g - 1 \\ 0 & \text{en otro caso} \end{cases}$$

Al grupo para el que no se crea una variable dicotómica se le llama grupo base o grupo de referencia. El grupo de base es arbitrario, pero en este caso y con el fin de simplificar, el grupo base que se ha elegido es el grupo para el cual $z = g$. Las variables dicotómicas z_1, z_2, \dots, z_{g-1} se incluyen al modelo de regresión lineal como variables independientes

$$y = \beta_0 + \delta_1 z_1 + \delta_2 z_2 + \dots + \delta_{g-1} z_{g-1} + \text{otros factores}$$

Los coeficientes tienen una interpretación sencilla: δ_1 es la diferencia promedio en la variable de resultado (permaneciendo los demás factores constantes) entre una unidad de observación perteneciente al grupo 1 y una perteneciente al grupo g ; δ_2 es la diferencia promedio en la variable de resultado entre una unidad de observación perteneciente al grupo 2 y una perteneciente al grupo g , etc. Todas las estimaciones de las variables binarias se interpretan con relación al grupo base.

Capítulo 2

Modelo de diferencias en diferencias

En este capítulo se presenta el modelo de diferencias en diferencias, se habla sobre el supuesto de tendencias paralelas y sobre las diversas formas de probar este supuesto. Posteriormente, se plantea el modelo que se emplea para estimar el impacto de la recepción de becas laborales sobre el ingreso, la horas de trabajo, las prestaciones y la colocación de empleo.

2.1. Planteamiento

La evaluación de impacto es un tipo particular de evaluación que tiene la finalidad de responder a una pregunta específica de causa y efecto: ¿Cuál es el impacto (o efecto causal) de un programa o fenómeno sobre un resultado de interés? De acuerdo con Bernal y Peña (2011), una evaluación de impacto tiene el propósito de medir el impacto de un programa o una política social sobre un conjunto de variables de resultado observadas en un conjunto de individuos. Estas variables de resultado son sobre las cuales se espera que el programa tenga un efecto.

En la práctica, la evaluación de impacto requiere dos grupos: uno de tratamiento y uno de control. El grupo de tratamiento fue afectado por el programa. El grupo de control no fue afectado por el programa, pero tiene similares características al grupo de tratamiento. El grupo de control representará lo que hubiera pasado con el grupo de tratamiento si no hubiera recibido el programa. La diferencia entre estos dos grupos nos indicaría el efecto del programa

El modelo de diferencias en diferencias es uno de los métodos de evaluación de impacto más utilizados. El planteamiento del método de diferencias en diferencias requiere observar a los grupos de tratamiento y control en al menos dos momentos distintos del tiempo: antes ($D = 0$) y después ($D = 1$) de la intervención del programa. Lo anterior permite comparar los cambios en la variable de resultados a lo largo del tiempo entre el grupo de tratamiento y el grupo de

comparación. Esto permite controlar por cualquier diferencia en características observables y no observables entre los grupos que sea constante a lo largo del tiempo.

El término “diferencias en diferencias” hace referencia a que este método calcula el impacto de la intervención de un programa como una diferencia entre dos diferencias. Específicamente, sea Y la variable de resultado sobre la que se estima el impacto del programa y sea T la variable indicadora del tratamiento, es decir, T toma el valor 1 si el individuo recibe el tratamiento y 0 en caso contrario.

Considerando a $Y_0 = Y(D = 0)$ como la observación de Y en el periodo anterior al tratamiento y a $Y_1 = Y(D = 1)$ como aquella en el periodo después del tratamiento, la primera diferencia de este procedimiento es la diferencia esperada en Y entre el periodo posterior y el periodo anterior a la implementación del programa en el grupo de tratamiento, $E(Y_1|T = 1) - E(Y_0|T = 1)$; la segunda sería la diferencia esperada en Y en el grupo de control entre el periodo posterior y anterior a la intervención del programa, $E(Y_1|T = 0) - E(Y_0|T = 0)$. Finalmente, a la primera diferencia se le resta la segunda, con esto, la metodología de diferencia en diferencias calcula qué tanto cambió en el tiempo el grupo de tratamiento respecto a qué tanto cambió en el tiempo el grupo de comparación. El impacto del programa por el método de diferencia en diferencias estaría dado por:

$$\gamma = [E(Y_1|T = 1) - E(Y_0|T = 1)] - [E(Y_1|T = 0) - E(Y_0|T = 0)]$$

Esto se puede observar gráficamente en la figura 2.1

El impacto se puede estimar usando una representación con base en el modelo de regresión lineal. Para ello se realizan dos regresiones: una sobre la variable de resultado en el periodo anterior y la otra sobre la variable de resultado en el periodo posterior a la implementación del programa. Las regresiones a estimar son:

$$Y_0 = \alpha_0 + \alpha_1 T + \delta^T \mathbf{X} + u$$

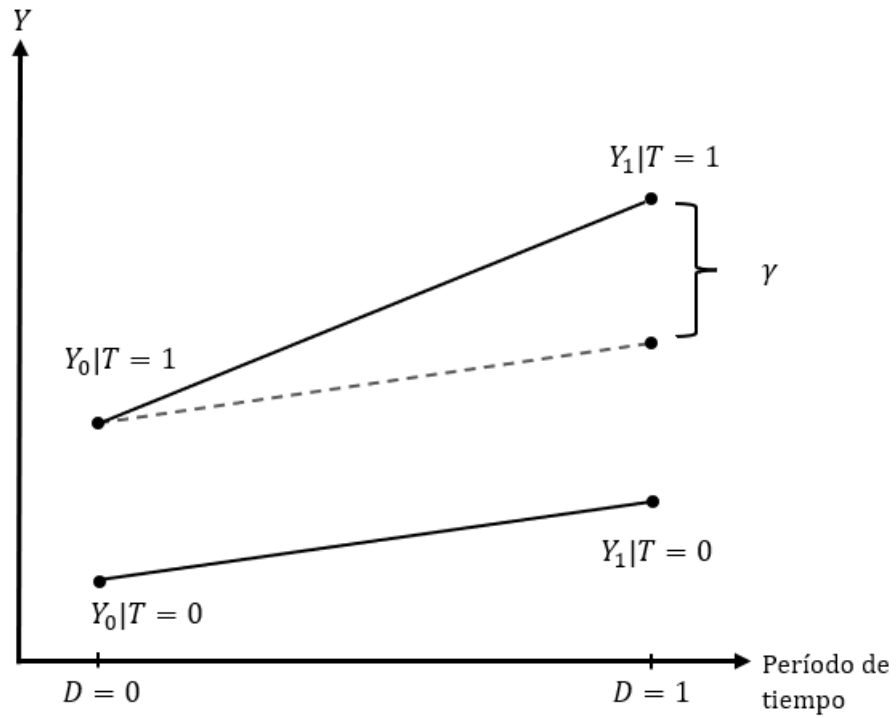
y

$$Y_1 = \beta_0 + \beta_1 T + \delta^T \mathbf{X} + u'$$

donde $\alpha_0, \alpha_1, \beta_0, \beta_1, \delta^T$ son los coeficientes de la regresión y u, u' son los términos de error, con los supuestos descritos en el capítulo previo y \mathbf{X} es un vector de características observables.

De este modo, el promedio en la variable de resultado en el grupo de tratamiento después de la intervención del programa es $E(Y_1|T = 1, \mathbf{X}) = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\delta}^T \mathbf{X}$, mientras que para este mismo grupo en el periodo anterior es $E(Y_0|T = 1, \mathbf{X}) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\delta}^T \mathbf{X}$. Análogamente, el promedio en la variable de resultado para el grupo de control después del programa es $E(Y_1|T = 0, X) = \hat{\beta}_0 + \hat{\delta}^T \mathbf{X}$, mientras que para el periodo anterior es $E(Y_0|T = 0, X) = \hat{\alpha}_0 + \hat{\delta}^T \mathbf{X}$.

Figura 2.1: Diferencias en diferencias



Fuente: elaboración propia

De modo que la diferencia en el grupo de tratamiento antes y después del programa es

$$[E(Y_1|T = 1, \mathbf{X}) - E(Y_0|T = 1, \mathbf{X})] = (\hat{\beta}_0 + \hat{\beta}_1) - (\hat{\alpha}_0 + \hat{\alpha}_1)$$

Y la diferencia en grupo de control antes y después del programa es

$$[E(Y_1|T = 0, \mathbf{X}) - E(Y_0|T = 0, \mathbf{X})] = \hat{\beta}_0 - \hat{\alpha}_0$$

Por lo tanto el efecto del programa esta dado por

$$\gamma = [E(Y_1|T = 1, \mathbf{X}) - E(Y_0|T = 1, \mathbf{X})] - [E(Y_2|T = 0, \mathbf{X}) - E(Y_1|T = 0, \mathbf{X})] = \hat{\beta}_1 - \hat{\alpha}_1 \quad (2.1)$$

Este mismo efecto se puede estimar usando una sólo regresión de la siguiente manera

$$Y = \alpha + \beta T + \lambda D + \gamma TD + \delta \mathbf{X} + u \quad (2.2)$$

donde

Y es la variable de resultado.

T es una variable dicotómica que toma el valor uno si el individuo participa en el programa (pertenecer al grupo de tratamiento) y el valor cero si el individuo no participa en el programa (pertenecer al grupo de control).

D es una variable dicotómica que indica el momento del tiempo en el que el individuo es observado, que podrá ser bien en un momento posterior al programa ($D = 1$) o bien en un momento previo ($D = 0$).

TD es la variable de interacción resultante de multiplicar las variables T y D .

\mathbf{X} es un vector de variables observables.

u es el término de error que cumple con los supuestos descritos en el capítulo previo.

Una vez estimados los coeficientes en (2.2) podemos identificar el efecto estimado del programa.

El promedio en la variable de resultado para el grupo de tratamiento antes de la intervención del programa es $E(Y|T = 1, D = 0, \mathbf{X}) = \hat{\alpha} + \hat{\beta} + \hat{\delta}\mathbf{X}$ y después de la intervención del programa es $E(Y|T = 1, D = 1, \mathbf{X}) = \hat{\alpha} + \hat{\beta} + \hat{\lambda} + \hat{\gamma} + \hat{\delta}\mathbf{X}$. Por lo tanto la diferencia en el grupo de tratamiento antes y después del programa es

$$(\hat{\alpha} + \hat{\beta} + \hat{\lambda} + \hat{\gamma} + \hat{\delta}\mathbf{X}) - (\hat{\alpha} + \hat{\beta} + \hat{\delta}\mathbf{X}) = \hat{\lambda} + \hat{\gamma}$$

Por otro lado, el promedio en la variable de resultado en el grupo de control antes de la intervención del programa es $E(Y|T = 0, D = 0, \mathbf{X}) = \hat{\alpha} + \hat{\delta}\mathbf{X}$ y después del programa es $E(Y|T = 0, D = 1, \mathbf{X}) = \hat{\alpha} + \hat{\lambda} + \hat{\delta}\mathbf{X}$. Por lo tanto la diferencia en el grupo de control antes y después del programa es

$$(\hat{\alpha} + \hat{\lambda} + \hat{\delta}\mathbf{X}) - (\hat{\alpha} + \hat{\delta}\mathbf{X}) = \hat{\lambda}$$

De modo que el efecto del programa está dado por

$$\hat{\lambda} + \hat{\gamma} - \hat{\lambda} = \hat{\gamma}$$

Entonces el coeficiente asociado a la variable de interacción de T y D mide el efecto del programa. Si $\hat{\gamma}$ es significativo, se puede decir que el programa tuvo un efecto estadísticamente significativo en la variable de resultado.

La ventaja de usar una sola regresión para conocer el efecto del programa, es que puede saberse de manera directa si el efecto es significativo, usando lo que se explicó en la subsección 1.2.2 sobre la inferencia de los parámetros de la regresión lineal múltiple, en cambio al realizar dos regresiones, debe calcularse el error estándar de la diferencia (2.1) para obtener el estadístico que permita probar la significancia del efecto, lo cual implica más trabajo. Además al hacer solo

una regresión no es necesario dividir la muestra en dos como se hace al utilizar dos regresiones, dividir la muestra puede afectar la normalidad en caso de que la muestra no sea lo suficientemente grande y esto a su vez puede afectar la validez y precisión del efecto.

2.2. Supuesto de tendencias paralelas

El modelo de diferencias en diferencias elimina las diferencias en características observables y no observables entre los grupos de tratamiento y comparación que son constantes a lo largo del tiempo, sin embargo no elimina aquellas diferencias que cambian con el tiempo. Para que el método proporcione una estimación válida, se debe suponer que no existen ese tipo de diferencias que varían en el tiempo entre los grupos de tratamiento y control. Por ello, el supuesto principal es que *en ausencia del programa, los resultados en el grupo que fue afectado por el programa hubieran tenido una tendencia igual a los del grupo de control*, es decir, sin el programa, los resultados tendrían que aumentar o disminuir en la misma medida en ambos grupos. Este supuesto se conoce como el supuesto de tendencias paralelas. La mayor parte de la información que se presenta en esta sección fue obtenida del libro de Gertler et al. (2017).

Ciertamente, no hay forma de demostrar que las diferencias entre los grupos de tratamiento y control habrían cambiado de manera paralela en ausencia del programa, por ello se debe suponer que, en ausencia del programa, los resultados en el grupo de tratamiento habrían cambiado de forma paralela con los resultados del grupo de control. Aunque este supuesto no se puede demostrar, existen maneras de verificar si es admisible que dicho supuesto se pudiera cumplir. En caso que no se cumpla este supuesto, el efecto estimado estaría sesgado.

Grupo de tratamiento falso

Una manera de verificar la validez del supuesto de las tendencias paralelas consiste en utilizar un grupo de tratamiento “falso” es decir, un grupo que no ha sido afectado por el programa. Para esta prueba, se realiza una segunda estimación de diferencias en diferencias, pero utilizando en este caso el grupo de tratamiento “falso”, dado que este grupo no es afectado por el programa, el impacto que se obtenga de la estimación debería ser no significativo, si resultara significativo, este impacto necesariamente se deberá a alguna diferencia en las tendencias entre el grupo de control y el grupo de tratamiento “falso” y esto conlleva a dudar sobre si es válido el supuesto de que los grupos de tratamiento y control tienen tendencias iguales en ausencia del programa.

Varios grupos de control

Otra manera de comprobar el supuesto de las tendencias iguales consiste en aplicar el método de diferencias en diferencias utilizando diferentes grupos de control. Si A y B son grupos de control, primero se realiza la estimación con el grupo A y después se realiza una segunda estimación con el grupo B . Lo que debería pasar para comprobar el supuesto de tendencias paralelas es que en

ambas estimaciones los impactos sean significativos y similares.

Efecto sobre otra variable de resultado

Una tercera manera de probar el supuesto de tendencias paralelas consiste en estimar el impacto del programa en un resultado falso, este es un resultado que, de acuerdo con lo que se sabe, no se ve afectado por el programa, si con la estimación del modelo de diferencias en diferencias se encuentra un efecto significativo del programa sobre dicho resultado, esto indica que de origen los grupos de tratamiento y control poseen diferentes características y que el modelo puede generar resultados falsos debido a factores no relacionados con el tratamiento.

Momentos previos al programa

Una cuarta opción para verificar la validez del supuesto es comparar los cambios en los resultados en los grupos de tratamiento y comparación en momentos (al menos dos) previos al comienzo del programa. Si se observa que los resultados evolucionaban de forma paralela antes del comienzo del programa, es razonable suponer que habrían seguido evolucionando de la misma manera después de la intervención del programa.

2.3. Modelo a estimar

El objetivo de este trabajo es estimar el impacto de las becas para la capacitación laboral sobre los ingresos, las horas trabajadas, las prestaciones de salud y las prestaciones laborales de las personas ocupadas que recibieron apoyos de becas de capacitación para el trabajo y sobre la colocación de empleo para los no ocupados que recibieron becas de capacitación laboral. A partir de los datos es posible observar a los grupos de tratamiento y control, antes y después de haber recibido la beca, por lo que se emplea un modelo de diferencias en diferencias de la siguiente manera

$$Y = \alpha + \beta Beca + \lambda Después + \gamma BecaDespués + \delta_1 \mathbf{X}_{CI} + \delta_2 \mathbf{X}_{CO} + u \quad (2.3)$$

donde

La variable dependiente Y representará en cada caso, el logaritmo del ingreso mensual, el número de horas de trabajo a la semana, una indicadora de que la persona ocupada recibe la prestación de salud por parte de su trabajo, una indicadora de que el ocupado recibe prestaciones laborales y una indicadora de que la persona desocupada obtuvo empleo.

$Beca$ es una variable dicotómica que toma el valor uno si la persona recibió beca y cero en otro caso.

$Después$ es una variable dicotómica que toma el valor uno en el periodo posterior al tratamiento

y cero en el periodo anterior.

\mathbf{X}_{CI} es el vector de características individuales tales como edad, género, escolaridad, estado civil y si vive en una localidad urbana

\mathbf{X}_{CO} es el vector de características de ocupación como el sector y la rama de ocupación. Este vector no se incluye para la estimación de la variable de resultado de colocación de empleo, ya que para esta estimación se seleccionó a la población económicamente activa que no tenía trabajo pero que se encontraba realizando acciones concretas de búsqueda.

u es el término de error, es una variable que recoge factores que no son observables y que pueden ser relevantes en la estimación, pero que en promedio valen cero.

Para cada variable de resultado el efecto estimado de interés está representado por γ .

Capítulo 3

Datos y métodos

En este capítulo se describe la fuente de origen de los datos. Enseguida se presenta la estadística descriptiva de las personas que reciben becas y se describen las variables que se emplean en este trabajo. Finalmente se especifican las diferentes muestras de estudio que se emplean para medir el impacto de las becas y se exhibe la respectiva estadística descriptiva.

3.1. Descripción de la ENOE

Los datos que se utilizan en el análisis provienen de la Encuesta Nacional de Ocupación y Empleo (ENOE) elaborada por el Instituto Nacional de Estadística y Geografía (INEGI). La ENOE es una encuesta trimestral y continua de esquema rotatorio, puesta en marcha en enero de 2005. Una de sus principales características es que se trata de un panel rotatorio en el que se reemplaza el 20 % de los entrevistados en cada trimestre. Los entrevistados seleccionados se visitan una vez cada trimestre durante cinco trimestres consecutivos y al término de este ciclo son sustituidos (Instituto Nacional de Estadística y Geografía (INEGI) 2023).

La ENOE emplea dos cuestionarios para recopilar información, el Cuestionario Sociodemográfico (CS) y el Cuestionario de Ocupación y Empleo (COE); por medio del primero, como lo indica su nombre, se recopilan datos sociodemográficos de los individuos seleccionados; y a través del segundo, se recopila información sobre la condición de actividad (Instituto Nacional de Estadística y Geografía (INEGI) 2023). A su vez el COE cuenta con dos versiones, una básica y una ampliada, la versión básica elimina algunas preguntas de la versión ampliada, esta última es la que recoge información sobre la recepción de becas a la capacitación para el trabajo. En cada ronda de entrevistas se aplican ambos cuestionarios (primero el CS seguido del COE); a partir del año 2009, únicamente en el primer trimestre de cada año se aplica la versión ampliada del COE.

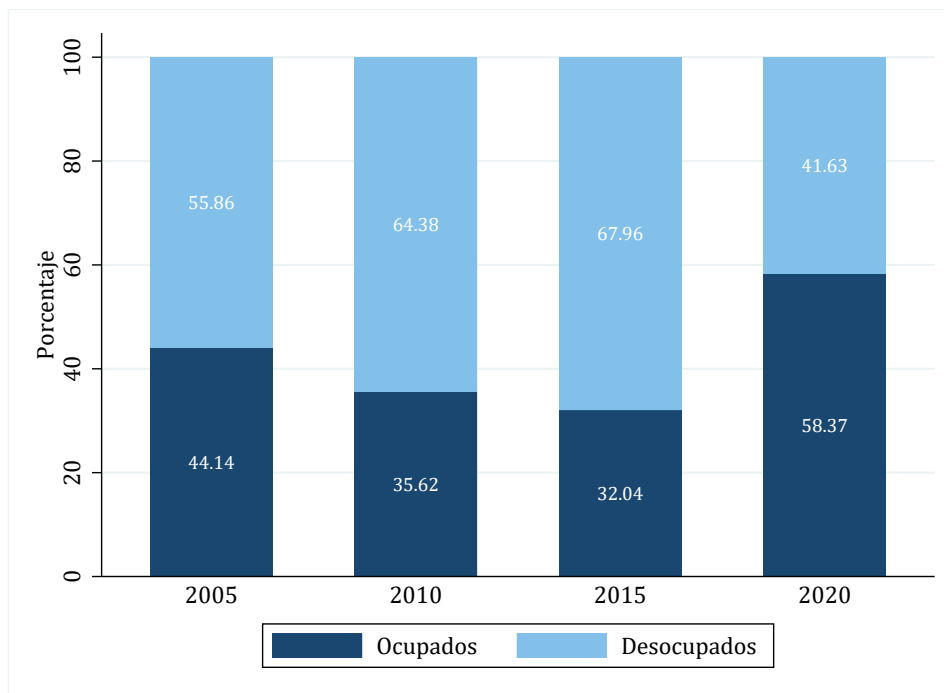
Con la finalidad de identificar la evolución en la composición del grupo de personas que han

recibido algún tipo de beca para la capacitación laboral, se construye una base de datos con datos de la ENOE del 2005 al 2020 a partir de la información del primer trimestre de cada año. La muestra de estudio en este caso es de personas de 16 años o más. Las características de interés son estatus de ocupación, género y edad.

3.1.1. Composición de los beneficiarios

En México los programas de becas para la capacitación para el trabajo han estado dirigidas principalmente a la población desempleada y subempleada.¹ De modo que la población beneficiaria está conformada por personas desocupadas y ocupadas. La figura 3.1 exhibe la composición de toda la población beneficiaria de la que se tiene registro, según la condición de la ocupación. En la figura se observa que en el año 2005, el 44 % de beneficiarios de becas estaban empleados, pero en el 2020 este porcentaje se incrementó a 58.4 %; en los años 2010 y 2015 se redujo el porcentaje de receptores de becas que también estaban ocupados.

Figura 3.1: Composición de la población beneficiaria de becas para la capacitación laboral según la condición de ocupación



Fuente: elaboración propia con datos de la ENOE. Información del primer trimestre de cada año. Muestra de personas de 16 años y más que recibieron beca de capacitación.

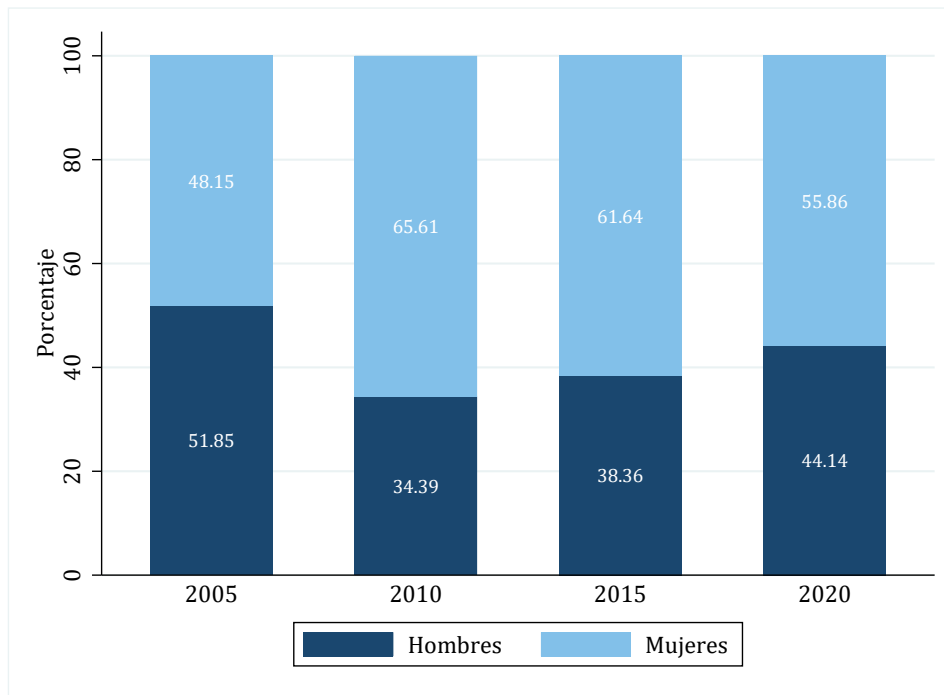
¹De acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI) (2023), la población subempleada o subocupada es aquella población ocupada que tiene la necesidad y disponibilidad de ofertar más tiempo de trabajo de lo que su ocupación actual le demanda.

En la Figura 3.2 se muestra la composición de la población beneficiaria de becas según el género; con excepción del 2005, se aprecia que la proporción de mujeres es mayor a la de hombres. En el año 2005, el 48 % de los beneficiarios de becas eran mujeres, para el año 2020 las mujeres representaron el 55.9 % de los beneficiarios. Encontrar trabajo es mucho más difícil para la mujer que para el hombre en toda América Latina, y cuando la mujer trabaja, suele estar ocupada en empleos informales, a tiempo parcial, con menor productividad y menor remuneración (Marchionni et al. 2019).

En cuanto a la composición por edad en la figura 3.3 se presenta esta información. Se observa que el grupo de los jóvenes de 16 a 29 años es el que representa el mayor porcentaje de los beneficiarios. En el año 2005 el grupo de los jóvenes representaba el 51.7 % de los beneficiarios de becas, este porcentaje ha ido en aumento a tal grado que, para el año 2020, este porcentaje fue del 86.5 %. Esto puede deberse a que los programas de apoyo al empleo en México han estado enfocados principalmente en la población joven, ya que el desempleo juvenil es uno de los retos más grandes a nivel mundial, los jóvenes se enfrentan a un mercado de trabajo que exige experiencia, y no les posibilita acceder a un primer empleo, encontrándose así en un círculo vicioso (A. Sánchez 2014). Otro punto a resaltar sobre el mayor porcentaje de becas recibidas por los jóvenes en el año 2020 es que esto se puede atribuir al más importante programa de becas a la capacitación para el trabajo implementado por el actual gobierno federal, el programa “Jóvenes construyendo el futuro”, el cual está dirigido específicamente a personas entre 18 y 29 años.

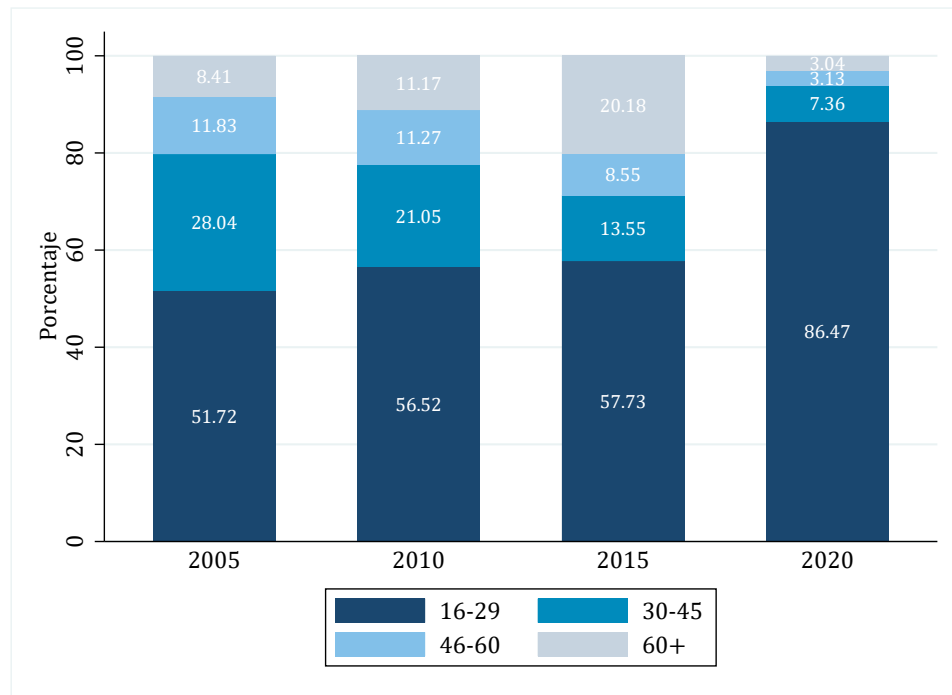
La precariedad laboral es un problema social, que incluye situaciones de desempleo, empleo inseguro, empleo con bajos ingresos, etc., y que afecta especialmente a los jóvenes y las mujeres (Vite 2018). Debido a esto, no es de extrañarse que sean los jóvenes y las mujeres quienes busquen acceder a programas sociales que les brinden apoyos económicos y a la vez les proporcionen herramientas para conseguir empleos dignos.

Figura 3.2: Composición de la población beneficiaria de becas para la capacitación laboral por sexo



Fuente: elaboración propia con datos de la ENOE. Información del primer trimestre de cada año. Muestra de personas de 16 años y más que recibieron beca de capacitación.

Figura 3.3: Composición de la población beneficiaria de becas para la capacitación laboral por grupo de edad



Fuente: elaboración propia con datos de la ENOE. Información del primer trimestre de cada año. Muestra de personas de 16 años y más que recibieron beca de capacitación.

3.2. Muestras de estudio para la evaluación de impacto

Los objetivos específicos de este trabajo son: estimar el impacto en los ingresos, las horas trabajadas y las prestaciones laborales de personas ocupadas que recibieron apoyos de becas de capacitación para el trabajo y estimar el impacto en la colocación de personas buscadoras de empleo que recibieron apoyos de becas. Para alcanzar los objetivos planteados y tomando en cuenta que la ENOE es una encuesta de tipo panel rotativo con periodicidad trimestral, entonces es posible observar a los mismos individuos durante 5 trimestres consecutivos,² pero únicamente se puede identificar si los individuos reciben algún tipo de beca a la capacitación para el trabajo en el primer trimestre de cada año. Dada esta restricción se construyen dos bases de datos con estructura tipo panel con rondas de observación trimestrales. Los dos paneles se construyen de la siguiente manera:

²Para poder identificar a un individuo en cada una de sus cinco entrevistas se utiliza un identificador individual construido con variables propias del individuo, estas variables son: ciudad autorrepresentada, entidad, identificación del control, número de vivienda seleccionada, número de hogar dentro de la vivienda, número de veces que el hogar se ha mudado, número de renglón del integrante del hogar.

Panel 1: Constituido por las cinco observaciones de la población cuya primera ronda de entrevistas se realizó en el trimestre 3 de cada año (por ende su quinta y última ronda fue realizada en el trimestre 3 del año siguiente). Para la construcción de este panel se utilizan las bases de datos desde el trimestre 3 de 2005 hasta el trimestre 3 de 2019.³ De modo que el panel 1 está compuesto por la unión de 14 subpaneles independientes (uno por cada año), los cuales, individualmente, observan a los mismos individuos durante cinco trimestres consecutivos, el primer subpanel contiene a los individuos entrevistados del trimestre 3 de 2005 al trimestre 3 de 2006; el segundo subpanel, a los entrevistados del trimestre 3 de 2006 al trimestre 3 de 2007; y así sucesivamente, hasta el subpanel número 14, el cual contiene a los individuos entrevistados del trimestre 3 de 2018 al trimestre 3 de 2019. Una ventaja de construir este panel de esta forma es que posibilita observar a los individuos dos trimestres antes y dos trimestres después de la recepción de becas.

Panel 2: Constituido por las cinco observaciones de la población cuya primera entrevista fue realizada en el trimestre 1 de cada año, de este modo podemos observar a los individuos por un periodo más largo, después de haber recibido la beca. Se utilizan las bases de datos del trimestre 1 de 2005 al trimestre 1 de 2020. Este panel está compuesto por la unión de 15 subpaneles independientes (uno por cada año), que observan a los mismos individuos durante cinco trimestres consecutivos, el primer subpanel contiene a los individuos entrevistados del trimestre 1 de 2005 al trimestre 1 de 2006; el segundo subpanel, a los entrevistados del trimestre 1 de 2006 al trimestre 1 de 2007; y así sucesivamente, hasta el subpanel número 15, el cual contiene a los individuos entrevistados del trimestre 1 de 2019 al trimestre 1 de 2020. La desventaja de la construcción de este panel es que no es posible observar a los individuos en periodos previos a la recepción de becas, una ventaja es que podemos observar a los individuos un año después de haber recibido la beca, lo cual nos permite observar el efecto de las becas de capacitación laboral en un plazo más largo.

3.2.1. Definición de las variables utilizadas en la modelación

Las variables de resultado para las cuales se presentará una estimación del impacto de las becas de capacitación al trabajo son: ingreso mensual, horas trabajadas, prestaciones de salud, prestaciones laborales y colocación de empleo. Enseguida se describe lo que cada una de las variables representa.

Ingreso mensual: Es el logaritmo natural del ingreso mensual en pesos convertido a precios constantes usando el Índice Nacional de Precios al Consumidor con base en el mes de diciembre del año 2022. El ingreso mensual es reportado por los individuos a partir de las preguntas: ¿Cada cuándo obtiene sus ingresos o le pagan? y ¿Cuánto ganó o en cuánto calcula sus ingresos?.

³No fue posible utilizar la información posterior al trimestre 3 de 2019 ya que el identificador individual sufrió modificaciones en su estructura en el trimestre 3 de 2020 y posteriormente en el trimestre 3 de 2021, lo cual imposibilita identificar a la población por 5 trimestres consecutivos.

Horas de trabajo: Es el número de horas trabajadas durante la última semana.

Prestaciones de salud: Esta variable de resultado indica si las personas ocupadas tienen acceso a atención médica por parte de su trabajo. Los sistemas de salud considerados son: seguro social (IMSS), ISSSTE, ISSSTE estatal (ISSSTELEON, ISSEMYM), hospital o clínica naval, militar o de Pemex, u otra institución médica. Esta es una variable dicotómica que toma el valor uno para aquellas personas que tienen acceso a algún sistema de salud y cero en otro caso.

Prestaciones: Esta variable considera las prestaciones laborales a las que tiene acceso el trabajador. La variable se codificó como una variable dicotómica, en donde el valor uno corresponde a las personas que declararon tener al menos una de las siguientes prestaciones en su trabajo: aguinaldo, vacaciones con goce de sueldo, o reparto de utilidades. Si no declararon tener alguna de las anteriores se imputó el valor cero.

Colocación de empleo: Esta variable se mide para aquellas personas que estaban desempleadas y que lograron obtener un empleo. Es una variable dicotómica que toma el valor uno si la persona reporta estar ocupada, y cero en otro caso.

Para estimar el impacto sobre el ingreso mensual, las horas de trabajo, las prestaciones de salud y las prestaciones laborales, se restringió la muestra en ambos paneles a las personas que reportaron estar ocupadas durante los cinco trimestres en los que son observadas, ya que estas variables de resultado únicamente se reportan para la población ocupada, pues se tratan de características que sólo presentan los individuos ocupados. En el caso de la estimación del impacto en la colocación de empleo se utilizó únicamente el panel 1, ya que para estimar dicho efecto se seleccionó a la población económicamente activa que no tenía trabajo pero que se encontraba realizando acciones concretas de búsqueda durante los dos trimestres anteriores al trimestre en el que reportó estar recibiendo beca. Realizar esta misma acción en el panel 2 no es posible debido a su estructura, ya que el primer trimestre en el que se observa a la población es el trimestre en el que se reporta la recepción de becas.

Se incluyen en la modelación algunas variables de características individuales como: género, edad, años de escolaridad, estado conyugal, tipo de localidad (urbana o rural). También se agregan variables de características de ocupación como el sector de actividad económica y la rama de actividad económica. La incorporación de estas variables se justifica ante la necesidad de considerar como controles a aquellos factores observables disponibles y que además han sido identificados en la bibliografía del tema como factores que intervienen en las condiciones laborales de un individuo (Muñoz, 2009; Román, 2013)

Las variables de control utilizadas se encuentran en la tabla 3.1, en la cual se muestra la definición y construcción de cada una de ellas.

Tabla 3.1: Definición de las variables de control

<i>Variable</i>	<i>Definición</i>
Edad	Se considera la edad en años cumplidos.
Hombre	Variable dicotómica que vale uno si el individuo es hombre y cero si es mujer.
Casado	Variable dicotómica construida con datos del estado conyugal. Toma el valor uno si el individuo declaró estar casado o vivir en unión libre; y el valor cero si dijo estar separado, divorciado, viudo o soltero.
Años escolaridad	Se consideran los años de estudio completados.
Localidad rural	Variable indicadora de que el tamaño de localidad de residencia del individuo es menor que 2,500 habitantes.
Sector de actividad económica	Variable categórica, cuyas categorías son: primario, secundario y terciario.
Rama de actividad económica	Variable categórica con categorías: agropecuario, construcción, industria manufacturera, comercio y servicio.

Nota: elaboración propia.

3.2.2. Estadística descriptiva de la muestra del Panel 1

La tabla 3.2 muestra las diferencias estimadas en las características individuales y de ocupación entre el grupo de los que recibieron beca (grupo de tratamiento) y el grupo de los que no recibieron beca (grupo de control), un trimestre previo a recibir la beca de capacitación (es decir, se emplean los datos de la segunda ronda de entrevistas). Los datos de la tabla corresponden a la muestra del panel 1 conformada por la población que estuvo ocupada durante las 5 rondas de entrevistas.

Al comparar las características individuales entre ambos grupos, se observa que, el 40% de los beneficiarios son hombres, mientras que, en el conjunto de los no beneficiarios, los hombres representan el 65%, de modo que el porcentaje de mujeres es mayor en el grupo de los que recibieron beca. Por otro lado, la edad promedio en el grupo de los que recibieron beca es mayor que en el grupo de los que no recibieron beca. Respecto a la escolaridad, estado civil y el tipo de localidad de residencia no se encontraron diferencias estadísticamente significativas entre los grupos. En cuanto a las características de ocupación, el porcentaje de personas que se desempeñan en activi-

dades agropecuarias y de comercio, es menor en el grupo de los que recibieron beca, mientras que el porcentaje de personas que se desempeñan en actividades de servicios es mayor en el grupo de los que recibieron beca.

La tabla 3.3 presenta información sobre la muestra conformada por la población económicamente activa que no tenía trabajo pero que se encontraba realizando acciones concretas de búsqueda durante los dos trimestres previos al recibimiento de beca. Se muestran las diferencias estimadas en las características entre el grupo de tratamiento y control en la segunda ronda de entrevistas. Respecto a las variables de género, edad, años de escolaridad y tipo de localidad de residencia no se observaron diferencias estadísticamente significativas en promedio entre los grupos. Los datos indican que el 78.4% de los beneficiarios están casados o viven en unión libre, mientras que en el grupo de los no beneficiarios el respectivo porcentaje es de 40.9%.

3.2.3. Estadística descriptiva de la muestra del Panel 2

Para analizar las características de la muestra del panel 2, en la tabla 3.4 se presentan las diferencias promedio en las características individuales y de ocupación entre el grupo de los que recibieron beca (grupo de tratamiento) y el grupo de los que no recibieron beca (grupo de control), en el primer trimestre a partir del cual son observadas las personas (ronda 1). Los datos de la tabla corresponden a la muestra conformada por la población que estuvo ocupada durante las 5 rondas en las que fue observada.

Hay mayor porcentaje de mujeres en el grupo de los beneficiarios. Respecto a la edad, el grupo de los individuos que recibieron beca tienen menor edad en promedio. En cuanto a los años de escolaridad, los individuos que recibieron beca presentan un mayor promedio de dicha variable relativo a los individuos que no recibieron beca. Sobre las características de ocupación, los datos señalan que el 14.2% de los beneficiarios se desempeñan en actividades del sector secundario, en el grupo de los no beneficiarios este porcentaje es 25.7%. También se observa que el 74.2% de los beneficiarios se desempeña en el sector terciario, mientras que en los no beneficiarios, el grupo de individuos en el sector terciario representa el 61.9%. El 0.2% de los beneficiarios se encuentra en la rama de la construcción, en los no beneficiarios el respectivo porcentaje es de 8.2%. Los datos también indican que el 11.1% de los beneficiarios se desempeñan en la rama del comercio y en los no beneficiarios este porcentaje es de 18.9%. La brecha en la rama de servicios es de 20.5 puntos porcentuales, siendo mayor para los beneficiarios.

En ambos paneles construidos se observan similitudes y diferencias en términos de las características observables. En ambos paneles el porcentaje de mujeres en el grupo de los beneficiarios es mayor al porcentaje de hombres; en la muestra del panel 2 los beneficiarios ocupados tienen menor edad que los no beneficiarios y sucede lo contrario en el panel 1. En los grupos de tratamiento y control en el panel 2 presentan diferencias significativas en más variables que en el panel 1. En comparación con el panel 2, en el panel 1 no es significativa la diferencia en el promedio

de los años de escolaridad, en el porcentaje de individuos que se desempeñan en actividades del sector secundario y terciario, y en el porcentaje de individuos que se emplean en la rama de construcción, sin embargo hay diferencia significativa en el porcentaje de individuos que se desempeñan en el sector primario y en la rama agropecuaria mientras que en el panel 2 no la hay.

En el análisis de la composición de los beneficiarios, sin restricción a los paneles, se observó que es mayor el porcentaje de mujeres en casi todos los años. Al restringir los datos a los paneles 1 y 2, también se observó que el porcentaje mujeres en el grupo de los beneficiarios es mayor.

Tabla 3.2: Diferencia de medias entre aquellos ocupados que tienen beca y no tienen beca. Panel 1.

Variable	Beca	Sin beca	Diferencia	Obs.
Hombre	0.404	0.652	-0.248*** (0.050)	215,420
Edad	40.901	37.379	3.522*** (1.150)	215,419
Años escolaridad	10.85	10.288	0.562 (0.429)	215,102
Casado	0.583	0.646	-0.063 (0.050)	215,420
Localidad urbana	0.824	0.816	0.008 (0.041)	215,420
Sector de actividad económica				
Primario	0.044	0.123	-0.079** (0.034)	215,420
Secundario	0.232	0.213	0.019 (0.043)	215,420
Terciario	0.724	0.659	0.065 (0.050)	215,420
Rama de actividad económica				
Agropecuario	0.044	0.124	-0.080** (0.034)	214,603
Construcción	0.004	0.021	-0.017 (0.015)	214,603
Industria mercantil	0.228	0.191	0.037 (0.041)	214,603
Comercio	0.107	0.194	-0.087** (0.041)	214,603
Servicios	0.617	0.468	0.149*** (0.052)	214,603

Nota: diferencia estadísticamente significativa al ***1 %, **5 %, *10 %. Errores estándar entre paréntesis. Estimaciones basadas en la segunda ronda de entrevistas en el panel 1. Muestra de personas de 16 años o más que están ocupados en todas las rondas de entrevista.

Tabla 3.3: Diferencia de medias entre buscadores de empleo que tienen o no beca

Variable	Beca	Sin beca	Diferencia	Obs.
Hombre	0.497	0.635	-0.138 (0.145)	14,512
Edad	37.366	32.306	5.060 (3.949)	14,505
Años de escolaridad	10.561	10.221	0.340 (1.249)	14,487
Casado	0.784	0.409	0.375** (0.148)	14,512
Localidad urbana	1	0.870	0.130 (0.101)	14,512

Nota: diferencia estadísticamente significativa al ***1 %, **5 %, *10 %. Errores estándar entre paréntesis. Estimaciones basadas en la segunda ronda de entrevistas en el panel 1. Muestra de personas de 16 años o más que están desocupados en la primera y segunda ronda de entrevistas.

Tabla 3.4: Diferencia de medias entre aquellos ocupados que tienen o no beca. Panel 2

Variable	Beca	Sin beca	Diferencia	Obs.
Hombre	0.498	0.653	-0.155*** (0.037)	281,394
Edad	34.696	36.740	-2.044** (0.886)	281,394
Años de escolaridad	11.398	10.007	1.391*** (0.315)	281,042
Casado	0.621	0.636	-0.015 (0.037)	281,394
Localidad urbana	0.833	0.807	0.026 (0.031)	281,394
Sector de actividad económica				
Primario	0.106	0.119	-0.013 (0.025)	281,394
Secundario	0.142	0.257	-0.115*** (0.034)	281,394
Terciario	0.742	0.619	0.123*** (0.038)	281,394
Rama de actividad económica				
Agropecuario	0.108	0.120	-0.012 (0.025)	280,102
Construcción	0.002	0.082	-0.080*** (0.021)	280,102
Industria mercantil	0.142	0.176	-0.034 (0.030)	280,102
Comercio	0.111	0.189	-0.078** (0.031)	280,102
Servicios	0.638	0.433	0.205*** (0.039)	280,102

Nota: diferencia estadísticamente significativa al ***1 %, **5 %, *10 %. Errores estándar entre paréntesis. Estimaciones basadas en la primera ronda de entrevistas en el panel 2. Muestra de personas de 16 años o más que están ocupados en todas las rondas de entrevista.

Capítulo 4

Resultados

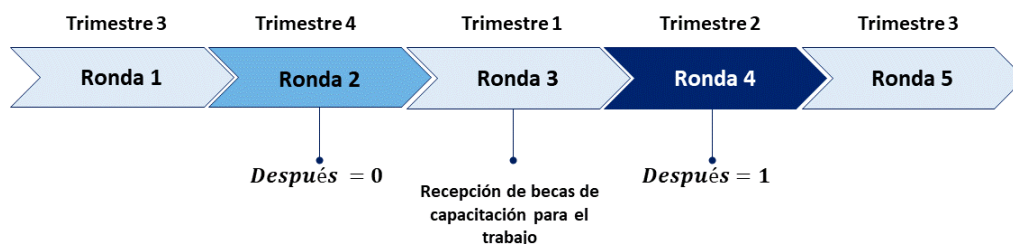
En este capítulo se presentan los resultados del modelo de diferencias en diferencias para cada uno de los paneles construidos, posteriormente se discuten los resultados y, al final, se muestra evidencia de la validez del supuesto de tendencias paralelas.

4.1. Resultados del panel 1

Primero se estima el impacto en el trimestre inmediato posterior al trimestre en el que se recibió la beca, empleando la muestra del panel 1. Como ya se mencionó previamente, el panel 1 es aquel constituido por las cinco observaciones de los individuos cuya primera ronda de entrevista se realizó en el trimestre 3 de cada año. Dado que es en la ronda 3 en la que se reporta la recepción de becas de capacitación, entonces el impacto se estima en la ronda 4, y se toma a la ronda 2 como el periodo anterior. Es decir que, para este caso, la variable *Después* en el modelo de diferencias en diferencias en (2.3) toma el valor uno para la ronda 4 y cero para la ronda 2. En la figura 4.1 se muestran gráficamente la estructura del panel 1 y los periodos que se utilizan para estimar el modelo de diferencias en diferencias para este caso.

En la tabla 4.1 se presentan los resultados de la estimación del coeficiente γ del modelo (2.3) considerando cada una de las variables de resultado, a saber, el logaritmo de ingreso mensual, las horas trabajadas, las prestaciones de salud, las prestaciones laborales y la colocación de empleo. La columna (1) contiene las estimaciones del modelo de diferencias en diferencias sin incluir variables de control; en la columna (2) se presentan las estimaciones al incluir características individuales, y en la columna (3) se muestran las estimaciones al incluir tanto características individuales como de ocupación. Las características de ocupación no se incluyen para la estimación de la variable colocación de empleo, ya que, como se mencionó anteriormente, para la estimación de esta variable se seleccionó a la población económicamente activa que no tenía trabajo pero que se encontraba realizando acciones de búsqueda durante el trimestre anterior al trimestre en el que recibió beca, de modo que durante ese periodo no presenta información sobre ocupación.

Figura 4.1: Descripción gráfica de las rondas empleadas para la construcción de cada subpanel que forma el panel 1 e identificación de la variable *Después* para medir el efecto al trimestre de haber recibido la beca



Fuente: elaboración propia

Los resultados indican un efecto significativo en los ingresos mensuales para los tres modelos y los efectos son similares. Considerando el modelo que incluye controles individuales y de ocupación, manteniendo las demás variables constantes, los beneficiarios de becas para la capacitación laboral presentan en promedio un aumento en sus ingresos mensuales del 14.43% relativo a los no beneficiarios, lo cual se puede ver en el renglón 2 de la tabla 4.1.

En las horas de trabajo se encontró un efecto mayor al incluir tanto características individuales como de ocupación. Los efectos son significativos con los tres modelos estimados. A partir del modelo con todos los controles se observa que, un trabajador que recibió beca aumentó 2.1 horas de trabajo a la semana en comparación con un trabajador que no recibió beca. Esto se observa en el renglón 3 de la tabla 4.1

Respecto a las prestaciones de salud, con el modelo sin controles se encuentra que, la probabilidad de tener acceso a prestaciones de salud incrementa en 1.9 puntos porcentuales para los individuos que recibieron becas de capacitación relativo a aquellas personas que no tuvieron acceso a las becas, esto se observa en la fila 4 de la tabla 4.1. No obstante, al incluir todos los controles se pierde la significancia que se obtuvo al no incluir ningún control y al incluir únicamente características individuales. De modo que, en este caso, el impacto de la recepción de becas sobre las prestaciones de salud es nulo estadísticamente.

Las estimaciones en las prestaciones laborales mantienen la significancia en los modelos. La magnitud del efecto es mayor al incluir características individuales y de ocupación. En la fila 5 de la tabla 4.1 se observa que el haber recibido beca aumenta en promedio 4.21 puntos porcentuales la probabilidad de que los trabajadores obtengan prestaciones laborales.

En cuanto a la colocación de empleo, vemos que con ambos modelos la magnitud y significancia del efecto estimado es similar. El modelo indica que el ser beneficiario de becas para la capacitación laboral no tiene un efecto positivo en la colocación en el empleo. La probabilidad de que los beneficiarios accedan a un empleo disminuye 6.8 puntos porcentuales relativo a los no beneficiarios, esto se puede ver en el renglón 6 de la tabla 4.1.

En resumen, con los datos del panel 1 se observa que el grupo de los beneficiarios ocupados presentan ingresos más altos, mayores horas de trabajo y son más probables de tener prestaciones laborales que aquellos no beneficiarios. Mientras que en la colocación de empleo, se observó un efecto negativo.

Tabla 4.1: Coeficientes estimados del modelo de diferencias en diferencias en el panel 1. El efecto es estimado un trimestre después de la recepción de becas.

Variable	(1)	(2)	(3)
Ingresos mensuales (logaritmo) ¹	0.1930*** (0.0072)	0.1370*** (0.0061)	0.1443*** (0.0059)
Horas de trabajo	1.8450*** (0.1502)	1.9316*** (0.1468)	2.1098*** (0.1431)
Prestaciones de salud	0.0193*** (0.0044)	0.0164*** (0.0041)	0.0050 (0.0039)
Prestaciones laborales	0.0336*** (0.0035)	0.0346*** (0.0034)	0.0421*** (0.0033)
Colocación en el empleo	-0.0668*** (0.0090)	-0.0680*** (0.0089)	
Control			
Características individuales	N	S	S
Características de ocupación	N	N	S

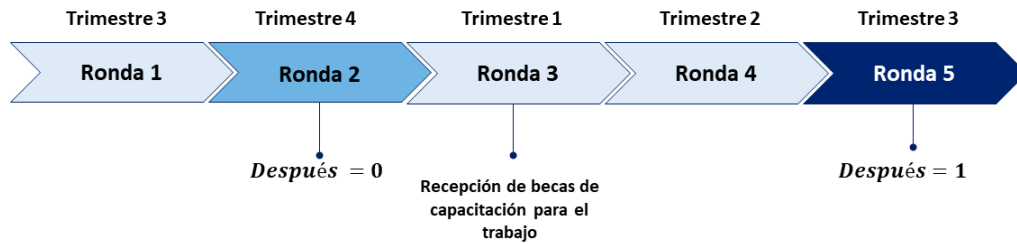
Nota: Coeficiente significativo al ***1%; **5%; *10%. Los errores estándar se presentan entre paréntesis.

¹ Ingresos a precios constantes, base diciembre de 2022. El efecto en el ingreso mensual, las horas de trabajo, las prestaciones de salud y las prestaciones se estimó sobre la muestra de individuos que estuvieron ocupados durante las 5 rondas de entrevistas, mientras que el efecto en la colocación de empleo se estimó sobre la muestra de individuos desocupados durante las rondas 1 y 2, que se encontraban buscando empleo. Estimaciones basadas en el panel 1.

Con la finalidad de identificar los efectos de la recepción de becas en el mediano plazo, se estimó el impacto dos trimestres después del trimestre en el que se recibió la beca, es decir, en la ronda 5. De modo que en este caso la variable *Después* en el modelo de diferencias en diferencias en (2.3) toma el valor uno en la ronda 5 y cero para la ronda 2. Esto se muestra gráficamente en la figura 4.2

Los resultados para este caso se muestran en la tabla 4.2. La estructura de esta tabla es análoga a la de la tabla 4.1. Se halla un efecto significativo en los ingresos cuando se estima el modelo sin controles, el efecto es similar y se mantiene la significancia al incluir solo características individuales y al incluir tanto características individuales como de ocupación. Manteniendo las demás variables constantes, los beneficiarios presentan en promedio un aumento en sus ingresos mensuales del 5.58% dos trimestres después de la recepción de la beca, relativo a los no beneficiarios,

Figura 4.2: Descripción gráfica de las rondas empleadas para la construcción de cada subpanel que forma el panel 1 e identificación de la variable *Después* para medir el efecto a los dos trimestres de haber recibido la beca



Fuente: elaboración propia.

lo cual se observa en la segunda fila de la tabla 4.2. La brecha entre este efecto y el obtenido un trimestre después de haber recibido la beca es de 8.85 puntos porcentuales, siendo mayor en el trimestre después. Es decir, parece ser que la recepción de becas para la capacitación laboral tiene su efecto de manera inmediata y conforme pasa el tiempo su efecto en el salario se va diluyendo.

Los coeficientes obtenidos para las horas de trabajo tienen significancia y magnitudes similares en las tres estimaciones. Los individuos que recibieron beca aumentaron en promedio 2.4 horas de trabajo a la semana dos trimestres después de la recepción de la beca, relativo a aquellos que no recibieron beca, esto se puede ver en la tercera fila de la tabla 4.2. De modo que, el impacto en las horas de trabajo es ligeramente mayor dos trimestres después del recibimiento de becas relativo al efecto inmediato.

En el caso de la variable de prestaciones de salud, vemos que al no incluir variables de control, la probabilidad de acceder a prestaciones de salud para los beneficiarios incrementa en 1.9 puntos porcentuales, relativo a los no beneficiarios. Sin embargo, al incluir todas las variables de control (características individuales y de ocupación), se pierde la significancia que se obtuvo al no incluir ningún control y al incluir únicamente características individuales, esto se observa en el renglón 4 de la tabla 4.2. De modo que en este caso el impacto de la recepción de becas sobre la probabilidad de tener un empleo con prestaciones de salud no es significativo a los dos trimestres de recibido el beneficio. De modo que no hay efectos significativos en las prestaciones de salud, en ninguno de los trimestres observados posteriores al recibimiento de las becas.

Para la variable de prestaciones laborales, en el renglón 5 de la tabla 4.2 se observa que el efecto es no significativo únicamente cuando no se incluyen controles. Al incluir tanto características individuales como características de empleo, se encuentra que el haber recibido beca aumenta en promedio 1.48 puntos porcentuales la probabilidad de que los trabajadores obtengan prestaciones laborales a los dos trimestres de haber recibido la beca. Esto muestra que el impacto en el trimestre posterior inmediato es mayor al obtenido dos trimestres después de la recepción de becas.

Sobre la colocación de empleo, se obtiene un efecto similar en magnitud y significancia en ambas estimaciones. El ser beneficiario disminuye 11.6 puntos porcentuales la probabilidad de acceder a un empleo a los dos trimestres de haber recibido el beneficio, lo cual se puede ver en el renglón 6 de la tabla 4.2. De modo que los efectos a los 3 y 6 meses de la recepción de becas de capacitación laboral sobre la probabilidad de acceder a un empleo son negativos. Vooren et al. (2019) menciona que los efectos negativos a corto plazo de programas de capacitación pueden asociarse a que durante dicho periodo los beneficiarios continúan en el programa de capacitación en lugar de insertarse en el mercado laboral. La demanda de tiempo de los programas de capacitación no les permite a los beneficiarios dedicarse a buscar empleo, con frecuencia, aquellos que participan en programas de capacitación dejan de lado su constante búsqueda de empleo y se centran en las actividades del programa (Card et al. 2017). Esto podría explicar por qué se obtienen peores resultados en el periodo posterior inmediato a la recepción de las becas de capacitación para los beneficiarios en comparación con los no beneficiarios.

Tabla 4.2: Coeficientes estimados del modelo de diferencias en diferencias en el panel 1. Efecto estimado a los dos trimestres de la recepción de becas

Variable	(1)	(2)	(3)
Ingresos mensuales (logaritmo) ¹	0.0691*** (0.0073)	0.0566*** (0.0062)	0.0558*** (0.0060)
Horas de trabajo	2.3453*** (0.1501)	2.4500*** (0.1465)	2.4580*** (0.1426)
Prestaciones de salud	0.0197*** (0.0044)	0.0126*** (0.0040)	0.0035 (0.0039)
Prestaciones laborales	0.0046 (0.0035)	0.0091*** (0.0034)	0.0148*** (0.0033)
Colocación en el empleo	-0.1184*** (0.0090)	-0.1166*** (0.0089)	
Control			
Características individuales	N	S	S
Características de ocupación	N	N	S

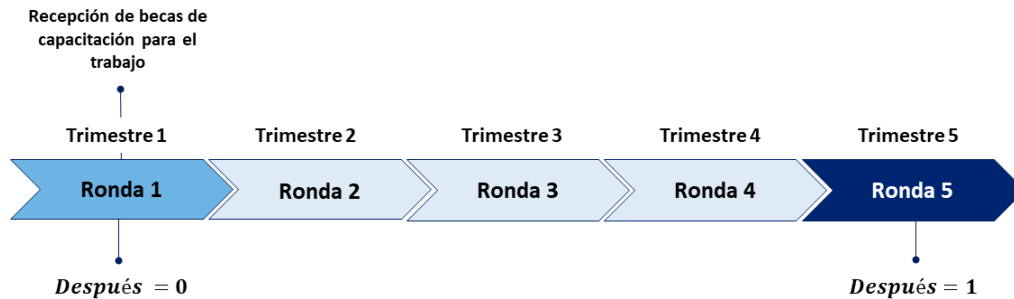
Nota: Coeficiente significativo al ***1%; **5%; *10%. Los errores estándar se presentan entre paréntesis.

¹ Ingresos a precios constantes base diciembre de 2022. El efecto en el ingreso mensual, las horas de trabajo, las prestaciones de salud y las prestaciones se estimó sobre la muestra de individuos que estuvieron ocupados durante las 5 rondas de entrevistas, mientras que el efecto en la colocación de empleo se estimó sobre la muestra de individuos desocupados durante las rondas 1 y 2, que se encontraban buscando empleo. Estimaciones basadas en el panel 1.

4.2. Resultados del panel 2

Con la finalidad de observar el impacto en el largo plazo de la recepción de becas en los ingresos, las horas de trabajo, las prestaciones de salud, y las prestaciones laborales, se emplea el panel 2. Como ya se mencionó, en este caso no se estima el impacto en la colocación de empleo, debido a la estructura del panel. El panel 2 está constituido por las cinco observaciones de la población cuya primera ronda de entrevista fue realizada en el trimestre 1 y este es el trimestre en el que se tuvo la recepción de becas, de modo que en el modelo de diferencias en diferencias, la variable *Después* se define como la variable dicotómica que toma el valor cero en la ronda 1 y el valor uno en la ronda 5. Esto se muestra gráficamente en la figura 4.3.

Figura 4.3: Descripción gráfica de las rondas empleadas para la construcción de cada subpanel que forma el panel 2 e identificación de la variable *Después* para medir el efecto al año de haber recibido la beca



Fuente: elaboración propia.

Los resultados de las estimaciones para este caso se presentan en la tabla 4.3. La columna (1) contiene las estimaciones del modelo de diferencias en diferencias sin incluir variables de control; en la columna (2) se presentan las estimaciones al incluir características individuales; y en la columna (3) se muestran las estimaciones al incluir tanto características individuales como de ocupación.

En cuanto a los ingresos mensuales se observa un efecto significativo en la estimación del modelo con y sin controles. En el caso sin controles se encuentra que el ingreso se reduce para los beneficiarios, pero al incluir controles el efecto es positivo. Las estimaciones indican que los beneficiarios de becas presentaron en promedio un aumento en sus ingresos mensuales del 9.4% después de un año de haber recibido la beca, manteniendo las demás variables constantes, lo cual se puede ver en el segundo renglón de la tabla 4.3.

Del mismo modo, en las horas de trabajo a la semana se encuentra que con los tres modelos el efecto es estadísticamente significativo y este es mayor al incluir tanto características individuales como de ocupación. Las estimaciones indican que en promedio, los individuos que recibieron beca aumentaron en 3.7 el número de horas de trabajo, un año después de haber recibido la beca, en comparación con aquellos trabajadores que no recibieron beca de capacitación, esto se observa en el tercer renglón de la tabla 4.3.

Respecto a las prestaciones de salud, en el cuarto renglón de la tabla 4.3 vemos que los coeficientes

correspondientes resultan significativos en todos los modelos, pero al incluir ambas características se obtiene un efecto mayor. Las estimaciones indican que el haber recibido beca aumenta en promedio 9.45 puntos porcentuales la probabilidad de acceder a prestaciones de salud por parte de su trabajo al año de haber recibido una beca de capacitación laboral.

En cuanto a las prestaciones laborales, en todos los modelos las estimaciones del efecto en términos de la magnitud y significancia son similares. Se observa que el haber recibido beca aumenta en promedio 2.43 puntos porcentuales la probabilidad de acceder a prestaciones laborales por parte del trabajo al año de ser beneficiario de la beca, esto se observa en el quinto renglón de la tabla 4.3.

En resumen, los efectos al año de haber recibido la beca son positivos. Es decir, el grupo de los beneficiarios presentan ingresos más altos y un mayor número de horas de trabajo a la semana que el grupo de los no beneficiarios. Además, los receptores de becas presentaron un aumento en la probabilidad de tener acceso a las prestaciones de salud y a las prestaciones laborales.

Tabla 4.3: Coeficientes estimados del modelo de diferencias en diferencias en el panel 2. Efectos estimados al año de haber recibido la beca

Variable	(1)	(2)	(3)
Ingresos mensuales (logaritmo) ¹	-0.0178*** (0.0058)	0.0605*** (0.0050)	0.0940*** (0.0049)
Horas de trabajo	3.3080*** (0.1136)	3.2297*** (0.1112)	3.7399*** (0.1099)
Prestaciones de salud	0.0782*** (0.0032)	0.0788*** (0.0029)	0.0945*** (0.0029)
Prestaciones laborales	0.0275*** (0.0027)	0.0278*** (0.0026)	0.0243*** (0.0026)
Control			
Características individuales	N	S	S
Características de ocupación	N	N	S

Nota: Coeficiente significativo al ***1%; **5%; *10%. Los errores estándar se presentan entre paréntesis.

¹ Ingresos a precios constantes base diciembre de 2022. Estimaciones sobre la muestra de individuos del panel 2 que estuvieron ocupados durante las 5 rondas de entrevistas.

4.3. Discusión

Los resultados indicaron efectos positivos en el ingreso en los tres periodos posteriores a la recepción de becas observados. En el corto plazo, es decir un trimestre después de la recepción de las becas de capacitación laboral, se observa un aumento promedio del 14.43 % en el ingreso mensual de los beneficiarios ocupados relativo al ingreso de los no beneficiarios ocupados. En el mediano plazo, es decir, a los dos trimestres de haber recibido beca, el efecto es un aumento promedio del 5.58 %, y en el largo plazo, es decir, al año de haber recibido la beca los beneficiarios presentan en promedio un aumento en sus ingresos mensuales del 9.4 % relativo a los no beneficiarios.

Este resultado es similar al encontrado por Van Gameren (2010) quien analizó el programa BÉCATE, uno de los programas de capacitación para el trabajo más relevantes en México, y encontró efectos positivos entre 700 y 1,075 pesos en el ingreso mensual después del apoyo.

En otros países también se han encontrado efectos positivos de la capacitación laboral sobre el ingreso. Un estudio, que recopila información sobre programas de capacitación similares en países de Sudamérica, como Argentina, Brasil, Chile, Colombia, Perú y Uruguay, muestra que varios de dichos programas tuvieron un impacto positivo en los ingresos de los beneficiarios (Abdala 2009).

De acuerdo con Ñopo et al. (2002), el aumento de los ingresos de las personas beneficiarias puede ser resultado de mejoras en su productividad debido a un mayor nivel de habilidades técnicas o a una mejora de desempeño en su trabajo como consecuencia de la capacitación recibida. Al adquirir nuevas habilidades y conocimientos, los beneficiarios se pueden desempeñar de manera más efectiva en su trabajo, lo que a su vez les brinda la oportunidad de negociar mejores condiciones salariales con sus empleadores o la oportunidad de acceder a puestos de trabajo con mayores responsabilidades y mejor remunerados dentro del lugar en el que trabajan. Además, es factible que después del programa, los beneficiarios que ya estaban empleados opten por cambiar de lugar de trabajo y se unan a una empresa donde puedan aumentar su productividad.

Con relación al número de horas trabajadas a la semana, se observa que los programas de becas para la capacitación laboral logran elevar el número promedio de horas trabajadas. Los efectos estimados en los tres periodos observados posteriores a la recepción de becas, son positivos y significativos estadísticamente. Un trimestre después de haber recibido la beca se observa que un trabajador beneficiario aumentó en promedio 2.1 horas de trabajo a la semana, en comparación con un trabajador no beneficiario. El efecto correspondiente a los dos trimestres de ser beneficiario de la beca es un aumento promedio de 2.4 horas, mientras que un año después, se observa un incremento promedio de 3.7 horas. Estudios previos encuentran efectos significativos en el número de horas trabajadas, por ejemplo Ñopo et al. (2002) estima que el programa de becas de capacitación laboral PROJOVEN en Perú tuvo un efecto de 5.5 % horas adicionales a la semana, equivalentes a 2.71 horas después de seis meses de concluido el programa y, menciona que, las horas adicionales de trabajo que los beneficiarios pueden alcanzar se podrían atribuir a las habilidades y conocimientos adquiridos a través de la capacitación. Esto puede llevar a que se les asignen nuevas responsabilidades, más tareas o proyectos adicionales, lo que puede resultar

en un aumento en las horas trabajadas.

También se encontraron efectos positivos en las prestaciones laborales en todos los periodos de tiempo posteriores a la recepción de becas (un trimestre, dos trimestres y un año). Mientras que para las prestaciones de salud, se encontró que los beneficiarios aumentaron sus probabilidades de acceder a prestaciones de salud pero solo en el largo plazo, es decir, las estimaciones son estadísticamente significativas un año después de la recepción de becas, pero se encuentra que no existe evidencia de efectos en el corto y mediano plazo, ya que con las estimaciones no son estadísticamente significativas un trimestre, ni dos trimestres después del tratamiento.

En cuanto a las prestaciones laborales, se encontró que el haber recibido beca aumenta en promedio 4.21 puntos porcentuales la probabilidad de que los trabajadores obtengan prestaciones laborales un trimestre después de haber recibido la beca. A los dos trimestres de haber recibido la beca, se encontró un aumento de 1.48 puntos porcentuales y al año se observa que el haber recibido beca aumenta en promedio 2.43 puntos porcentuales la probabilidad de acceder a prestaciones laborales. Sobre las prestaciones de salud, se encontró que el haber recibido beca aumenta en promedio 9.45 puntos porcentuales la probabilidad de acceder a prestaciones de salud por parte del trabajo al año de haber recibido una beca de capacitación laboral.

Van Gameren (2010) estudia también los efectos del programa BÉCATE, sobre las prestaciones laborales y de salud de los mexicanos, y en general, encuentra efectos significativos y positivos para las prestaciones laborales y el acceso a los servicios de salud, a las 13 y 26 semanas posteriores al programa, pero las estimaciones no son muy estables en las diferentes modalidades del programa. El autor sugiere que la probabilidad de encontrar efectos positivos es más grande para programas que combinan aspectos de capacitación y subsidio salarial. Algunos de los programas que se han implementado en México combinan ambos aspectos, sin embargo, en nuestro caso no es posible identificar cuáles son los programas específicos que recibieron los beneficiarios.

Generalmente las prestaciones laborales y las prestaciones de salud están asociadas a la formalidad en el empleo. En muchos países, las prestaciones laborales y de salud suelen estar vinculadas al empleo formal, donde los trabajadores están contratados de acuerdo con las leyes laborales y están registrados ante las instituciones correspondientes. Algunos estudios en América Latina hallan efectos positivos de los programas de capacitación laboral en la probabilidad de acceder a un empleo formal. Por ejemplo, una evaluación del programa Chile Joven realizado por Aedo y Pizarro (2004), encuentra que, en cuanto a la probabilidad de tener un empleo formal, los resultados sugieren una respuesta positiva, el impacto obtenido es de 19 puntos porcentuales. Así mismo, Suaya (2018) encontró que Proyecto Joven en Argentina aumentó la probabilidad de emplearse en el mercado formal un año después de haber finalizado el programa. Se encontró que la probabilidad de acceder a un empleo formal fue 2.15 puntos porcentuales superior en el grupo de los beneficiarios respecto al grupo de los no beneficiarios. Del mismo modo, Hernani-Limarino y Villarroel (2015) presentan una evaluación de impacto del programa Mi Primer Empleo Digno

(MPED) implementado en Bolivia, en donde encuentra que los beneficiarios del programa aumentaron sus probabilidades de tener un empleo formal un trimestre después de finalizado el programa, la magnitud del efecto varía en un rango entre 4 y 7 puntos porcentuales.

De manera adicional se encontraron efectos negativos en el corto plazo y mediano plazo, de la recepción de becas de capacitación laboral sobre la probabilidad de acceder a un empleo, es decir, aquellas personas que recibieron las becas fueron menos propensas a encontrar un empleo a los 3 y 6 meses de haber sido beneficiarios. Esto coincide con algunos estudios anteriores sobre el impacto de programas de capacitación en los que encuentran efectos negativos a corto plazo de los programas de capacitación en la obtención de empleo, no obstante estos efectos se vuelven positivos después de 2 años (Card et al. 2010; Card et al. 2017). Desafortunadamente, con los datos disponibles sólo es posible dar seguimiento a las personas a lo más por cinco trimestres por lo que no es posible estimar el efecto en periodos de tiempo más largos.

4.4. Comprobación del supuesto de tendencias paralelas

Aunque el método de diferencias en diferencias toma en consideración las variaciones constantes en el tiempo entre los grupos de tratamiento y control, no logra eliminar las diferencias que cambian a lo largo del tiempo entre estos grupos. Si los beneficiarios de becas de capacitación para el trabajo también se ven beneficiados de otras acciones independientes a la recepción de becas de capacitación para el trabajo al mismo tiempo que reciben dichas becas, el efecto de estas no podría separarse de las influencias externas.

Por ello, el método de diferencias en diferencias está basado en el supuesto de tendencias paralelas, este supuesto es clave para poder validar los efectos estimados. En este contexto el supuesto de tendencias paralelas significa, que sin las becas de capacitación para el trabajo, las diferencias en los resultados entre los grupos de tratamiento y control tendrían que evolucionar de forma paralela. No existe manera de demostrar que las diferencias entre los grupos de tratamiento y comparación habrían evolucionado de manera paralela si no se hubieran otorgado las becas de capacitación ya que no se puede observar lo que habría pasado con el grupo de tratamiento, sin embargo existen maneras de verificar si es admisible que dicho supuesto se pudiera cumplir.

Una manera de comprobar el supuesto de tendencias paralelas consiste en estimar el impacto de las becas de capacitación para el trabajo en un resultado falso, este es un resultado que se espera que no esté relacionado o no se vea afectado por el hecho de haber recibido beca. Si se encuentra un efecto significativo de las becas en una variable que no se espera que sea afectada, podría indicar que hay diferencias sistemáticas entre los grupos de tratamiento y control antes de la implementación del tratamiento. Esto sugiere que los resultados encontrados en el modelo original pueden ser falsos, ya que los efectos observados pueden estar influidos no solo por las becas de capacitación, sino también por las diferencias entre los grupos. Esta prueba es una estrategia importante para evaluar la validez de los resultados.

Se consideró como variable de resultado falso a la variable edad, la edad es una variable socio-demográfica que se refiere al tiempo transcurrido desde el nacimiento de una persona y, por lo tanto, está intrínsecamente relacionada con el paso del tiempo. No existen otras variables que puedan cambiar la edad de una persona. Por lo tanto se estima el siguiente modelo

$$Edad = \alpha + \beta Beca + \lambda Después + \gamma BecaDespués + \delta_1 \mathbf{X}_{CI} + \delta_2 \mathbf{X}_{CO} + u \quad (4.1)$$

En la tabla 4.4 se muestra la estimación del coeficiente γ sobre la muestra del panel 1 (en una de las dos rondas posteriores a la recepción de becas) y sobre la muestra del panel 2. El resultado en (1) es el efecto en el trimestre posterior inmediato a la recepción de becas, es decir, es la estimación del modelo (4.1) cuando la variable *Después* toma el valor uno en la ronda 3 y cero en la ronda 2; en la columna (2) se muestra el efecto a los dos trimestres de la recepción de becas, es decir, cuando la variable *Después* toma el valor uno en la ronda 4 y cero en la ronda 2; y en la columna (3) se muestra el efecto al año de la recepción de becas, la estimación se hace en el panel 2, en este caso la variable *Después* toma el valor uno en la ronda 5 y cero en la ronda 1.

Se observa un efecto no significativo de las becas de capacitación laboral en la edad en las dos rondas posteriores a la recepción de la beca. Esto brinda evidencia sobre la validez de los resultados obtenidos en el panel 1. Mientras que en la estimación del panel 2, se halla un efecto significativo de las becas de capacitación en la edad de los individuos. Esto indica que los resultados encontrados para las variables de resultado originales en el panel 2 pueden no tener validez y que pueden existir otros factores que están influyendo en los resultados.

Tabla 4.4: Efecto de las becas de capacitación para el trabajo en la edad

	(1)	(2)	(3)
Edad	0.0954 (0.0898)	0.0549 (0.0892)	0.5040*** (0.0678)

Nota: Se incluyen las variables de control de características individuales y de ocupación. Coeficiente significativo al ***1%; **5%; *10%. Los errores estándar se presentan entre paréntesis. Efecto un trimestre, dos trimestres y un año después de la recepción de becas en las columnas (1), (2) y (3), respectivamente. (1) y (2) estimaciones basadas en el panel 1; (3) estimación basada en el panel 2.

Otra manera de comprobar el supuesto de tendencias paralelas es observar los cambios en los resultados en los grupos de tratamiento y control en momentos previos a la recepción de becas. Si se observa que los resultados evolucionaban de forma paralela antes de la recepción de las becas, podría suponerse que los resultados habrían seguido evolucionando de la misma manera después. Si las tendencias de los grupos no son paralelas, es decir, si los grupos de tratamiento y control siguen trayectorias diferentes antes de la implementación del tratamiento, entonces es difícil

conocer el verdadero efecto de las becas, ya que el efecto estimado sería inválido o estaría sesgado.

Notemos que esta forma de comprobar el supuesto de tendencias paralelas solo se puede emplear en el panel 1, ya que en este panel es posible observar a los individuos dos trimestres antes de la recepción de las becas. Para probar el supuesto de tendencias paralelas en el modelo de diferencias en diferencias con momentos previos al tratamiento se pueden visualizar los datos de manera gráfica lo cual permite observar si las tendencias de los grupos de tratamiento y control son paralelas antes de la intervención, o bien se puede emplear una prueba estadística.

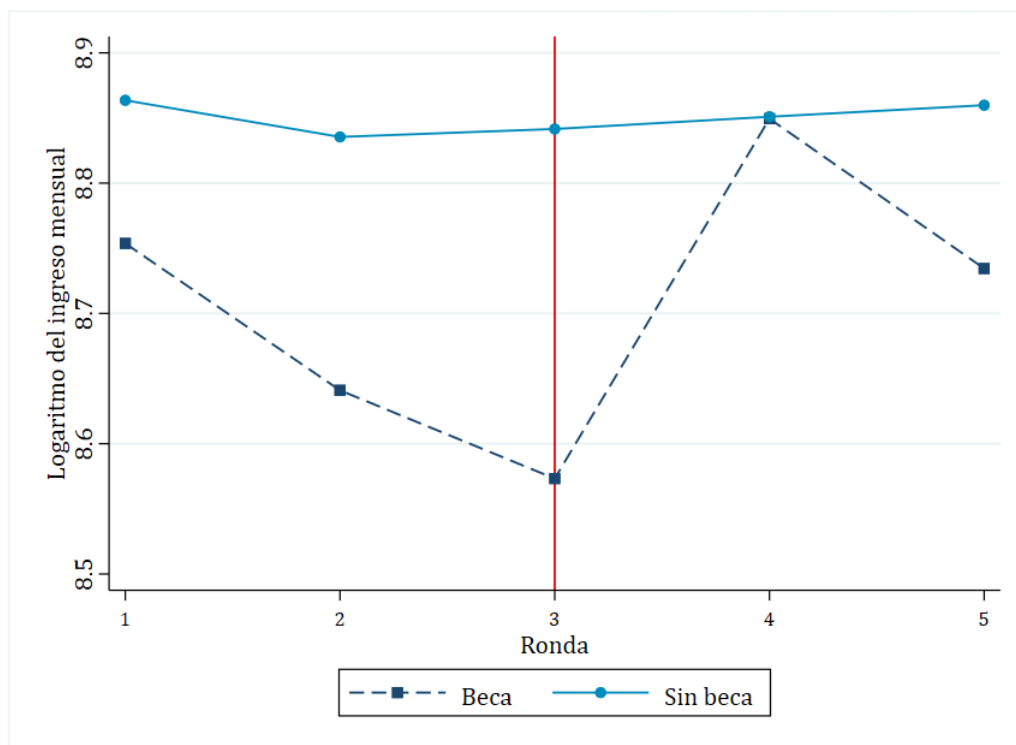
Primero, se examina gráficamente si las tendencias de las variables de resultado en los grupos de tratamiento y control son paralelas durante las rondas en el periodo previo a la recepción de becas y posteriormente se realiza una prueba estadística para evaluar si las tendencias son paralelas en el período previo al tratamiento. A continuación, se presentan las gráficas que muestran las tendencias en las variables de resultado analizadas.

La figura 4.4 muestra el incumplimiento del supuesto de tendencias paralelas en la variable de ingreso mensual, ya que se observa que en promedio las trayectorias del grupo de tratamiento y control, no se comportaban paralelamente en el periodo previo a la recepción de las becas. Por otro lado, en la figura 4.5 se observa que las horas de trabajo seguían una tendencia paralela entre el grupo de tratamiento y el de control en las rondas previas al recibimiento de las becas; en la ronda posterior al recibimiento de becas esa tendencia paralela se rompe. Esto sugiere que en cuanto a horas de trabajo el grupo de beneficiarios y el de no beneficiarios podrían haber seguido evolucionando de la misma manera.

En la figura 4.6 se observa que, el porcentaje de individuos con acceso a prestaciones de salud en el grupo de beneficiarios y, el respectivo porcentaje en grupo de los no beneficiarios, no mantenían una tendencia paralela. Sobre las prestaciones laborales, en la figura 4.7 se puede observar que los grupos de tratamiento y control muestran tendencias no paralelas en los porcentajes de individuos que tienen acceso a prestaciones laborales.

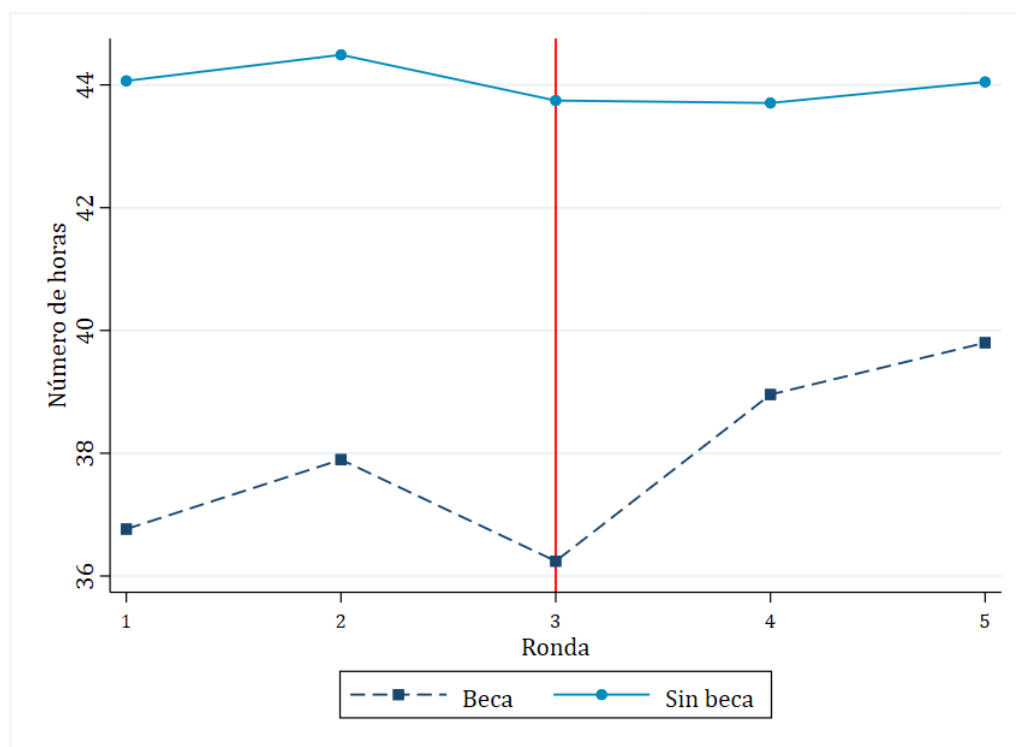
Puesto que para estimar el efecto en la colocación de empleo se restringió la muestra del panel 1 a aquellos individuos que estaban desocupados y se encontraban buscando empleo de manera activa durante la primera y segunda ronda de entrevistas, entonces todos los individuos de la muestra resultante parten de la misma condición inicial, todos se encuentran desocupados. De este modo se cumple que la tendencia en la colocación de empleo previa a la recepción de becas seguida por los individuos beneficiarios es paralela a la tendencia previa a la recepción de becas seguida por los individuos no beneficiarios (ver figura 4.8), esto brinda confianza en la veracidad del supuesto de tendencias paralelas.

Figura 4.4: Promedio del logaritmo del ingreso mensual en cada una de las rondas de observación



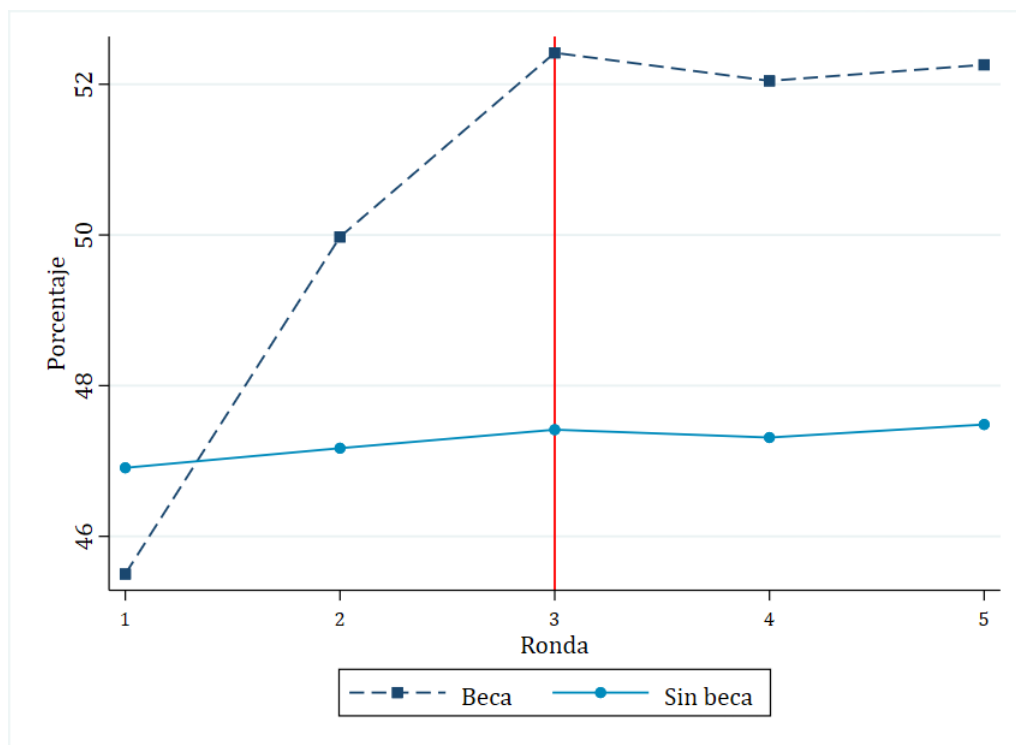
Fuente: elaboración propia con datos de la ENOE. La línea roja indica el momento del tiempo en el que se recibieron las becas de capacitación laboral.

Figura 4.5: Promedio de las horas trabajadas a la semana en cada una de las rondas de observación



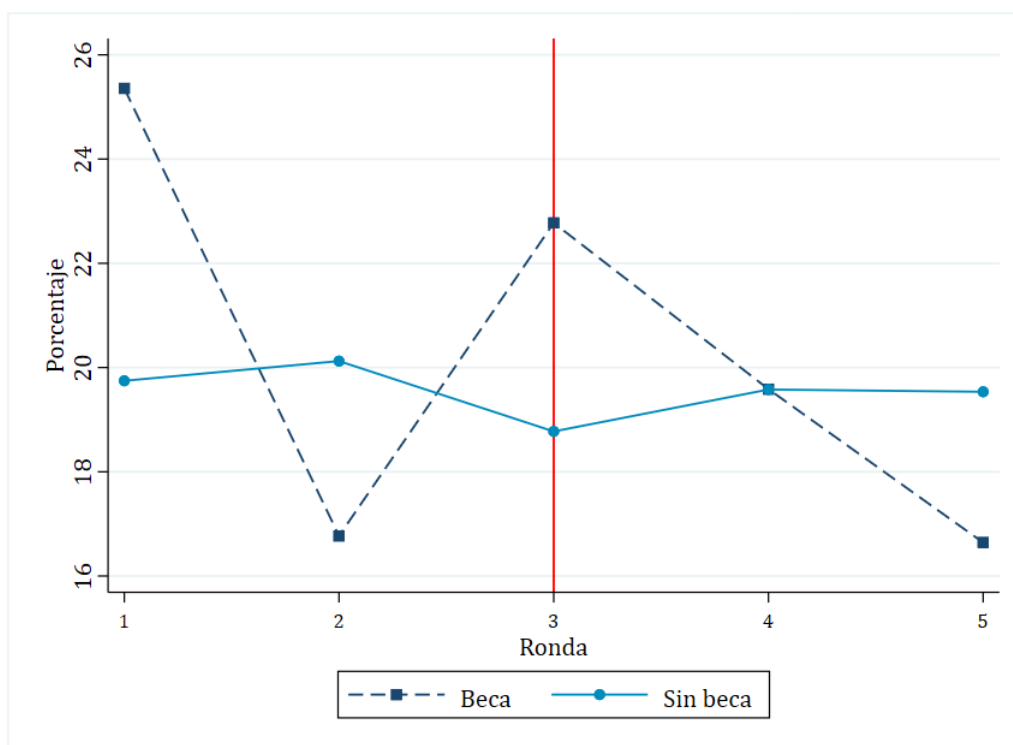
Fuente: elaboración propia con datos de la ENOE. La línea roja indica el momento del tiempo en el que se recibieron las becas de capacitación laboral.

Figura 4.6: Porcentaje de individuos con acceso a prestaciones de salud en cada ronda de observación



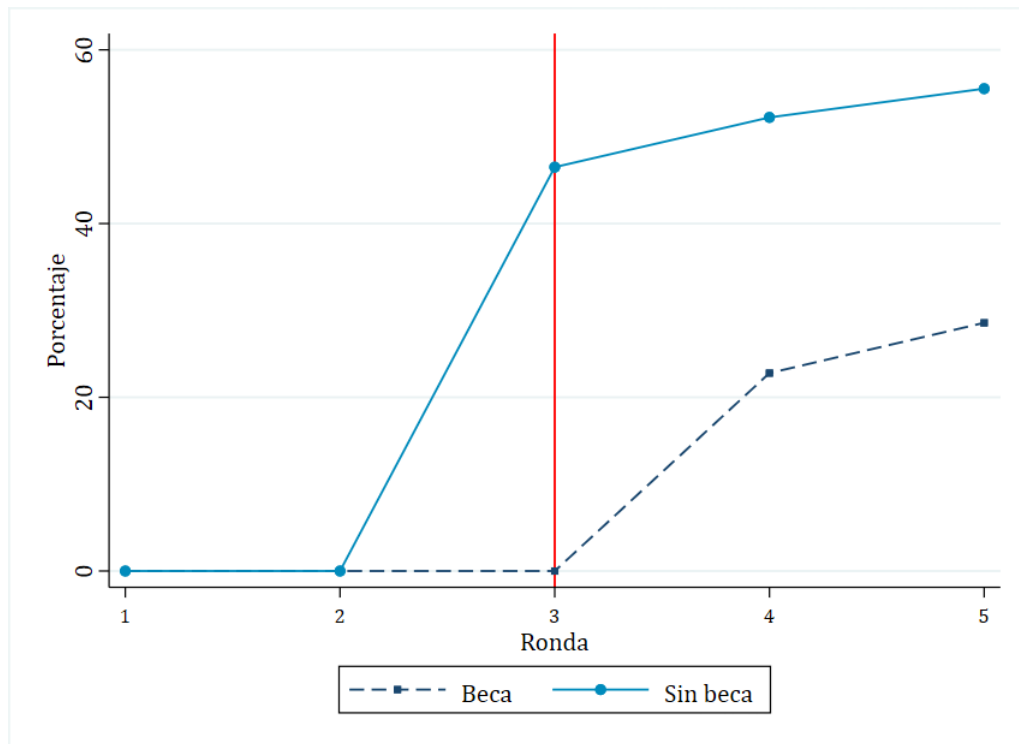
Fuente: elaboración propia con datos de la ENOE. La línea roja indica el momento del tiempo en el que se recibieron las becas de capacitación laboral.

Figura 4.7: Porcentaje de individuos con acceso a prestaciones laborales en cada ronda de observación



Fuente: elaboración propia con datos de la ENOE. La línea roja indica el momento del tiempo en el que se recibieron las becas de capacitación laboral.

Figura 4.8: Porcentaje de individuos ocupados



Fuente: elaboración propia con datos de la ENOE. La línea roja indica el momento del tiempo en el que se recibieron las becas de capacitación laboral.

Para probar el supuesto de tendencias paralelas empleando tiempos previos a la recepción de becas, se realiza también una prueba estadística para evaluar si las tendencias son paralelas utilizando los datos de tiempos previos al tratamiento. Para ello se emplea un modelo de diferencias en diferencias para cada una de las variables de resultado. En cada modelo la variable *Después* toma el valor cero para la ronda 1, y el valor uno para la ronda 2. Se examina si el coeficiente de la interacción entre la variable *Después* y la variable *Beca* es estadísticamente significativo. Si el coeficiente no es significativo, esto indica que no hay diferencias en las tendencias entre los grupos de tratamiento y control antes de la recepción de becas. El modelo que se estima es

$$Y = \alpha + \beta Beca + \lambda Después + \gamma Beca * Después + \delta_1 \mathbf{X}_{CI} + \delta_2 \mathbf{X}_{CO} + u \quad (4.2)$$

donde

La variable dependiente Y representará en cada caso, el logaritmo del ingreso mensual, horas de trabajo, prestaciones de salud, prestaciones laborales y colocación de empleo.

Beca es la variable dicotómica que toma el valor 1 si la persona recibió beca y 0 en otro caso.

Después es una variable dicotómica que toma el valor 1 en la segunda ronda de entrevistas y 0 en la primera ronda.

\mathbf{X}_{CI} es el vector de características individuales tales como edad, género, escolaridad, estado civil y si vive en localidad urbana

\mathbf{X}_{CO} es el vector de características de ocupación como, el sector y la rama de ocupación. Este vector no se incluye para la estimación de la variable de resultado de colocación de empleo, ya que para esta estimación se seleccionó a la población económicamente activa que no tenía trabajo pero que se encontraba realizando acciones concretas de búsqueda durante los dos trimestres anteriores al trimestre en el que reportó estar recibiendo beca, de modo que durante ese periodo no presenta información sobre ocupación.

u es el término de error, una variable que recoge factores que no son observables y que pueden ser relevantes en la estimación, pero que en promedio valen cero.

En la tabla 4.5, se presenta la estimación del coeficiente γ para cada una de las variables de resultado. Se observa que los coeficientes asociados a las variables ingreso mensual, prestaciones de salud, prestaciones laborales y colocación en el empleo son significativos, de modo que en este caso no se pueden validar los resultados para estas variables que se muestran en las tablas 4.1 y 4.2, ya que esto indica que puede haber diferencias entre los grupos de tratamiento y control. El coeficiente asociado a las horas de trabajo es el único que resulta ser no significativo, lo cual señala que, en el período previo a la recepción de becas, no hubo diferencias significativas en el valor promedio de las horas de trabajo, entre los grupos de tratamiento y control. Esto sugiere que las tendencias podrían haber seguido evolucionando de la misma manera en el periodo posterior y, por lo tanto, los dos grupos pueden compararse.

Lo observado sobre las horas de trabajo, gráficamente y con el modelo de diferencias en diferencias en 4.2, proporciona mayor evidencia de que las tendencias eran paralelas antes de la recepción de las becas de capacitación para el trabajo, y esto fortalece la validez de los resultados del modelo de diferencias en diferencia y proporciona más confianza en las inferencias causales que se pueden hacer.

Tabla 4.5: Tendencias paralelas

Variable	DD
Ingresos mensuales (logaritmo) ¹	-0.0704*** (0.0061)
Horas de trabajo	0.8603 (0.1473)
Prestaciones de salud	-0.1043*** (0.0034)
Prestaciones laborales	0.0672*** (0.0039)
Colocación en el empleo	0.0000 (0.0000)

Nota: : Se incluyen las variables de control de características individuales y de ocupación. Coeficiente significativo al ***1%; **5%; *10%. Los errores estándar se presentan entre paréntesis.

¹ Ingresos a precios constantes base diciembre de 2022.

Capítulo 5

Conclusiones

Algunos de los programas de capacitación que han existido en México se focalizan en las capacidades demandadas por una empresa o un trabajo específico, y colocan a los beneficiarios en actividades o puestos específicos para que adquieran experiencia laboral a través de su ocupación productiva en las empresas, con el fin de facilitar su inserción laboral. Mientras que algunos otros programas se centran en capacidades generales, brindando capacitación para el autoempleo y vales de capacitación para que los beneficiarios se incorporen a cursos de capacitación en cualquier institución privada o pública que la ofrezca. Sin embargo, a partir de los datos disponibles sólo es posible identificar a los individuos que son receptores de becas para la capacitación laboral pero se desconoce cuál es el tipo de beca que reciben. Ante esta limitación, el objetivo de este trabajo fue estimar el impacto que tiene la recepción de becas para la capacitación laboral sobre la mejora en las condiciones laborales en términos de algunas variables de interés como los ingresos, las horas de trabajo, las prestaciones recibidas y la colocación de empleo. Para ello se emplearon datos de la ENOE del 2005 al 2020.

Los resultados indicaron efectos positivos en el ingreso en los tres periodos posteriores a la recepción de becas observados. En el corto plazo, es decir un trimestre después de la recepción de las becas de capacitación laboral, se observa un aumento promedio del 14.43 % en el ingreso mensual de los beneficiarios ocupados relativo al ingreso de los no beneficiarios ocupados. En el mediano plazo, es decir, a los dos trimestres de haber recibido beca, el efecto es un aumento promedio del 5.58 %, y en el largo plazo, es decir, al año de haber recibido la beca los beneficiarios presentan en promedio un aumento en sus ingresos mensuales del 9.4 % relativo a los no beneficiarios.

Con relación al número de horas trabajadas a la semana, se observa que los programas de becas para la capacitación laboral logran elevar el número promedio de horas trabajadas. Los efectos estimados en los tres periodos observados posteriores a la recepción de becas, son positivos y significativos estadísticamente. Un trimestre después de haber recibido la beca se observa que un trabajador beneficiario aumentó en promedio 2.1 horas de trabajo a la semana, en comparación con un trabajador no beneficiario. El efecto correspondiente a los dos trimestres de ser benefi-

ciario de la beca es un aumento promedio de 2.4 horas, mientras que un año después, se observa un incremento promedio de 3.7 horas.

En el presente trabajo también se encontraron efectos positivos en las prestaciones laborales en todos los periodos de tiempo posteriores a la recepción de becas (un trimestre, dos trimestres y un año). Mientras que para las prestaciones de salud, se encontró que los beneficiarios aumentaron sus probabilidades de acceder a prestaciones de salud pero solo en el largo plazo, es decir, las estimaciones son estadísticamente significativas un año después de la recepción de becas, pero se encuentra que no existe evidencia de efectos en el corto y mediano plazo, ya que con las estimaciones no son estadísticamente significativas un trimestre, ni dos trimestres después del tratamiento.

En cuanto a las prestaciones laborales, se encontró que el haber recibido beca aumenta en promedio 4.21 puntos porcentuales la probabilidad de que los trabajadores obtengan prestaciones laborales un trimestre después de haber recibido la beca. A los dos trimestres de haber recibido la beca, se encontró un aumento de 1.48 puntos porcentuales y al año se observa que el haber recibido beca aumenta en promedio 2.43 puntos porcentuales la probabilidad de acceder a prestaciones laborales. Sobre las prestaciones de salud, se encontró que el haber recibido beca aumenta en promedio 9.45 puntos porcentuales la probabilidad de acceder a prestaciones de salud por parte del trabajo al año de haber recibido una beca de capacitación laboral.

De manera adicional se encontraron efectos negativos en el corto plazo y mediano plazo, de la recepción de becas de capacitación laboral sobre la probabilidad de acceder a un empleo, es decir, aquellas personas que recibieron las becas fueron menos propensas a encontrar un empleo a los 3 y 6 meses de haber sido beneficiarios.

La verificación del supuesto de tendencias paralelas es esencial para validar los resultados obtenidos mediante un modelo de diferencias en diferencias. Por esta razón, los resultados con mayor solidez son los obtenidos para las horas de trabajo al trimestre y dos trimestres después de haber recibido becas de capacitación laboral. Pese a que las estimaciones al año de la recepción de becas fueron significativas, la única prueba que fue posible realizarse para verificar el supuesto de tendencias paralelas no fue favorable, por lo que no se confía en los resultados obtenidos a largo plazo, los resultados obtenidos pueden ser sesgados y no reflejar con precisión el efecto de las becas de capacitación, ya que puede haber otros factores no observados que estén influyendo en los resultados. Esto significa que las diferencias observadas entre los grupos de tratamiento y control después de la recepción de becas podrían no deberse exclusivamente al efecto de las becas en sí, sino a diferencias preexistentes en las tendencias o a otros factores externos.

Para el resto de las variables tampoco fue posible validar el supuesto de las tendencias paralelas. Cuando no se cumple el supuesto de tendencias paralelas, es importante considerar enfoques alternativos o técnicas adicionales para abordar este problema. Una opción viable en este caso es el método de emparejamiento por puntajes de propensión, cuyo objetivo es establecer grupos de

tratamiento y control comparables en términos de sus características observadas. Esto es parte del trabajo futuro por desarrollar.

Una limitación de este trabajo es que no es posible emplear otras maneras para probar el supuesto de tendencias paralelas, como el uso de un tratamiento falso u otros grupos de control, para poder dar mayor solidez de los efectos estimados en las demás variables de resultado estimadas. Otra limitación de los datos empleados es que no es posible saber si antes de la ronda en la que el individuo reporta estar recibiendo beca de capacitación para el trabajo, el individuo ya se encontraba recibiendo beca, ya que la información sobre becas para la capacitación laboral solo se recopila en el primer trimestre de cada año. Es importante resaltar la importancia de recolectar la información disponible sobre los programas de capacitación laboral por cada trimestre de la ENOE, pues conocer esta información puede ayudar en la mejora de este tipo de programas que buscan mejorar ampliar las oportunidades laborales de los mexicanos.

Bibliografía

- Abdala, E. (2009). La evaluación de los programas de capacitación laboral para jóvenes en Sudamérica. *Papeles de población*, 15(59), 11-82.
- Aedo, C., & Pizarro, M. (2004). Rentabilidad económica del programa de capacitación laboral de jóvenes Chile Joven.
- Aparicio, A. (2006). Reducción del desempleo a través de ALMP: el caso de México. *Análisis Económico*, 21(47), 261-279.
- Bernal, R., & Peña, X. (2011). *Guía práctica para la evaluación de impacto*. Universidad de los Andes.
- Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *The economic journal*, 120(548), F452-F477.
- Card, D., Kluve, J., & Weber, A. (2017). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3), 894-931.
- Cervantes, D. (2011). Las políticas activas de empleo en México: el caso de los programas de formación. *Revista de Estudios Empresariales. Segunda época*, (1), 21-45.
- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2017). *La evaluación de impacto en la práctica*. World Bank Publications.
- Hernani-Limarino, W. L., & Villarroel, P. M. (2015). Capacitación Laboral y Empleabilidad Evidencia de Mi Primer Empleo Digno. *Revista Latinoamericana de Desarrollo Económico*, 35-76.
- Hogg, R. V., McKean, J. W., & Craig, A. T. (2019). *Introduction to mathematical statistics*. Pearson Education.
- Instituto Nacional de Estadística y Geografía (INEGI), (2023). Encuesta Nacional de Ocupación y Empleo. <https://www.inegi.org.mx/programas/enoe/15ymas/>
- Marchionni, M., Gasparini, L., & Edo, M. (2019). *Brechas de género en América Latina. Un estado de situación*. CAF Banco de Desarrollo de América Latina.

- Ñopo, H., Robles, M., & Saavedra, J. (2002). *Una medición del impacto del programa de capacitación laboral juvenil Projoven*. Documento de trabajo No. 36. GRADE.
- Rubio, G. J., Razo, L. A., & Loredó, L. A. (2022). Impacto de Jóvenes Construyendo el Futuro y desempleo juvenil de México. *Política y Cultura*, (57), 109-134.
- Sánchez, A. (2014). Los jóvenes frente al empleo y el desempleo: la necesaria construcción de soluciones multidimensionales y multifactoriales. *Revista latinoamericana de derecho social*, (19), 133-162.
- Sánchez, R. M., Manzano, L. D., & Maturana, L. A. (2022). Informalidad laboral, pobreza monetaria y multidimensional en Bogotá y el Área Metropolitana. *Problemas del Desarrollo. Revista Latinoamericana de Economía*, 53(208), 31-63.
- Suaya, A. (2018). *Evaluando el efecto de programas de capacitación laboral para jóvenes: el caso de Proyecto Joven*. (Tesis de maestría, Universidad de San Andrés). Repositorio digital San Andrés. <http://hdl.handle.net/10908/16026>
- Van Gameren, E. (2010). Evaluación de impacto del programa de apoyo al empleo. *Secretaría de Trabajo y Previsión Social (STPS), Ciudad de México*.
- Vite, M. Á. (2018). Género, vulnerabilidad y precariedad de los jóvenes de la Ciudad de México en la reproducción de la desigualdad social. *Espiral*, 25(71), 193-224.
- Vooren, M., Haelermans, C., Groot, W., & Maassen van den Brink, H. (2019). The effectiveness of active labor market policies: a meta-analysis. *Journal of Economic Surveys*, 33(1), 125-149.
- Wooldridge, J. M. (2010). *Introducción a la econometría. Un enfoque moderno*. Cengage Learning Editores.
- Yacila, M. d. P. (2021). Los programas de capacitación laboral para jóvenes como respuesta al desempleo juvenil. *Revista de Sociología*, (32), 77-88.